Developing and Characterising Genetically Encoded Vibrational Probes for Next-Generation Bioimaging Applications using Electronic Pre-Resonant Raman Spectroscopy

Ozan Aksakal

A thesis submitted to Cardiff University for the degree of

Doctor of Philosophy

November 2024

Molecular Biosciences Division

Biophotonics & Quantum Optoelectronics

School of Biosciences

Cardiff University

Engineering and Physical Sciences Research Council

CARDIFF UNIVERSITY

PRIFYSGOL CAERDYDD

Acknowledgements

First and foremost, I would like to give my greatest thank you to my primary supervisor, Prof. D. Dafydd Jones, for the last six years through my undergraduate studies to the end of my PhD. You have helped and supported me throughout my early academic career and on a personal level when I needed it most! Your guidance and advice have been indispensable and your trust in allowing me to take ownership of my projects and steer the direction of my thesis has been invaluable. I would also like to say thanks for giving me the opportunity for this project as this was the perfect multidisciplinary work that I was seeking after my Master's degree. I would like to thank my secondary supervisors Prof. Wolfgang Langbein and Prof. Paola Borri for the contributions and knowledge in the world of physics and Raman, coming from a biochemistry background was challenging but incredibly rewarding and you both helped me bridge that gap. And finally, thank you to my last supervisor Prof. Pete Watson, for the advice throughout my time at Cardiff and of course for the cell biology aspects of this work and looking after the mammalian cell cultures. Thank you also to the EPSRC, this whole project was possible due to the funding you provided.

One of the most important individuals in this project was Dr David Regan for all the imaging aspects within this thesis, I know how hard and long the days were imaging each sample, so for that I would like to say a big thank you! A special mention must go to Athena Zitti for being by my side for the last five years and working on this project with me, what a ride it has been! Also thank you to other members of the DDJ lab, past and present, Dr Rebecca Gwyther, Dr Rochelle Ahmed, Dr Lainey Williamson, Dr Drew Mack, Dr. Krithika Ramakrishnan, Hannah Read, Danoo Vitsupakorn and John McLarnon for their massive help and support. I would like to also thank Dr Georgina Menzies and James Davies for all the MD support I have ever needed and more. Thank you also to Dr Thomas Malcomson for the QM knowledge to supplement this thesis. And last but not least, a big thank you to Dr Pierre Rizkallah, for all your wisdom and guidance on crystallography and structure refinement.

I would like to extend my thanks to the collaborators who have supported my work in this project too. Thank you to Dr Magdalena Lipka-Lloyd in the MDI and Dr Anna Warren at DLS for the crystallisation and micro focussing crystallography. And also, thanks to Dr Andreas Zumbusch, Dr Andrea Pruccoli and Dr Franzika Rabold at Kontanz University for the epr-RS imaging of the mCherry variants.

Finally, I would like to extend my gratitude to my family and friends. My friends (you know who you are!) have been amazing in Cardiff and, at times, carried me through this PhD, so thank you for that! And a big final shout out to my parents for everything you have done in supporting my time at Cardiff University, this would not have been possible without you both, I couldn't be more thankful!

Publications

Aksakal, O., Regan, D., Zitti, A., Lipka-Lloyd, M., Rizkallah, P.J., Warren, A.J., Watson, P.D., Langbein, W., Borri, P., Jones, D.D., 2024. Genetically encoding stimulated Raman-scattering probes for cell imaging using infrared fluorescent proteins. *Pending publication.*

Abstract

The advancement of molecular imaging techniques is paramount for revealing cellular and molecular processes in their native environments. This thesis explores the development of genetically encoded Raman-active probes, utilising far-red and near-infrared fluorescent proteins (FPs) and non-canonical amino acids (ncAAs) to enhance Raman spectroscopy applications. By integrating experimental and computational methodologies, this work addresses limitations of fluorescence microscopy, including photobleaching, cytotoxicity, and limited multiplexing capabilities, paving the way for next-generation imaging techniques.

In Chapter 3, selected far-red and near-IR FPs were characterised and spectroscopically analysed to identify the most suitable candidates for Raman-active probes. Structural insights into mRhubarb720 were obtained through micro-focused crystallography, revealing key residues within its biliverdin (BV)-binding pocket. Chapter 4 examines Raman enhancement in near-IR FPs, focusing on chromophore properties and their compatibility with Raman techniques. The study evaluates the suitability of these proteins for electronic pre-resonant Raman scattering (epr-RS) in bacterial and mammalian systems.

Chapter 5 explores the incorporation of para-cyano-phenylalanine in attempt to induce through-space Raman enhancement within the biologically silent region. Molecular dynamics (MD) simulations were employed to assess structural and dynamic effects of the ncAA incorporation, while quantum mechanics (QM) simulations predicted electronic and vibrational properties of BV in protein-bound and free states. These simulations provided insights into unexplored vibrational modes and informed future epr-RS studies.

Finally, Chapter 6 evaluates the incorporation of ncAAs into mCherry chromophores, analysing changes in spectral and structural properties. MD simulations highlighted structural stability, while QM methods elucidated optical behaviour, emphasising how ncAA modifications influence chromophore functionality.

This thesis highlights the cooperation between experimental and computational approaches in developing Raman-active probes, offering a robust framework for enhancing molecular imaging. The findings contribute to a deeper understanding of chromophore interactions within biological environments and demonstrate the potential of ncAAs and far-red/near-IR FPs to transform Raman-based imaging technologies.

Table of Contents

List of Abbreviations

The following table describes the various abbreviations and acronyms used throughout this thesis. Chemical or buffer acronyms will be defined in the Methods section and not listed here. Names of proteins are omitted also as they are clearly defined in the main text. The page on which abbreviation or acronym is defined or first used is also given.

Abbreviation	Definition	Page
ACPYPE	AnteChamber Python Parser interfacE	64
AM	Amplitude Modulation	61
AOM	Acousto-Optic Modulator	61
a.u.	Arbitrary units	54
Å	Angstrom (10^{-10} meters)	6
BV	Biliverdin XI-alpha	9
CARS	Coherent Anti-stokes Raman Spectroscopy	19
CHARMM	Chemistry at HARvard Macromolecular Mechanics	31
CPU	Central Processing Unit	62
D	Debye (unit of dipole moment)	189
Da	Dalton (atomic mass unit)	56
DBS	Dichroic Beam Splitter	61
DFT	Density Functional Theory	33
DLS	Diamond Light Source	56
Emλ	Emission wavelength	9
epr-RS	Electronic pre-resonant stimulated Raman spectroscopy	22
Exλ	Excitation wavelength	9
FLIM	Fluorescence Lifetime Imaging Microscopy	1
FP	Fluorescent Protein	4
FPLC	Fast Protein Liquid Chromatography	53
FRET	Förster Resonance Energy Transfer	1
fs	Femtosecond (10^{-15} seconds)	60
GPU	Graphics Processing Unit	62
GROMACS	Groningen Machine for Chemical Simulations	30
HOMO	Highest Occupied Molecular Orbital	38
HPC	High Performance Computing	62
IR	Infrared	24
K	Kelvin	65
kJ/mol/nm	Kilojoules per mole per nanometer	64
LUMO	Lowest Unoccupied Molecular Orbital	38
M	Molar (mol/L)	43
MD	Molecular Dynamics	30
Mel	Mega-electron volts ($x10^6$)	115
MO	Microscope Objective	61
mV	Milivolt	108
ncAA	Non-canonical Amino Acid	25
nm	Nanometre (10^{-9} meters)	6

ns	Nanoseconds (10^{-9} seconds)	65
OD_{600}	Optical Density at 600 nm	52
PCA	Principle Component Analysis	176
PCR	Polymerase Chain Reaction	49
PD	Photodiode	61
PDB	Protein Data Bank	7
PMT	Photomultiplier Tube	61
ps	Picosecond (10^{-12} seconds)	67
qDIC	Quantitative Differential Interference Contrast	60
QM	Quantum Mechanics	30
QY	Quantum Yield	55
RAM	Random Access Memory	62
RMSD	Root Mean Square Deviation	32
RMSF	Root Mean Square Fluctuation	32
Rg	Radius of Gyration	32
ROIs	Regions of Interest	124
ROS	Reactive Oxygen Species	14
RPM	Revolutions per Minute	50
SASA	Solvent Accessible Surface Area	32
SEC	Size Exclusion Chromatography	53
SRS	Stimulated Raman Spectroscopy	19
TD-DFT	Time Dependent Density Functional Theory	38
μM	Micromolar (10^{-6} mol/L)	45
UV-Vis	Ultraviolet-Visible	54
WT	Wild-Type	46

1 Introduction

1.1 General background to the field of research

The field of molecular imaging has transformed the ability to observe and understand biological processes at molecular and cellular levels. It enables the visualisation of specific molecules and structures within their native environments, bridging the gap between structural biology and functional studies. This capability is critical for uncovering the dynamic interactions that drive cellular behaviour, such as protein-protein interactions, intracellular signalling pathways, and organelle dynamics. At its core, molecular imaging aims to answer fundamental questions about the inner workings of cells. How are molecular structures organised within the cell? What triggers specific biochemical events, and how do these events propagate? How do pathological processes, such as cancer or neurodegeneration, alter these dynamics? Addressing these questions requires tools capable of high sensitivity, specificity, and spatial resolution while minimising perturbations to the system being studied.

Traditional imaging techniques, such as fluorescence microscopy, have been instrumental in advancing the field. For example, fluorescent proteins and dyes, have allowed researchers to tag and track specific molecules, making fluorescence microscopy a foundation of cell imaging. However, fluorescence techniques are not without limitations, including spectral overlap due to broad emission peaks restricted to the narrow visible light region of the electromagnetic spectrum and photobleaching due to the energy input required to observe these probes due the reactive nature of the excited state (Hickey et al. 2021). Complementary approaches, such as Raman spectroscopy, have emerged to overcome these challenges by exploiting the vibrational properties of specific bonds and molecules. Raman spectroscopy provides unique spectral fingerprints that can be used to identify and quantify biomolecules, which can function outside of the visible light spectrum with less energetic (red-shifted) input in the infrared (IR) spectrum, allowing for more probes to be detected while reducing phototoxicity (Shipp et al. 2017).

Although both fluorescence and Raman techniques have proven invaluable for molecular imaging, they excel in distinct applications. Fluorescence-based approaches, especially when combined with time-resolved methods like Fluorescence Lifetime Imaging Microscopy (FLIM) and interaction-sensitive techniques such as Förster Resonance Energy

Transfer (FRET), are ideally suited for capturing rapid, subtle changes in protein-protein interactions and metabolic dynamics. In contrast, Raman-based imaging is not primarily used to monitor instantaneous molecular events but offers advantages for long-term, low-phototoxicity imaging. Its capacity for generating discrete spectral fingerprints enables superior multiplexing, allowing multiple probes to be tracked simultaneously during extended acquisition times without significant sample degradation. This complementary balance, in which fluorescence offers high sensitivity to dynamic interactions and Raman provides enhanced stability and multiplex imaging capability, forms the basis for developing next generation bioimaging tools.

As biomolecular imaging continues to evolve, the focus is shifting toward more precise, multiplexed, and dynamic approaches. These innovations not only enhance our understanding of fundamental biology but also hold promise for applications in medicine, such as disease diagnosis and drug discovery. The work presented in this thesis builds on these advancements, contributing to the development of next generation bioimaging techniques that address current limitations and open new frontiers in the field.

1.2 Fluorescence microscopy

1.2.1 Principles of fluorescence

Fluorescence is a process where a molecule absorbs light at one wavelength and subsequently emits light at a longer, red-shifted wavelength. This process is fundamental to many imaging techniques and relies on the unique photophysical properties of fluorophores/chromophores, which are molecules capable of undergoing electronic excitation that release energy as emitted light on returning to the ground state. Fluorescence has become indispensable in biology, particularly for real-time, live-cell imaging, where its high sensitivity and specificity offer unparalleled insights into dynamic biological processes.

The fluorescence process begins when a chromophore absorbs a photon, promoting an electron from the ground state to an excited state. This transition typically occurs within femtoseconds, as the energy of the absorbed light matches the energy gap between the electronic states (Datta et al. 2020). Once in the excited state, the molecule undergoes relaxation processes, such as internal conversion and vibrational relaxation, losing some

energy as heat before returning to the ground state by emitting a photon, typically within nanoseconds (Figure 1.1). This emitted photon has a lower energy (longer wavelength) than the absorbed photon, a property known as the Stokes shift. For most organic-based chromophores, the ability to undergo excitation and emission is intrinsically linked to its conjugated bond system. Conjugated systems consist of alternating single and double bonds, enabling the delocalisation of π-electrons across multiple atoms. This delocalisation lowers the energy gap between the ground and excited states, allowing the chromophore to absorb photons in the visible or near-visible spectrum.

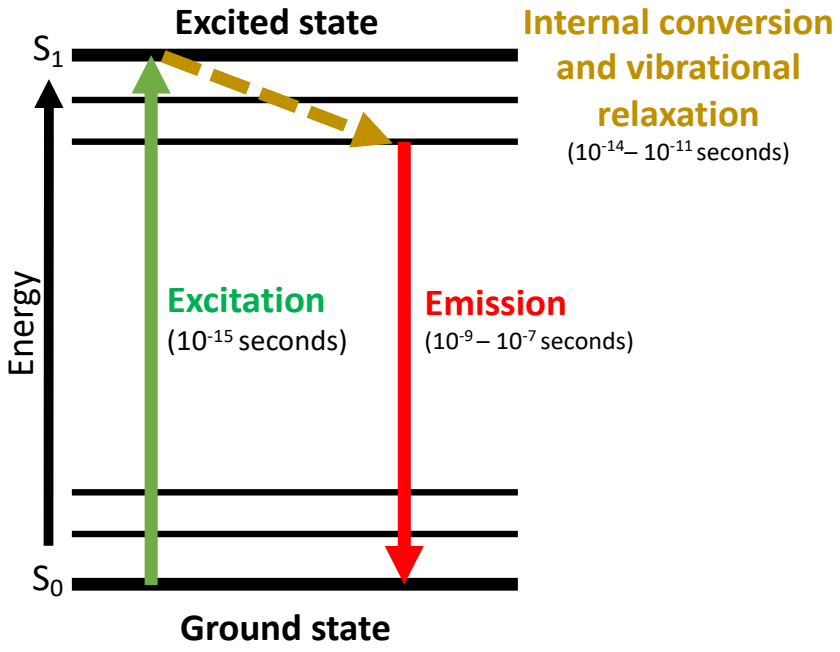


Figure 1.1: Jablonski energy diagram presenting the process of fluorescence. Excitation (green arrow) of electrons from the ground state (S_0) to the excited state (S_1) occurs within 10^{-15} seconds after the input of energy via a photon (black arrow). Internal heat conversion and vibrational relaxation occurs via non-radiative transition (dashed gold arrow) to a lower excited energy state, where emission occurs (red arrow), causing the electron return to the ground state, releasing a red-shifted photon in the process.

The efficiency of the fluorescence process is quantified by the quantum yield, which represents the ratio of emitted photons to absorbed photons. A high quantum yield indicates efficient emission, making the chromophore bright and suitable for imaging applications. There are many key factors influencing the quantum yield including the

molecular size (usually inversely proportionally) and geometry of the chromophore, the surrounding environment (e.g. solvent, temperature and pH), and the degree of non-radiative energy loss (Cranfill et al. 2016; Omary and Patterson 2017). Another characteristic inherent to fluorescence is the molar extinction coefficient, also known as molar absorption (measured in $M^{-1}cm^{-1}$), which is a measure of how likely an electronic transition from ground to excited state will occur. This indicates how much light is absorbed at a specific wavelength and is a component of calculating brightness of a fluorescent probe, as a product of the quantum yield. These quantities outline the properties of fluorescent probes, critical for deciding which probes to use in fluorescent microscopy.

1.2.2 Fluorescent proteins

The discovery of green fluorescent protein (GFP) from *Aequorea victoria* in the 1960s marked a revolution in bioimaging. The fact that GFP is a protein and can fluoresce without requiring additional cofactors made it an ideal tool for genetically tagging and tracking proteins in living cells (Shimomura et al. 1962; Shimomura 1979; Chalfie et al. 1994; Tsien 1998). The ground breaking work of Tsien, Shimomura and Chalfie earned them the 2008 Nobel Prize in Chemistry, emphasising GFP's impact on biology. GFP has since been extensively engineered to create a spectrum of derivatives with varying excitation and emission properties. These include yellow fluorescent protein (YFP) and cyan fluorescent protein (CFP). Red-shifted variants like mCherry are derived from the later discovered DsRed fluorescent protein. Protein engineering has also led to improved brightness, photostability, and pH sensitivity, expanding their applicability in diverse biological systems.

Fluorescent proteins (FPs) are unique because they can be genetically encoded and fused to specific targets, a feature that synthetic fluorescent dyes lack (Figure 1.2). Synthetic dyes often require secondary antibody conjugation or other chemical labelling steps, which can increase experimental complexity and risk altering the biological function of the target molecule. In contrast, FPs enable seamless integration into living systems, making them particularly advantageous for applications such as spatial tracking of specific biomolecules, measuring gene expression by quantifying the fluorescence output, and dynamic studies via fluorescence resonance energy transfer (FRET) to monitor protein-protein interactions.

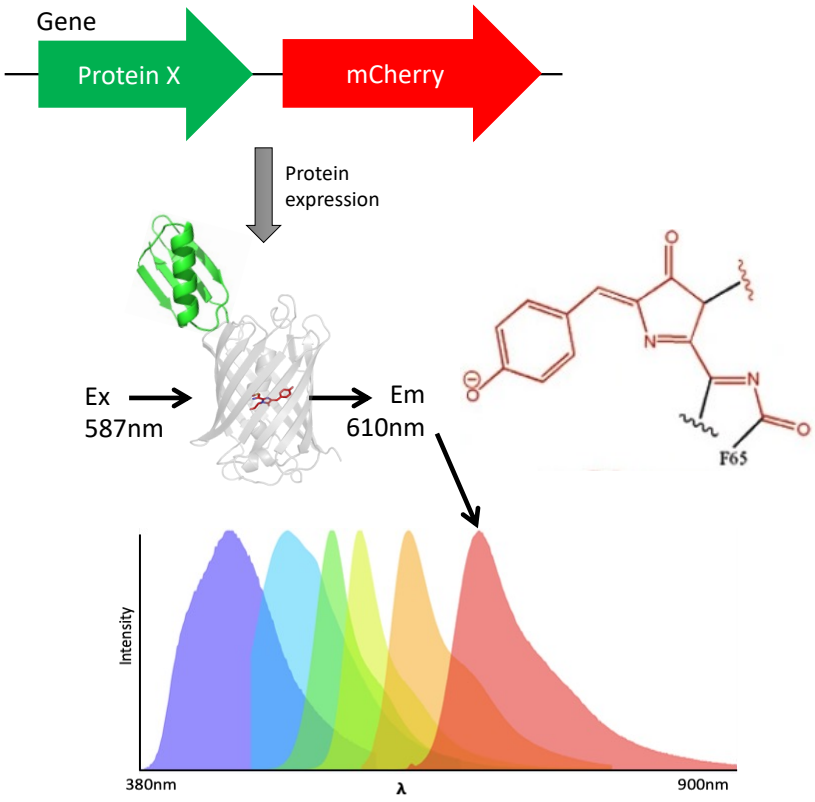


Figure 1.2: Schematic representation of the genetic encoding and spectral characteristics of fluorescent proteins, using mCherry as an example. The gene encoding Protein X is fused to the mCherry gene, resulting in the expression of a fusion protein. Upon expression, the mCherry chromophore matures (red chemical structure) within the β-barrel structure. The visible light spectrum is also shown with various emission profiles of common fluorescent proteins, with mCherry being the furthest red-shift profile.

For most β-barrel FPs, like the ones previously described, the chromophore lies at the centre of the β-barrel structure and is formed through the covalent rearrangement of three amino acids. This triad, collectively referred to as the XYG motif, typically consists of a variable residue (X), tyrosine and a glycine. The chromophore forms via a maturation process that involves three key steps: first, cyclisation of the peptide backbone occurs between the amine nitrogen of glycine and the carbonyl carbon of the variable residue. This is followed by dehydration to create a C=N double bond within the imidazolinone ring, and finally, an oxidation step forms the β-methylene bridge that links the phenolic and imidazolinone rings of the chromophore (Tsien 1998). The glycine residue is strictly conserved, while the tyrosine can be replaced by other aromatic amino acids without disrupting chromophore maturation.

The chemical maturation of the chromophore is critical to the fluorescence process, as it establishes the conjugated bond system responsible for its electronic properties. In GFP, this conjugated system is sufficient for green fluorescence.

An additional advantage of fluorescence microscopy lies in its capacity for time-resolved imaging through techniques such as FLIM combined with FRET. FLIM-FRET not only enables the precise mapping of spatial distributions but is especially proficient at capturing dynamic molecular interactions in real time (Kaufmann et al. 2020). This approach is highly sensitive to subtle changes in endogenous fluorescence lifetimes, for instance the variations in NADH and NADPH, which serve as critical indicators of shifts in cellular metabolic activity. While monitoring these lifetime changes, FLIM-FRET provides quantitative insights into protein-protein interactions and structural conformational changes, making it an invaluable tool for dissecting cellular signalling pathways (Blacker et al. 2014). Such capabilities extend the utility of fluorescence microscopy beyond static imaging, offering robust, real-time analysis of the molecular events driving cellular function.

1.2.2.1 mCherry

A main feature of this thesis is the mCherry protein, a widely used red fluorescent protein (RFP) derived from the *Discosoma* species. The mCherry was developed through extensive mutagenesis of its predecessor, DsRed, to address challenges such as slow maturation and oligomerisation (Shaner et al. 2004). These efforts resulted in a monomeric protein with faster chromophore maturation, and excitation and emission maxima at 587 nm and 610 nm, respectively. The molar extinction coefficient is relatively high compared to other blue-shifted variants with a reported 72,000 $M^{-1}cm^{-1}$, however, its quantum yield is lower with 22% efficiency compared to its predecessor DsRed with 68% (Matz et al. 1999; Shu et al. 2006). Since its introduction as part of the mFruit series, mCherry has become one of the most popular RFPs in biological research, offering a versatile tool for studying dynamic processes in living cells.

Structurally, mCherry adopts the canonical β-barrel fold characteristic of FPs, measuring approximately 28 Å in diameter and 42 Å in height (Figure 1.3). The β-barrel provides a stable, protective environment for the chromophore, shielding it from quenching by water or other polar molecules. The chromophore in mCherry undergoes an additional

oxidation step that dehydrogenates the Cα-N bond of the variable residue (methionine in the case of mCherry), forming an acylimine group. This acylimine group extends the conjugated π-electron system of the chromophore, lowering the energy gap between the ground and excited states and resulting in excitation and emission wavelengths that are red-shifted compared to GFP (Gross et al. 2000).

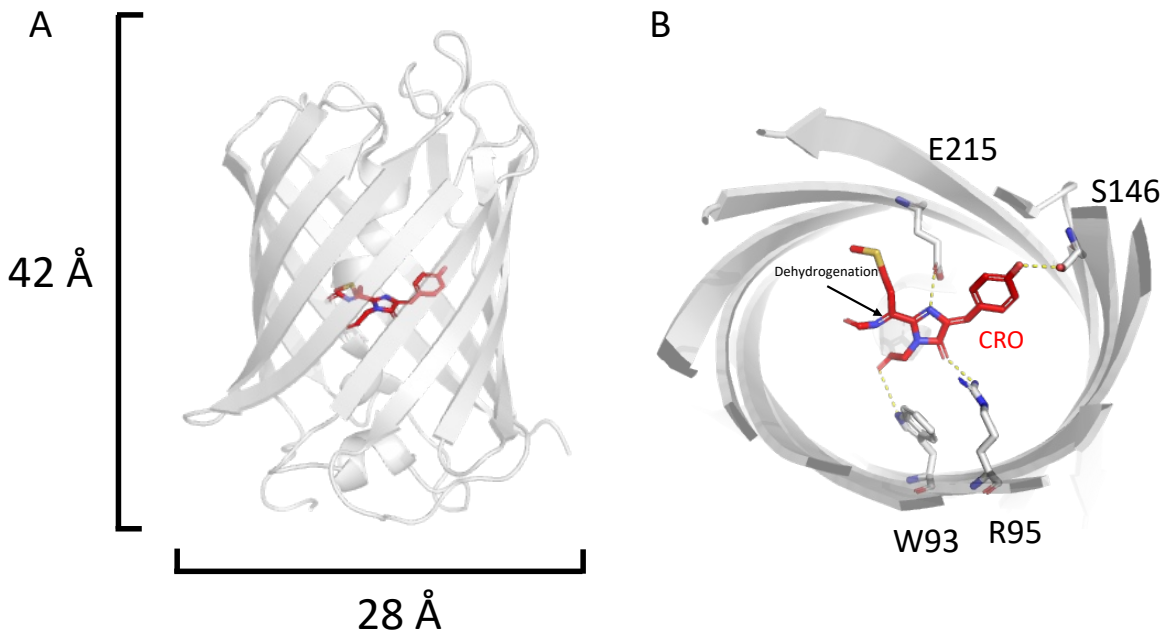


Figure 1.3: mCherry structure. (A) Side view and (B) top-down view of the crystal structure (PDB: 2H5Q), top-down view displays the polar contacts from interacting residues only (yellow dashed lines). The C=N bond that is formed via dehydrogenation is also labelled by the black arrow.

The β-barrel structure of fluorescent proteins plays a pivotal role in stabilising the chromophore and tuning the spectral properties. It protects the chromophore from quenching by external factors, such as an overload of water or oxygen, and minimises energy loss through non-radiative decay (Stepanenko et al. 2013). Interactions between the chromophore and specific residues within the β-barrel further fine-tune its optical properties. For instance, residues S146 and E215 form direct hydrogen bonds with the conjugated bond system of the chromophore, stabilising its electronic states and influencing its spectral characteristics (Figure 1.3.B). Additionally, residues W93 and R95 are positioned to stabilise the chromophore in a rigid conformation, ensuring structural integrity and

minimising potential conformational flexibility. Specific water molecules within the barrel also contribute additional hydrogen bonds, collectively ensuring the chromophore's photophysical stability (Subach and Verkhusha 2012).

These structural and functional features make mCherry a valuable tool for fluorescence-based applications. Its monomeric nature, rapid maturation, and red-shifted emission have been pivotal for multi-colour imaging, tracking protein localisation, and studying protein-protein interactions via FRET. The capacity of mCherry to be genetically encoded and fused to specific targets has further solidified its role as a cornerstone in fluorescence microscopy.

1.2.2.2 Near-IR FPs

A significant class of FPs relevant to this thesis is the near-infrared (near-IR) FPs. These proteins are structurally very different from β-barrel FPs and are engineered primarily from bacterial phytochromes, which are light-sensitive proteins found in cyanobacteria and other photosynthetic organisms. Phytochromes naturally function as photoreceptors, detecting changes in light conditions to regulate biological processes such as photomorphogenesis, seed germination, and shade avoidance in plants (Robson et al. 1996; Li et al. 2011). Their light-sensing ability arises from the covalently bound bilin chromophores, which undergo conformational and spectral changes between two states: Pr (red-absorbing) and Pfr (far-red-absorbing). In the Pr state, the chromophore absorbs red light, which induces a photochemical isomerisation of the bilin structure, transitioning the phytochrome to the Pfr state. The Pfr state, in turn, absorbs far-red light, allowing the chromophore to revert back to the Pr state. This reversible photoconversion enables phytochromes to act as molecular switches, finely tuned to sense environmental light quality and intensity. The Pr and Pfr states correspond to distinct structural conformations of the chromophore and the protein scaffold, with each state triggering specific downstream signalling pathways (Yang et al. 2009; Li et al. 2010; Macaluso et al. 2021).

This bilin-binding capability has been utilised to engineer near-IR FPs, which absorb and emit light in the near-IR spectrum, making them highly advantageous for bioimaging applications. By adapting the bilin-binding properties of phytochromes, researchers have developed near-IR FPs with biliverdin IX-alpha (BV) as their chromophore. These proteins

exploit the extended conjugated system of BV to achieve red-shifted absorbance and fluorescence properties, tailored specifically for imaging in the 650–720 nm range. Near-IR FPs are particularly useful for fluorescence imaging in living tissues, as their spectral properties reduce background autofluorescence and allow for greater tissue penetration compared to visible-spectrum FPs (Schmidt et al. 2024). However, these proteins often exhibit relatively low quantum yields compared to their visible-wavelength counterparts, presenting a trade-off between spectral advantages and photophysical performance.

A central focus of this thesis is the exploration of near-IR FPs for bioimaging applications. While their fluorescence quantum yields may be suboptimal, these proteins possess unique properties that make them invaluable for techniques beyond conventional fluorescence imaging. Specifically, their BV chromophores exhibit strong Raman scattering signals, which can be harnessed for enhanced photophysical studies. This thesis further investigates the potential of near-IR FPs in electronic pre-resonant Raman scattering, a technique that combines the high specificity of Raman spectroscopy with the benefits of pre-resonance enhancement offered by the near-IR chromophores. Although these aspects will be discussed in detail later, it is important to highlight that the unique spectral properties of near-IR FPs extend their applications beyond standard fluorescence imaging into other complementary spectroscopic techniques.

The near-IR fluorescent proteins discussed in this thesis are emiRFP670 (Excitationλ [Exλ]: 642 nm, Emission [Emλ]: 670 nm), miRFP670nano3 (Exλ: 645 nm, Emλ: 670 nm) and mRhubarb720 (Exλ: 701 nm, Emλ: 720 nm), which share the common chromophore of BV. BV is a linear tetrapyrrole pigment formed through the catabolic breakdown of heme (Wegele et al. 2004; Dammeyer and Frankenberg-Dinkel 2008). This process begins with the action of heme oxygenase, which cleaves the heme at α-methene bridge to produce BV, along with carbon monoxide and ferrous iron as by-products. As a bilin-based chromophore, BV is characterised by its extended conjugated π-system, which underpins the red-shifted absorbance and fluorescence properties of near-IR FPs. These near-IR FPs were chosen for their advantageous features, such as their monomeric nature, which ensures minimal interference with the biological systems they are used to study and makes them particularly suitable as genetically encoded probes. Additionally, their spectral properties are red-shifted into the near-IR region, enhancing compatibility with specific imaging systems. Another critical factor in their selection is their high extinction coefficients, which enhance their

absorbance properties and fluorescence output. While their quantum yields are relatively low compared to visible-spectrum fluorescent proteins, these proteins remain highly effective with the new imaging applications explored in this thesis. The miRFP670nano3 (~17 kDa) was also selected for its compact size, which is approximately half that of emiRFP670 (~32 kDa), making it advantageous for applications requiring smaller probes. The following spectral properties are previously reported in the literature, where emiRFP670 is reported to have an extinction coefficient of 87,400 M^{-1} cm^{-1} and a quantum yield of 14% (Matlashov et al. 2020). The miRFP670nano3 has an improved extinction coefficient of 129,000 M^{-1} cm^{-1} and a quantum yield of 18.5%, making it brighter than emiRFP670 (Oliinyk et al. 2022). Whereas mRhubarb720 has an extinction coefficient of 94,941 M^{-1} cm^{-1} and a quantum yield of 6.46% (Rogers et al. 2019). Despite its lower quantum yield, mRhubarb720 was chosen for its highly red-shifted spectral profile.

Two of the near-IR FPs discussed in this thesis, emiRFP670 and miRFP670nano3, have associated crystal structures (Figure 1.4). The emiRFP670 is a truncated variant of its precursor, miRFP670, with a deletion of the first 19 residues. The crystal structure of miRFP670 serves as the modified structure used in this thesis (PDB: 5VIV, Baloban et al. 2017). Both emiRFP670 and miRFP670 feature BV as their chromophore, which can form thioester linkages at one of two alternative attachment sites, C20 or C253. This results in a mixed population of chromophore configurations. The stabilisation of BV within the protein scaffold is mediated by specific hydrogen bonding interactions involving four key residues. The Y210 and R248 residues form hydrogen bonds with the carboxyl arms of BV extending from ring B, while R216 interacts with the carboxyl group protruding from ring C. Additionally, H285 forms a hydrogen bond with ring D, anchoring the chromophore within a Pr conformation, so that this ring is facing away from the other pyrroles. These interactions stabilise BV by restricting its movement within the protein scaffold and creating an environment favourable for fluorescence (Ghosh et al. 2021).

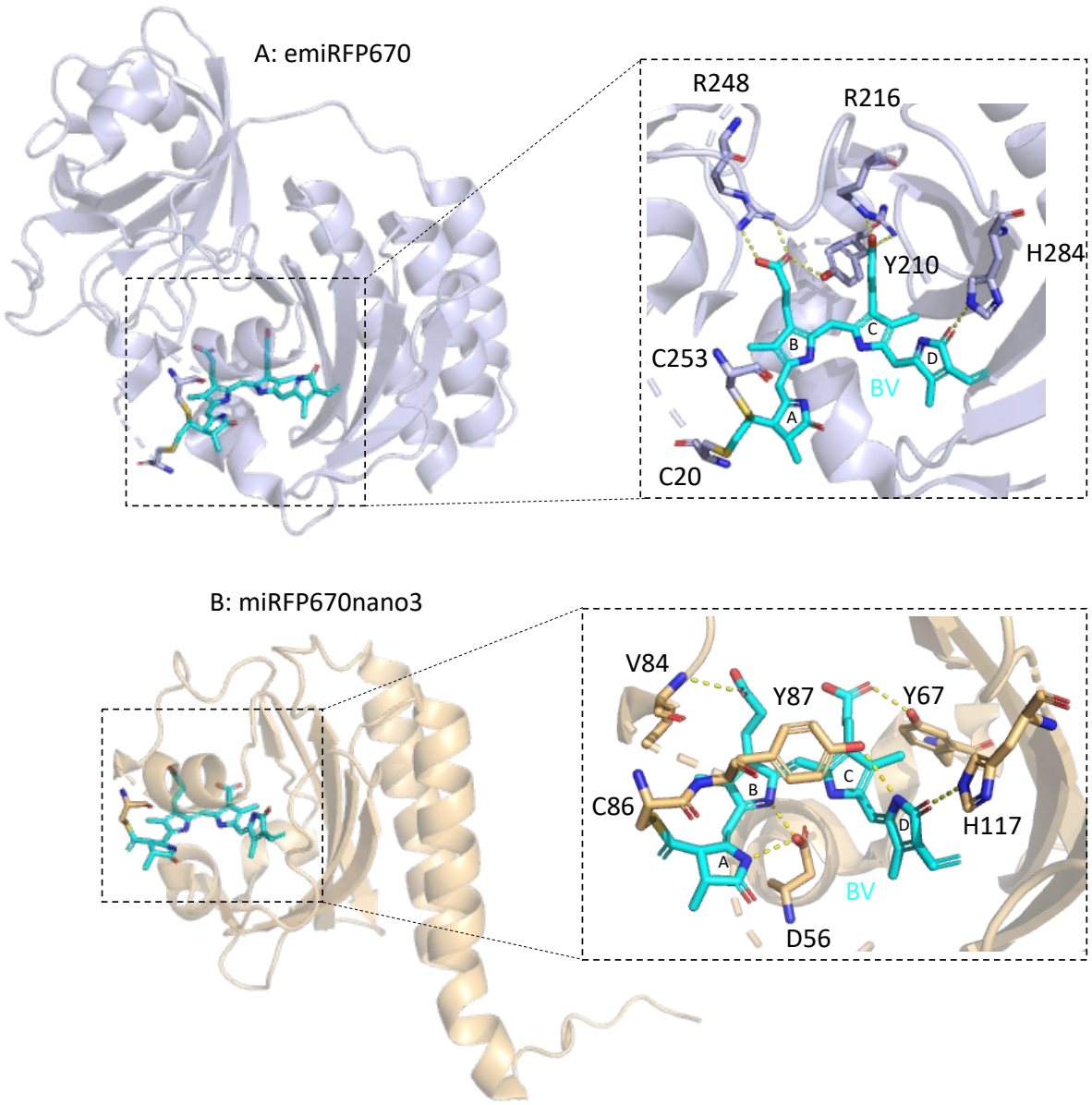


Figure 1.4: Crystal structures of two of the near-IR FPs explored in this thesis. (A) emiRFP670, modified from the miRFP670 crystal structure (PDB: 5VIV, Baloban et al. 2017) and (B) miRFP670nano3 crystal structure (PDB: 7LSC, Pletnev 2022). Polar contacts from interacting residues (yellow dashed lines) are displayed in the zoomed images of each structure. The BV is highlighted in cyan for each FP.

As mentioned, miRFP670nano3 is a smaller fluorescent protein compared to emiRFP670, yet it retains similar spectral characteristics. The reduced size of miRFP670nano3 results in a distinct set of residues interacting with the BV chromophore, contributing to its stabilisation and spectral tuning. These interactions include hydrogen bonding and covalent attachment, which collectively secure the chromophore within the

binding pocket. The backbone of V84 forms a hydrogen bond with the carboxyl group extending from ring B, while Y67 interacts with the carboxyl group extending from ring C. Additionally, D56 plays a unique role by directly hydrogen bonding to the nitrogen atoms in rings A and B, likely contributing to the rigidity of the chromophore. The covalent attachment of BV in miRFP670nano3 occurs at residue C86 via a thioester bond, analogous to the attachment at C253 in emiRFP670. This covalent linkage, along with the hydrogen bonding network, ensures that BV is firmly positioned and stabilised within the protein scaffold, enabling its optical properties. Similar to emiRFP670, H117 hydrogen bonds with ring D, restricting the chromophore to a Pr state, which is an important feature of both miRFP670nano3 and emiRFP670. This interaction secures BV within the binding pocket and maintaining the chromophore in the Pr state. The Pr state represents the red-absorbing conformation of the chromophore, which is critical for the spectral properties of these proteins. This stabilisation prevents the alternative Pfr conformations observed in bacterial phytochromes, ensuring consistent excitation and emission characteristics.

In contrast, mRhubarb720 lacks an associated crystal structure, but its BV-binding mechanism is expected to involve conserved interactions similar to those observed in emiRFP670 and miRFP670nano3. Specifically, a conserved cysteine residue is anticipated to form a covalent bond with the BV, along with a histidine residue stabilising ring D in the Pr state. Additional interactions with nearby residues are also expected to contribute to the stabilisation and spectral tuning of the chromophore. These aspects are explored in detail in Chapter 3, with crystal trails to provide insights into the structure-function relationship of mRhubarb720.

While each protein exhibits distinct structural features and spectral characteristics, their shared ability to stabilise BV within a protein scaffold highlights the importance of residue-chromophore interactions in tuning optical properties. The near-IR FPs discussed in this thesis not only highlight the potential of this class of proteins for fluorescence imaging but also extend their applicability to complementary techniques such as Raman scattering, which will be explored further.

1.2.3 Limitations of fluorescence microscopy

Despite its transformative impact, fluorescence imaging is not without limitations. These challenges are particularly relevant when attempting to extend fluorescence

techniques to more complex, longer timescale or multiplexed imaging applications. Firstly, the broad excitation and emission spectra of many FPs are intrinsically broad (50-100 nm) and lead to spectral overlap. As observed in Figure 1.2, researchers are restricted in terms of the number of FPs that can be imaged in a single sample. Typically, only 5-7 FPs per sample can be imaged using a highly specialised fluorescent microscope, which is inadequate if the aim is to analyse a whole biochemical pathway with over 10 biomolecules. Increasing the number of FPs above this threshold in fluorescence microscopy will induce the FRET phenomena, which is fine for analysing protein-protein interactions, but if the aim was to spatially track these proteins, then fluorescence emission will be skewed and undetectable per probe.

The irreversible loss of fluorescence known as photobleaching can also occur as chromophores undergo detrimental photochemical reactions upon prolonged exposure to light, rendering them non-fluorescent. This issue is exacerbated during long-term imaging or high-intensity illumination, limiting the duration and quality of fluorescence-based experiments. The by-product of photobleaching is cytotoxicity as photobleached probes release reactive oxygen species (ROS) into the cell causing apoptosis (Figure 1.5).

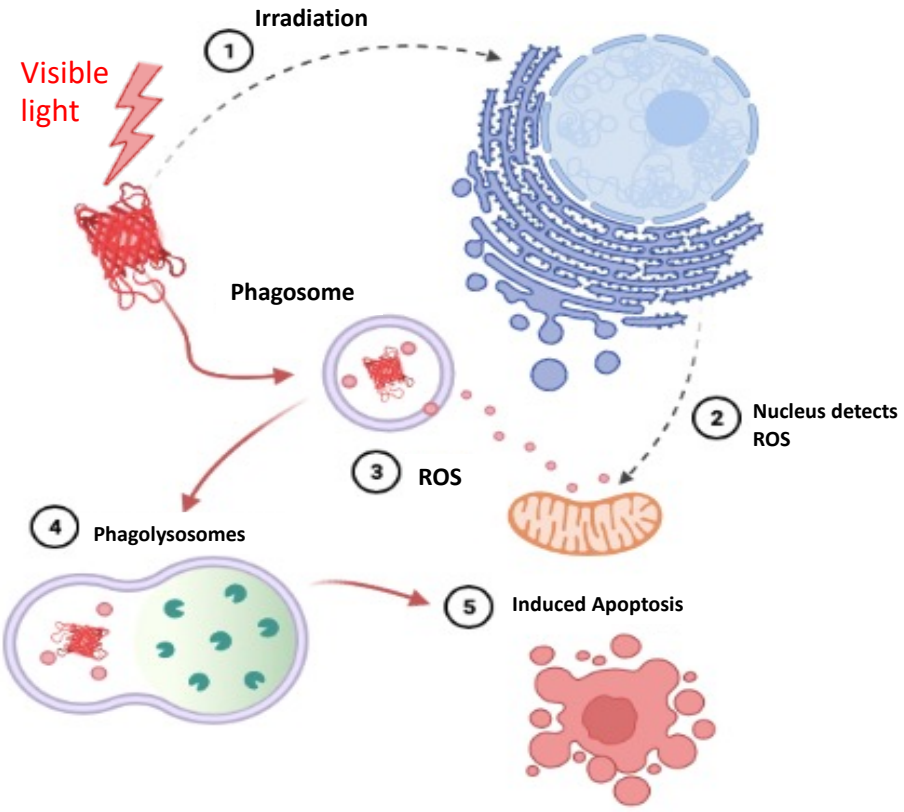


Figure 1.5: Cytotoxic effects of photobleaching during fluorescence imaging. Visible light irradiation of fluorescent proteins (Step 1) leads to photobleaching and the generation of reactive oxygen species (ROS). These ROS trigger oxidative stress responses within the cell, detected by the nucleus and mitochondria (Step 2). This initiates downstream effects, including the recruitment of phagosomes (Step 3), which subsequently form phagolysosomes (Step 4) to degrade oxidative by-products. Excessive ROS levels ultimately result in apoptosis (Step 5), a programmed cell death pathway that limits cell viability during fluorescence-based imaging.

The ROS generated during photobleaching is recognised by the cell's nucleus and mitochondria as a signal of oxidative stress (Figure 1.5, Step 2). This detection prompts a cascade of cellular responses, including the activation of pathways that lead to the formation of phagolysosomes (Figure 1.5, Step 4). Within these vesicles, damaged cellular components and oxidative by-products are degraded in an attempt to mitigate further damage. However, if the ROS levels exceed the cell's antioxidative capacity, the process concludes in the induction of apoptosis (Figure 1.5, Step 5) (Ansari et al. 2016). This form of programmed cell death is a protective mechanism to remove heavily damaged cells, but it also limits the viability of fluorescence-based experiments in living cells. The figure

emphasises the critical need to minimise photobleaching in fluorescence microscopy by optimising light intensity, exposure time, and the choice of fluorophores to reduce cytotoxic effects and preserve cellular integrity during imaging.

These limitations underscore the need for complementary imaging techniques, such as Raman spectroscopy, which offers sharper spectral peaks, minimised photobleaching, and access to other detectable areas of the electromagnetic spectrum. Together, fluorescence and Raman methods can address the challenges of modern molecular imaging, offering a more comprehensive toolkit for studying complex biological systems.

1.3 Raman spectroscopy

1.3.1 Principles of Raman scattering

The phenomenon of Raman scattering was first discovered in 1928 by the physicist Chandrasekhara Venkata Raman, a discovery that earned him the Nobel Prize in Physics in 1930. Raman's work demonstrated that when light interacts with a molecule, a small fraction of the scattered light undergoes a change in frequency, providing a direct link to the vibrational modes of the molecule (Raman and Krishnan 1928). This breakthrough laid the foundation for Raman spectroscopy, a technique that has since become an indispensable tool in chemistry, physics, biology, and materials science. Over the decades, Raman spectroscopy has enabled advances across a wide range of disciplines, from identifying unknown compounds to monitoring cellular processes in living organisms via synthetic dyes (Smith et al. 2016). In molecular biology, Raman spectroscopy combined with microscopy has emerged as a powerful imaging tool due to its label-free nature, molecular specificity, and ability to probe chemical structures with sub-cellular spatial resolution. Notably, the Raman signal is largely independent of the local molecular environment, variations in parameters such as pH, temperature, or solvent polarity exert minimal effect on its vibrational fingerprint. This inertness enables significantly longer imaging sessions with reduced photodamage and photobleaching, making Raman probes particularly well-suited for tracking metabolic processes and other dynamic cellular events over extended time periods. Together, these features make Raman micro-spectroscopy a cornerstone of vibrational micro-spectroscopy, with applications spanning drug development, cancer diagnosis and molecular imaging (Qi et al. 2023).

Raman scattering arises from the interaction of light with the vibrational energy levels of molecules. When monochromatic light, typically from a laser, illuminates a sample, most photons are elastically scattered (Rayleigh scattering), where their energy remains unchanged. However, a small fraction of photons interacts with the vibrational modes of the molecule, resulting in inelastic Raman scattering. The energy of the scattered photon either decreases (Stokes shift) or increases (anti-Stokes shift) relative to the incident photon (Figure 1.6). These energy shifts correspond to the vibrational frequencies of molecular bonds, providing a unique fingerprint of the molecule. Vibrational modes, such as stretching, bending, or torsion of bonds, can be directly identified through the Raman spectrum, which consists of sharp and well-defined peaks of Raman scattered intensity versus light wavelength or frequency, corresponding to these specific vibrations.

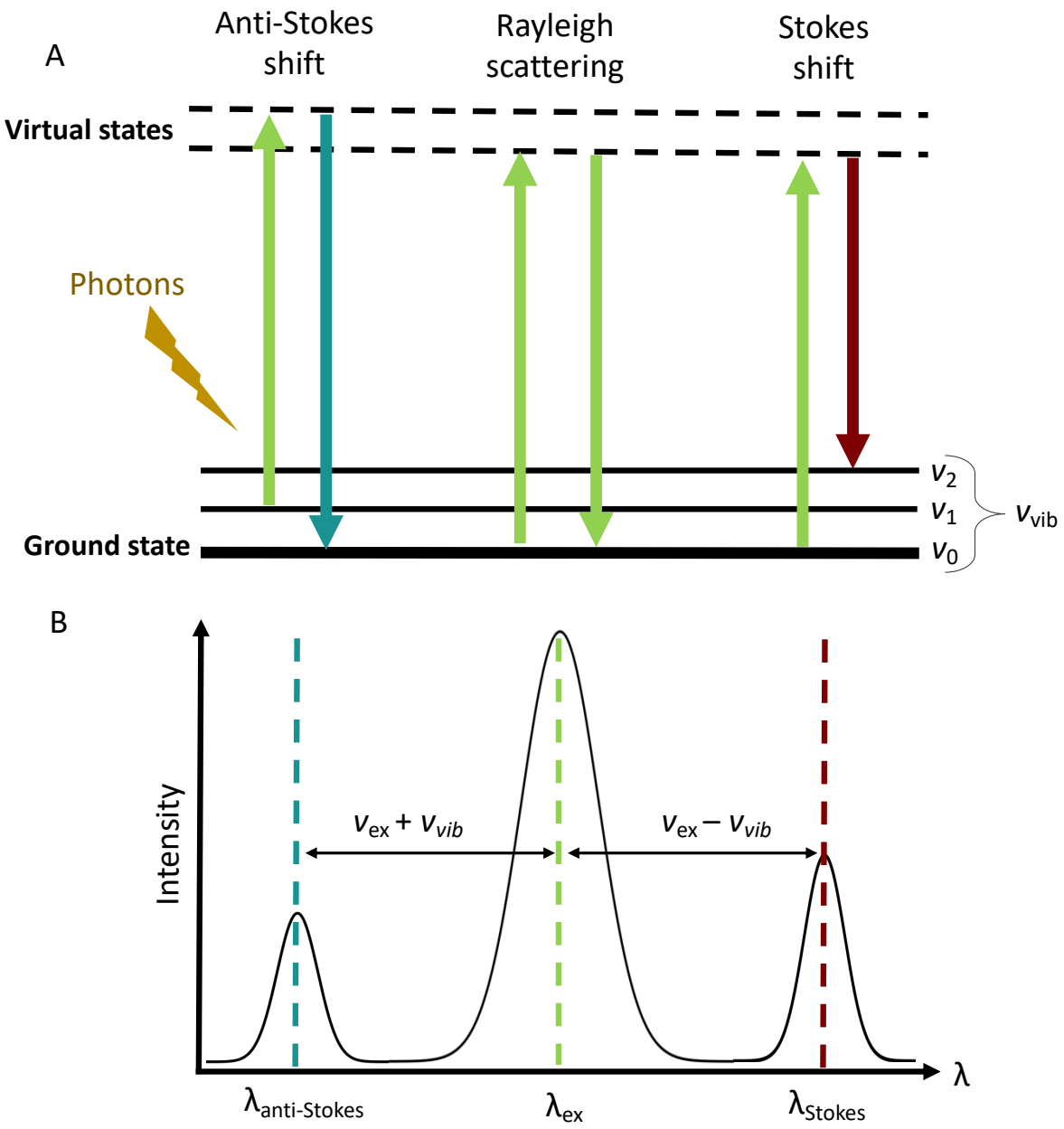


Figure 1.6: Raman scattering and types of shifts. (A) Energy diagram illustrating the different types of Raman scattering in the presence of incident photons (orange). Anti-Stokes scattering (blue arrow) occurs when photons gain energy from the excited vibrational states of a molecule (v_1), while Stokes scattering (red arrow) results in the photon losing energy to the molecule, populating a higher vibrational state (v_2). Rayleigh scattering (green arrows) represents the elastic scattering where photon energy remains unchanged. (B) Corresponding Raman spectrum, where the Stokes shift is observed at longer wavelengths and has higher intensity than the anti-Stokes shift, which occurs at shorter wavelengths.

Raman scattering provides valuable insight into the vibrational states of molecules, where the interaction of light induces energy transfer between quantised vibrational energy levels (Campanella et al. 2021). The molecules energy states are typically divided into the ground state and excited vibrational states. In Figure 1.6.A, these vibrational states are labelled v_0, v_1 and v_2, corresponding to the ground state and the first two excited vibrational states, respectively. Figure 1.6.A also shows the concept of virtual states, which are states driven non-resonantly hence shown as dotted energy levels, which do not correspond to a real energy level. Photons interacting with the molecules can drive the system to this virtual state, after which the molecule relaxes returns to a lower vibrational state, by scattering a photon with higher energy (anti-Stokes) or lower energy (Stokes) than the incident photon. In Raman spectra, the Stokes shift is generally observed with higher intensity than the anti-Stokes shift because, at standard temperatures, most molecules are typically in the ground state (v_0), making the population of molecules in excited states like v_1 or v_2 less probable (Keresztury 2006). As a result, Stokes scattering, where the photon loses energy and the molecule transitions to a higher vibrational state, is more probable. Therefore, the greater population in the ground state relative to the excited states explains why the Stokes peak tends to be more intense.

When the incident light interacts with the vibrational modes of a molecule, a Raman shift is induced, referring to the difference in frequency between the incident light and the scattered light. The Raman shift is typically expressed in wavenumbers (cm^{-1}) rather than wavelength. This is because wavenumber is a unit of frequency, hence directly proportional to the vibrational energy of chemical bonds in molecules and provides a useful quantity to interpret molecular vibrations. The Raman shift spectrum spans a range of wavenumbers, with different regions revealing important information about the molecular structure and environment. One of the most significant areas of the spectrum is the so-called “fingerprint region”, typically occurring between 400-1800 cm^{-1}. This region is rich in vibrational modes specific to the bonds and functional groups within a molecule, such as C=C (1500-1900 cm^{-1}), C=N (1610-1680 cm^{-1}) and C=O (1680-1800 cm^{-1}) stretches present in many organic compounds. These vibrational modes are highly specific to the molecular structure and the fingerprint region serves as a unique molecular identifier, akin to a ‘fingerprint’ for the molecule. It provides valuable insights into the composition, conformation, and interactions of biological molecules like proteins, lipids, and nucleic acids, without the need for labelling

or chemical treatments (Zhang et al. 2023). In contrast, the so-called “biologically silent region”, typically found between 1800-2400 cm^{-1}, is free from strong endogenous Raman signals from biological components. Incorporating non-natural chemical bonds in this region, such as nitriles (2150-2250 cm^{-1}) and ethynyls (2100-2150 cm^{-1}), should induce clear and sharp Raman peaks with improved signal-to-background ratio over the fingerprint region (Vardaki et al. 2024).

1.3.2 Coherent Raman scattering microscopy

Both Stokes and anti-Stokes shifts are central to understanding advanced spectroscopic techniques like Coherent Anti-Stokes Raman Scattering (CARS) and Stimulated Raman Scattering (SRS). These label-free methods harness the inelastic scattering of light to provide unique insights into molecular vibrations, but go beyond the spontaneous Raman scattering process described in the previous section, and rely on the interaction of two laser pulses, named pump and Stokes, onto the sample. The underlying light-matter interactions are crucial for understanding how these techniques enhance Raman signals and their respective advantages. CARS and SRS are both non-linear processes (third-order nonlinearity). In these processes, pump and Stokes laser fields having slightly different optical frequencies (ν_p, ν_S respectively) interact with the sample, driving molecules to vibrate at the corresponding frequency difference $\nu_{vib}=\nu_p-\nu_S$. As a result, identical molecules in the focal volume are driven to vibrate in sync, hence coherently. This coherence results in a constructive interference of Raman scattered light, compared to spontaneous Raman scattering, producing strong CARS and SRS signals with improved signal-to-noise ratios. In turn, CARS and SRS are capable of detecting low micromolar concentrations or can be used with high-speed acquisition, making them particularly useful for label-free imaging in living cells (Rigneault and Gachet 2012; Sun et al. 2024).

A notable challenge in CARS is the non-resonant background, caused by instantaneous electronic nonlinearities, which generates signals unrelated to vibrational resonance. This background can obscure molecular contrast and complicate data interpretation, particularly in complex biological samples (Luo et al. 2024). To address this, polarization-sensitive CARS (P-CARS) has been developed, leveraging the differing polarization properties of resonant and non-resonant signals to suppress the latter. By

adjusting pump and Stokes beam polarizations and filtering the detected signal through a polarizer, the non-resonant background is effectively eliminated while retaining vibrationally resonant signals (Oudar et al. 1979; Cheng et al. 2001). This technique enhances signal-to-background ratios, achieving extinction ratios up to ~600:1 in live-cell imaging setups and enabling clearer visualization of molecular vibrations. The use of polarization adds structural imaging capabilities, allowing researchers to probe molecular orientation and symmetry alongside chemical specificity (Brasselet 2011).

In contrast, SRS is free from this background and generates a signal that is proportional to the concentration of the target species. This proportionality makes SRS particularly easy to interpret, and often a preferred high-speed, high-resolution imaging modality, especially in complex biological systems where contrast and sensitivity are crucial (Ozeki et al. 2009).

Overall, both CARS and SRS techniques offer distinct advantages, and their complementary strengths make them well-suited to different imaging needs. Owing to the non-resonant contribution, CARS has an intrinsic referencing, which allows for quantitative image analysis (Masia et al. 2013). In contrast, SRS provides clearer contrast by eliminating the background. Together, the addition of polarization techniques in CARS and the background-free nature of SRS have significantly expanded the capabilities of Raman-based imaging, enabling label-free, chemically specific studies with high-speed and resolution. An energy schematic comparing the SRS and CARS processes illustrates the distinct mechanisms underlying these light-matter interactions (Figure 1.7).

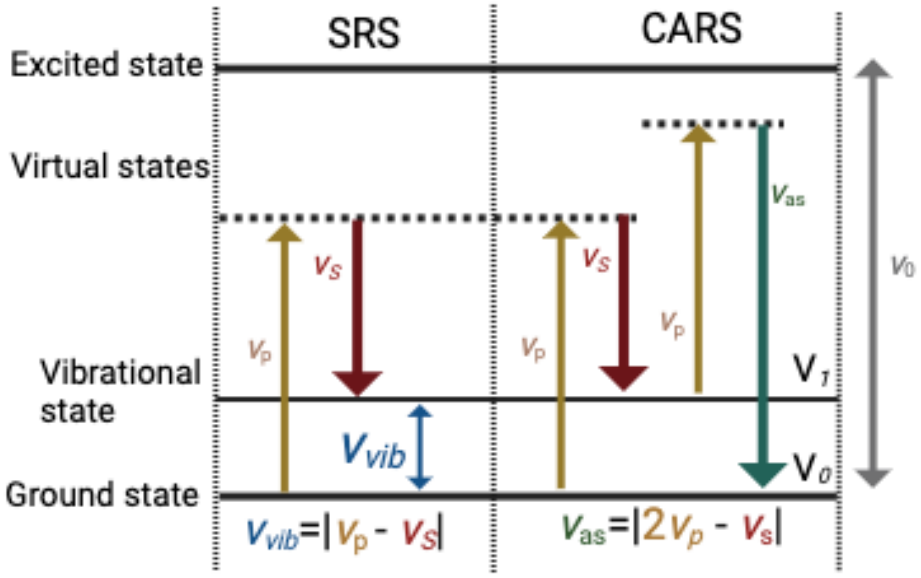


Figure 1.7: Energy diagram comparing SRS and CARS. In SRS, the interaction of pump (v_p) and Stokes (v_S) laser fields, excites molecular vibrations, where the vibrational frequency is given by $v_{vib} = v_p - v_S$. This vibrational excitation results in a measurable gain or loss in the intensity of the transmitted pump and Stokes fields, depending on the coupling to the vibrational state. In CARS, the pump field acts as a probe, producing a new coherent anti-Stokes signal at $v_{as} = 2v_p - v_S$. The grey arrow (v_0) represents the electronic energy gap between the ground state (V_0) and the excited electronic state. Virtual states represented by the dashed lines mediate the energy transfer in both processes.

As mentioned, CARS and SRS originate from the same driving process where the interference between pump and Stokes fields drive molecules to vibrate at $v_{vib} = v_p - v_s$. In two-beam CARS, the pump also acts as a probe, and the signal is detected at the anti-Stokes Raman shift $v_{as} = v_p + v_{vib}$. Conversely, in SRS, the signal manifests as a loss or gain of the incident pump and Stokes beams respectively. This requires more sophisticated measurements techniques (using high-frequency modulation) to separate the loss/gain from the strong incoming beams (Rigneault and Gachet 2012).

1.3.3 Electronic pre-resonant Raman scattering

In Raman spectroscopy, one of the primary challenges is detecting weak signals from low-concentration species, especially when they are present in complex mixtures or surrounded by other scattering and/or fluorescent molecules. The inherent Raman cross-section for many biomolecular vibrations is relatively small (typically in the range of 10^{-29} cm^2), making it difficult to obtain high-quality spectra for species that are not present in high

concentrations. Moreover, vibrational modes that scatter weakly can be easily overshadowed by background or stronger signals from other components in the sample. To address this limitation, the combination of techniques such as CARS or SRS and electronically enhanced Raman spectroscopy using chromophores provides a powerful solution. As already mentioned, both CARS and SRS offer coherent enhancement via the constructive interference of the Raman scattered light, owing to the coherent driving of vibrations by the pump and Stokes lasers, to boost the Raman signal, significantly increasing the sensitivity of the technique. Further amplifying this effect, the use of chromophores allows to exploit the so-called electronic pre-resonance enhancement effect, whereby the Raman scattering cross-section may be significantly increased for the bonds within the chromophore (Shi et al. 2018). By tuning the Pump laser frequency near the electronic absorption transition of the chromophore, the Raman cross-section is significantly increased for the species of interest (Figure 1.8).

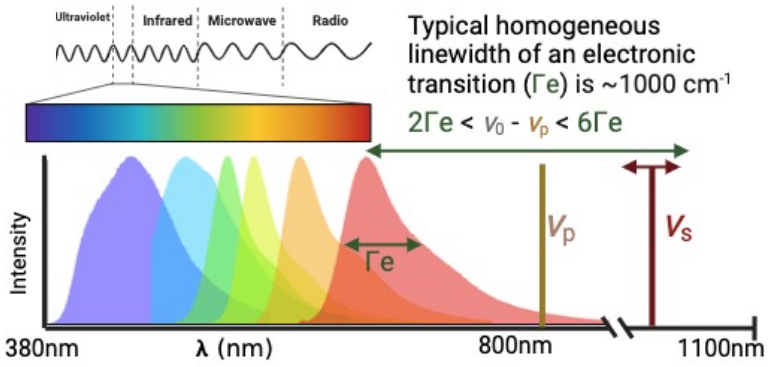


Figure 1.8: Electronic pre-resonant laser conditions. The diagram depicts the pump laser frequency (ν_p) and the Stokes scattered frequency (ν_s), illustrating the pre-resonance conditions where the pump laser frequency is tuned to lie within the linewidth of the electronic transition.

An important quantity for electronic pre-resonant Raman scattering (epr-RS) is the linewidth of an electronic transition (Γe), which represents the width of the absorption peak, typically on the order of 1000 cm^{-1} for fluorescent proteins. In epr-RS, the key is to tune the pump laser frequency such that it lies near, yet below an electronic transition frequency ν_0. This is referred to as electronic pre-resonance; in the literature, it was empirically shown that this occurs when the pump laser frequency (ν_p) is in the region where $2\Gamma e < \nu_0 - \nu_p < 6\Gamma e$, significantly enhancing the Raman signal (Wei and Min 2018). The central frequency of the

electronic transition v_0 represents the peak frequency of the absorption maximum, around which the electronic transition occurs and reaches the excited electronic state (Figure 1.7). The difference between this and the pump frequency needs to be within the pre-resonance condition to amplify the Raman cross-section, making it easier to detect low-concentration species and subtle vibrational modes. Tuning the pump frequency below the electronic transition frequency also minimises fluorescence background and reduces the risk of photodamage, both of which can obscure the Raman signal or harm the sample. This thesis will be focusing on using this technique of epr-RS in respect to both CARS and SRS to mitigate the limitations observed in fluorescence microscopy.

1.3.4 Genetically encoded Raman-active probes

The current technology in this field is dominated by synthetic Raman-active dyes that possess vibrational modes with strong Raman scattering signals, allowing for their detection in complex biological samples. These dyes have been widely used in Raman spectroscopy for imaging and spectroscopic analysis. One paper has even expanded the vibrational palette consisting of 24 resolvable reporters without photobleaching in the near-IR using epr-SRS (Wei et al. 2017). Therefore, these probes have shown to mitigate limitations of multiplexing and cytotoxicity in fluorescence microscopy. However, their use often requires sample fixation, which can limit the ability to perform real-time imaging of dynamic biological processes. Additionally, these synthetic dyes have to be supplied externally, and suffer from issues such as non-specific binding and reduced cell viability, especially in long-term experiments (Zanetti-Domingues et al. 2013).

To overcome the limitations associated with synthetic Raman-active dyes, the aim is to utilise Raman probes that can be expressed genetically, adopting the “Raman equivalent” strategy of genetically encodable fluorescent protein tags. Specifically, we have investigated monomeric far-red FPs and the near-IR FPs previously described. These proteins have been selected due to their high molar absorption coefficients (72,000 $M^{-1}cm^{-1}$ and above), which indicates a strong coupling of their electronic transitions to light. This strong coupling is important for enhancing the efficiency of pre-resonant Raman scattering. Moreover, considering the near-IR wavelength of the pump beam in the SRS set-up used in this thesis, it was important to select FPs with an absorption wavelength matching the pre-resonant

conditions, hence in the far-red to near-IR regions of the spectrum. These wavelengths are also beneficial because IR photons are less scattered (by elastic Rayleigh scattering) than visible or UV photons, thus improving penetration depth. Notably, far-red to near-IR FPs exhibit low quantum yields, which limits their use in fluorescence applications; yet this characteristic is actually advantageous in the context of epr-RS. The low quantum yield minimises fluorescence background, particularly in techniques such as CARS, where fluorescence can interfere with the Raman signal. Furthermore, a low quantum yield can contribute to increased photostability, which is essential for long-term, live-cell imaging and monitoring of dynamic biological processes.

In this thesis, two types of FPs will be considered as potential candidates for genetically encoded Raman-active probes. The first group consists of β-barrel proteins, such as the previously described mCherry, which are well-known for their stability and ease of expression in cellular systems. The second group includes bacterial phytochrome proteins, such as emiRFP670, miRFP670nano3 and mRhubarb720, that binds BV which is a natural heme breakdown product that absorbs in the 650-705 nm range (Rogers et al. 2019). As previously mentioned, when attached to its protein partner, BV has a high extinction coefficient with low quantum yield, making it an ideal candidate for Raman-active probes. These phytochrome-based systems offer significant potential for epr-RS due to their ability to absorb light in the near-IR region and their coupling to the pump pulse. These proteins are widely used as fluorescent markers but are also suitable for Raman-based imaging, especially in the fingerprint region, where their characteristic vibrational modes of the chromophores can be exploited (Figure 1.9).

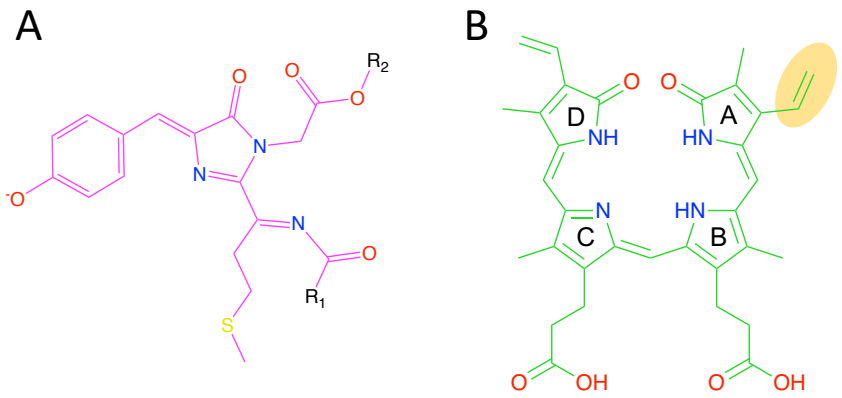


Figure 1.9: Chemical structures of the chromophores used in this thesis. (A) mCherry chromophore and (B) biliverdin XI-alpha, common in many near-IR FPs. The rings of biliverdin XI-alpha are labelled with the C=C bond highlighted in orange forms the thioester bond binding to the protein.

Both chromophores feature a conjugated double bond system, which extends electron delocalisation and facilitates lower-energy electronic transitions, making them well-suited for epr-RS with a near-IR pump. The BV contains two additional double bonds compared to the mCherry chromophore, which further red-shifts its spectral properties, enabling even lower-energy transitions. For both chromophores, the goal is to enhance bonds that comprise the conjugated system, with a particular focus on C=C stretching vibrations in the fingerprint region. While biomolecules within cells may also have C=C double bonds, they should be specifically identifiable when tagged with these chromophores, as the Raman signal from the electronic pre-resonant effect with the chromophore will stand out. Though lipids and carbohydrates also do contain C=C double bonds, the electronic pre-resonance enhancement effect of the chromophore is expected to significantly amplify the Raman signal, making it more detectable than those of these other biomolecules.

1.3.5 Genetic encoding of vibrational tags via non-canonical amino acids

The chromophores of FPs can be used to probe the fingerprint region of the Raman spectrum. However, this region contains various molecular signatures from other biological compounds, which can complicate the analysis. To address this, there is also a focus on the biologically silent region, which helps minimise interference from these other compounds. To achieve this, non-canonical amino acids (ncAAs) containing Raman-active bonds that can

be selectively enhanced in the biologically silent region need to be incorporated into these FPs. The ncAAs are modified amino acids that are not found in the standard genetic code but can be incorporated into proteins through genetic engineering. The ability to expand the chemical toolkit of amino acids allows researchers to introduce new functional groups, enabling the creation of proteins with properties not possible with the 20 natural amino acids. This expansion opens new possibilities for studying protein structure, function, and dynamics, as well as for creating biomolecular probes with unique spectroscopic properties, such as those required for Raman spectroscopy.

To incorporate ncAAs into proteins, the genetic code is reprogrammed using amber codon suppression. This method exploits the amber stop codon (TAG), which normally terminates translation of mRNA on the ribosome, and repurposes it to incorporate an ncAA at a specific site in the protein. This is achieved using orthogonal tRNA/aminoacyl-tRNA synthetase pairs, which are engineered to specifically recognise the amber codon and their corresponding ncAA without interfering with the host cell's natural translation machinery (Bartoschek et al. 2021). The modified tRNAs are charged with the desired ncAAs, and when they encounter the amber stop codon during translation, they then insert the ncAA at the designated position in the growing peptide chain. The orthogonality of the tRNA-synthetase pair ensures that only the modified tRNA recognizes the amber codon, avoiding interference with the natural translational process.

For this thesis, the ncAAs chosen include para-cyano-phenylalanine (pCNPhe), para-ethynyl-phenylalanine (pCCPhe) and meta-nitrosyl-tyrosine (3-NO_2-Tyr), which possess unique functional groups such as the nitrile, ethynyl and nitro groups, respectively (Figure 1.10). These modifications introduce unique Raman-active bonds not present in other biological molecules, allowing them to be probed in Raman spectroscopy. The ability to incorporate these Raman-active bonds into proteins enables selective tagging of target proteins and observing within the biologically silent region of the Raman spectrum, particularly for pCNPhe (2150-2250 cm^{-1}) and pCCPhe (2100-2150 cm^{-1}). In contrast, the 3-NO_2-Tyr (1340-1380 cm^{-1} and 1510-1550 cm^{-1}) produces two characteristic Raman peaks in the fingerprint region, allowing for a distinct double peak pattern which could be useful for identifying nitro-containing compounds.

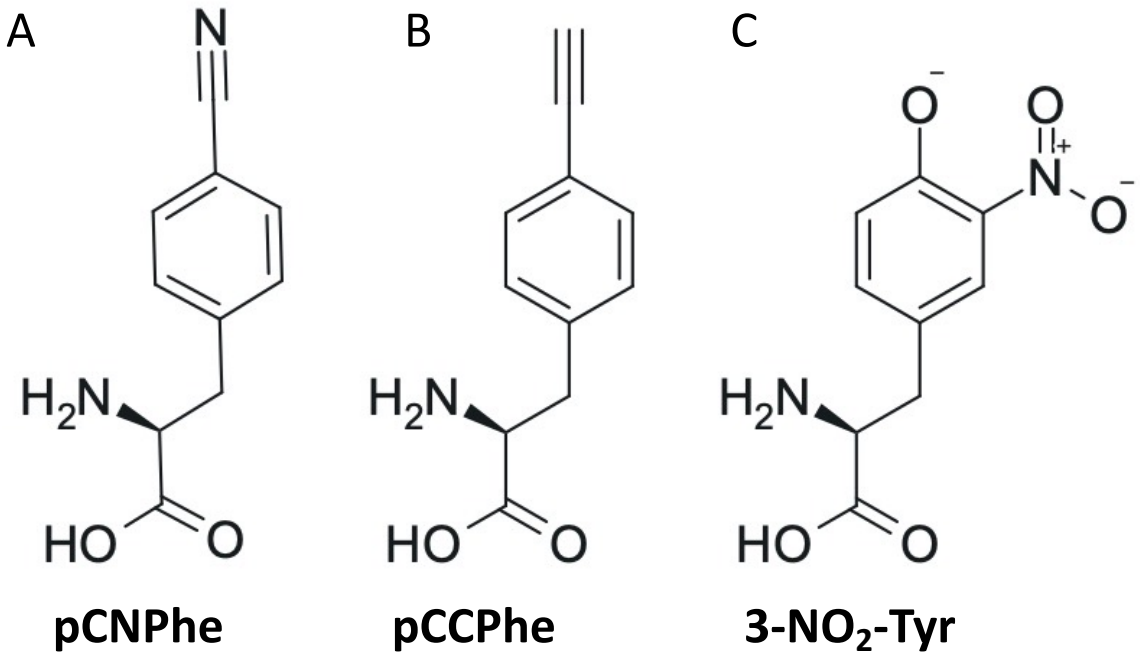


Figure 1.10: Chemical structures of the non-canonical amino acids in this thesis. (A) pCNPhe (Pearson et al. 1996), (B) pCCPhe (Stokes et al. 2000), and (C) 3-NO_2-Tyr (Katchalski 1966).

This thesis explores two distinct approaches for enhancing the Raman-active bonds within ncAAs via epr-RS. The first approach involves the direct incorporation of the ncAAs into the chromophore of FPs, particularly focusing on the chemistry of β-barrel FPs, such as mCherry. This method replaces the tyrosine residue in the chromophore with the ncAA, a strategy that has been successfully applied to various FPs and ncAAs without compromising protein function (Reddington et al. 2013; Reddington et al. 2015; Hartley et al. 2016). The second approach, through-space enhancement, aims to exploit the intrinsic electronic transitions of the chromophore in near-IR bacterial phytochrome proteins that bind BV. As BV is a cofactor, we cannot easily directly incorporate the chemistry in Figure 1.10 into the chromophore. Instead, ncAAs are placed at strategic positions adjacent to BV to potentially couple to its electronic transitions and thereby enhance the Raman signal via spatial proximity, rather than direct chemical bonding. This coupling occurs when vibrational modes of nearby bonds interact with the transition dipole moment of the chromophore's electronic transition (Dietzek et al. 2018). Vibrational modes that modulate the electronic energy landscape can effectively link the vibrational and electronic states, amplifying the Raman signal through vibronic coupling.

To clarify, fluorescent protein chromophores possess two distinct types of dipole moments: the permanent dipole moments, which reflect static charge separation in the ground or excited state, and the transition dipole moments, which describe charge redistribution during an electronic transition (e.g. $S_0 \rightarrow S_1$). Raman scattering intensity is governed not by changes in permanent dipoles, but by how vibrations alter the molecule's polarizability, which is a property that is strongly influenced by the strength and orientation of the transition dipole moments when under resonant excitation. While calculations in this thesis report permanent dipole moments, which reflect the overall polarity of the chromophore and contribute to electrostatic interactions, it is the transition dipole moment of the $S_0 \rightarrow S_1$ transition that plays a central role in enhancing Raman scattering. This transition dipole moment governs the oscillator strength of the transition and the extent to which the chromophore can couple to nearby vibrational modes. Hence, when ncAAs are placed near BV, they may benefit from this coupling pathway, leading to enhanced Raman signals, even without being directly incorporated into the chromophore.

These two approaches highlight the versatility of genetically encoded ncAAs as a powerful tool for Raman spectroscopy. Harnessing the ability to precisely position Raman-active bonds within proteins, offers epr-RS a unique opportunity to study protein dynamics with high specificity and minimal background interference when using the biologically silent region.

1.3.6 Comparative overview of detection limits

Fluorescence-based imaging techniques have demonstrated exceptional sensitivity in live-cell applications. For instance, FLIM-FRET is capable of detecting minute changes in endogenous fluorescence lifetimes, with detection limits down to approximately 10 nM (Dezhurov et al. 2011; Lai et al. 2022). This high sensitivity enables precise monitoring of rapid molecular interactions, such as protein-protein tracking and enzyme-cofactor binding, providing real-time insights into dynamic cellular processes.

In contrast, Raman-based imaging methods, including SRS/CARS, and particularly epr-RS, generally exhibit higher detection thresholds, typically in the 0.1 – 1 μM range (Miao et al. 2021; Qian and Min 2022). Although this represents a lower intrinsic sensitivity compared to FLIM-FRET, Raman spectroscopy offers significant compensatory advantages.

The vibrational fingerprints in Raman spectra are largely independent of the local molecular environment; variations in factors such as pH, temperature, or solvent polarity exert minimal influence on these signals. This environmental inertness allows Raman-based methods to support extended, long-duration imaging sessions without the severe photobleaching or phototoxicity issues that can limit fluorescence approaches. Consequently, Raman probes are particularly well-suited for non-invasive, long-term tracking of metabolic processes and other dynamic events in complex biological systems. In addition, the superior spectral resolution of Raman techniques enables the simultaneous detection of multiple molecular species, facilitating multiplexing even in densely populated samples.

Furthermore, advanced signal enhancement strategies such as epr-RS have been shown to boost the Raman cross-section, thereby partially mitigating sensitivity limitations (Wei et al. 2017; Wei and Min 2018; Pruccoli et al. 2023). These enhancements can narrow the gap between the lower intrinsic sensitivity of Raman scattering and the rapid, low-concentration detection capabilities of fluorescence. In applications where prolonged, non-destructive imaging and high multiplexing capability are essential, the trade-off in absolute sensitivity is offset by these advantages. Therefore, while fluorescence methods like FLIM-FRET excel in resolving transient, low-concentration molecular events, Raman imaging is optimised for robust, extended observation with high chemical specificity, a combination that makes it an invaluable complementary tool in modern molecular imaging.

1.4 Computational methods

1.4.1 *In silico* modelling

In silico modelling refers to the use of computational techniques to simulate and analyse molecular systems. This approach has become an indispensable tool in modern research, allowing scientists to explore molecular behaviour in a virtual environment, reducing the need for extensive experimental work and enabling predictions that would be difficult or impossible to observe directly. The development of *in silico* modelling can be traced back to the early days of computational chemistry, with the arrival of molecular dynamics simulations and quantum mechanical calculations, which provided powerful methods for exploring atomic and molecular interactions in great detail. Historically, the field of *in silico* modelling began with the development of basic computational methods to solve

the Schrödinger equation for small molecules (Stillinger et al. 1995). As computational power increased, more sophisticated techniques emerged, allowing for the simulation of larger and more complex molecular systems. The introduction of high-performance computing in the latter half of the 20th century marked a significant turning point, enabling the application of these methods to real-world biological and chemical systems. The introduction of software such as GROMACS (GROningen Machine for Chemical Simulations), a tool for molecular dynamics (MD) simulations, and Gaussian09, which uses quantum mechanical (QM) methods, further expanded the possibilities of computational studies in various scientific disciplines, including chemistry, biophysics, and materials science.

The fundamental principle behind *in silico* modelling lies in representing the system of interest in a computational framework, whether a small molecule, a protein or a complex material. These models are based on physical laws, such as classical mechanics for molecular dynamics simulations, and quantum mechanics for electronic structure calculations, which provide insights into the behaviour of atoms and molecules at both the electronic and atomic levels. Solving the equations that govern these interactions can help researchers predict properties such as protein dynamics, geometry of molecules, energy levels, optical properties and reactivity, offering invaluable insights into the systems behaviour. Furthermore, *in silico* modelling can be used to simulate experimental conditions, predict reaction pathways, and suggest modifications to optimise molecular interactions or material properties. The applications of *in silico* modelling are vast, ranging from drug design, where it is used to predict the binding affinity of drug molecules to target proteins, to materials science, where it can help design novel compounds with desired properties. In biochemistry and biophysics, *in silico* models enable the study of protein folding, ligand binding, and other dynamic and electronic processes that occur on timescales that are often too fast or too short to study directly through experimental means (Sasidharan et al. 2023). These simulations can also be used to complement experimental data, helping to interpret and refine findings from techniques like X-ray crystallography, NMR, and Raman spectroscopy.

To perform these simulations, researchers rely on specialised software to visualise molecular structures. For example, PyMOL is a widely used molecular graphics tool that allows users to manipulate and analyse 3D structures of proteins, nucleic acids, and small molecules (Schrödinger 2015). Its ability to display large structures in three dimensions, analyse aspects of the structure and generate high-quality renderings makes it an essential

tool in structural biology. On the other hand, GaussView, which interfaces with Gaussian09, is used for quantum mechanical simulations, enabling researchers to perform electronic structure calculations, predict molecular and optical properties, and explore reaction mechanisms (M. J. Frisch et al. 2016). Together, these tools provide researchers with powerful capabilities to investigate molecular behaviour *in silico* and guide experimental design, optimising both time and resources.

1.4.2 Molecular dynamics simulations

Molecular Dynamics (MD) simulations represent a powerful computational approach to studying the motion of atoms and molecules over time. Solving Newton's equations of motion for a system of particles allows MD simulations provide a dynamic, atomic-scale view of molecular systems. Unlike static models, MD simulations track the temporal progression of molecules, capturing the full range of motions at atomic-scale resolution. This allows for the exploration of molecular behaviour over extended timescales, from femtoseconds to microseconds, depending on the system of interest. At the heart of MD simulations lies the concept of forcefields, which are mathematical descriptions of the interatomic interactions governing the system (Badar et al. 2022). These force fields define how atoms within a molecule interact with one another and how their positions evolve over time. Commonly used forcefields in MD simulations include CHARMM (Chemistry at HARvard Macromolecular Mechanics) and Amber (Assisted Model Building and Energy Refinement) (Brooks et al. 2009; Bjelkmar et al. 2010; Salomon-Ferrer et al. 2013). These forcefields provide parameters that describe bond lengths, angles, torsional rotations, van der Waals forces and electrostatic interactions, allowing the simulation of molecular systems with a high degree of accuracy. Although force fields cannot fully capture all quantum mechanical effects, they offer an efficient way to model large systems, such as proteins, lipids and nucleic acids, over realistic timescales, providing critical insights into their behaviour in environments that closely resemble experimental conditions (Guvench and MacKerell 2008).

One of the major strengths of MD simulations lies in their ability to study protein dynamics and for this thesis, chromophore stability. Proteins are highly dynamic molecules, constantly undergoing conformational changes that are central to their biological function. MD simulations provide a detailed picture of these dynamical processes, allowing

researchers to observe the movement of proteins, conformational changes upon ligand binding and the impact of mutations on protein stability. The types of analysis used in this thesis includes the Root Mean Square Deviation (RMSD), which tracks the overall structural stability of the protein over time; Root Mean Square Fluctuation (RMSF), that measures the flexibility of individual residues; Solvent Accessible Surface Area (SASA), to provide insights into how the protein's surface interacts with the solvent; Radius of Gyration (Rg), which indicates the compactness of the protein structure; and hydrogen bond counting, that helps track the stability and interactions between specific residues or between the protein and ligands. These analyses together offer a comprehensive understanding of the protein's dynamic behaviour and its interactions with the surrounding environment. Similarly, MD simulations are invaluable in understanding the behaviour of chromophores, as simulating the protein/solvent environment around a chromophore, researchers can gain insights into its stability, interactions with surrounding residues, and its role in the overall function of the protein, using the described analyses.

In addition to providing detailed structural and dynamical insights, MD simulations are increasingly being used to guide experimental design. Simulating a wide range of experimental conditions, researchers can identify the most promising configurations or modifications for further investigation, reducing the time and cost associated with experimental trials. For example, MD can be used to predict how a particular ligand or ncAA will interact with a protein, allowing researchers to select the best candidates for synthesis and experimental validation. In this thesis, MD simulations were also used to generate starting models for QM simulations of the chromophores, providing a solid foundation for more detailed electronic structure calculations. While the approach in this thesis involved conducting experiments first and analysing them computationally afterward, which is an approach that may not be the optimal strategy, this methodology provided valuable insights into the into the interpretation of the experimental results and helped inform the design and refinement of future experiments.

1.4.3 Quantum mechanics

Quantum mechanics (QM) emerged in the early 20th century, revolutionising the understanding of atomic and molecular behaviour. The initial breakthroughs came from the

development of the quantum theory of radiation by Max Planck and Albert Einstein's explanation of the photoelectric effect, which demonstrated the particle-like behaviour of light (Planck 1901; Einstein 1905). In the years that followed, Werner Heisenberg's uncertainty principle, Erwin Schrödinger's wave equation, and Niels Bohr's model of the atom provided a more detailed understanding of the behaviour of particles at atomic and molecular scales (Bohr 1913; Schrödinger 1926; Heisenberg 1927). The application of QM to chemical systems began after Schrödinger formulated his wave equation in 1926, which describes the wave function of a particle and how it evolves over time. As computational power grew in the mid to late 20th century, QM simulations became feasible for larger molecular systems. The development of more efficient algorithms and computational techniques, such as Hartree-Fock (HF) and density functional theory (DFT), allowed researchers to apply quantum mechanical principles to a wide variety of chemical problems (Ketterson 2016). The introduction of quantum mechanical simulation software, such as Gaussian09, made these powerful methods accessible to the broader scientific community (M. J. Frisch et al. 2016). Today, QM simulations are routinely used to predict molecular and optical properties, reaction mechanisms, and electronic structures with high accuracy.

Central to all of quantum mechanics is the wavefunction (Ψ), a mathematical description of the quantum state of a system. The wavefunction contains all the information about a system's possible states and evolves according to the Schrödinger equation, which describes how this state changes over time. The wavefunction is a complex-valued function that describes the probability amplitude for a particle's position, momentum, or other measurable properties. While the wavefunction itself doesn't correspond to a direct physical observation, its square magnitude ($|\Psi|^2$) represents the probability density of finding a particle in a particular region of space. Additionally, ($|\Psi(x,t)|^2$) gives the probability of finding the particle at position, x, and time, t (Griffiths and Schroeter 2018; Zwiebach 2022). This probabilistic interpretation is a hallmark of QM and distinguishes it from classical physics, where particles have well defined positions and velocities.

An essential concept in QM is the electron density, which describes the spatial distribution of electrons within a molecule. The electron density is directly related to the wavefunction, as the probability of finding an electron in a specific region of space is given by $|\Psi|^2$. In the case of chromophores and molecular systems, electron density plays a

crucial role in determining chemical reactivity and the molecule's interaction with light. The distribution of electron density around a chromophore influences how the molecule absorbs light, interacts with other molecules and undergoes photophysical processes. Calculating the electron density distribution using QM simulations can predict the likelihood of electron transitions within a molecule, which are essential for understanding polarisability and dipole moments, two key concepts for Raman scattering (Tzeliou et al. 2022).

Closely related to the electron density is the concept of molecular orbitals, which describes the spatial distribution of electrons in a molecule. Molecular orbitals arise from the combination of atomic orbitals from each atom in the system and they dictate how electrons are distributed across the molecule. Bonding orbitals, where the electron density is concentrated between atoms, lowers the energy of the system and stabilise the molecule (Schmidt et al. 2014). The conjugation of electrons in chromophores, such as those found in β-barrel FPs or BV, plays a significant role in lowering the energy of electronic transitions, which contributes to their characteristic absorption in the red to near-IR region.

The concept of quantum states, otherwise known as energy levels, is integral to understanding the behaviour of chromophores. In QM, systems exist in discrete energy states and the transition between these states governs the absorption and emission of light (Tallents 2018). As previously mentioned, the ground state represents the lowest energy configuration, while the excited states correspond to higher energy levels that are accessed when the molecule absorbs photons. The energy difference between these quantum states determines the wavelength of light absorbed or emitted by the chromophore. The arrangement of electrons in molecular orbitals and the associated energy levels influence the optical properties of chromophores, including their ability to absorb light and undergo photophysical processes like fluorescence or Raman scattering (Waddar et al. 2024).

Electron density, molecular orbitals, and quantum states collectively offer a detailed understanding of the electronic structure of chromophores. These concepts are essential for predicting and interpreting the behaviour of chromophores under experimental conditions, especially regarding their interactions with light. Through QM simulations, these properties can be accurately modelled, allowing for the design of chromophores with specific, desired characteristics, such as tailoring the absorption spectra or improving the Raman scattering.

1.4.3.1 Parameters of QM simulations

In this thesis, gaussian09 was implemented to conduct the QM simulations as this computational chemistry software package offers methods to study the electronic structures, optical properties and vibrational frequencies of the chromophores. This software provides an effective balance between computational cost and accuracy, making it an essential tool in both fundamental research and applied studies, particularly in modelling chromophores and their interactions.

At the core of QM simulations are the functionals, which are mathematical approximations used to describe the exchange-correlation energy of a system (Tsuneda 2014). The exchange-correlation energy accounts for the complex interactions between electrons, which include the effects of exchange (arising from the Pauli exclusion principle) and correlation (arising from electron-electron repulsion). In density functional theory (DFT), this energy is treated as a function of the electron density, which reduces the computational cost compared to solving the many-body wavefunction directly (Pauli 1925; Bechstedt 2015). One of the most used functionals in QM simulations is B3LYP, a hybrid functional that has become a standard choice in computational chemistry due to its balance between computational efficiency and accuracy. B3LYP combines the Becke three-parameter exchange functional (B3) with the Lee-Yang-Parr correlation functional (LYP) (Lee et al. 1988; Becke 1993; Stephens et al. 1994). The hybrid nature of B3LYP arises from incorporating a fraction of exact exchange energy calculated from the Hartree-Fock method into the exchange-correlation energy, which improves its ability to describe a wide variety of systems. This blend of components allows B3LYP to provide reliable results for molecular geometries, reaction energies and spectroscopic properties, making it particularly well-suited for studying large systems like chromophores and their environments (El-Saady et al. 2023). B3LYP is particularly useful for modelling the electronic structure of chromophores, which are characterised by conjugated systems of delocalised electrons. As previously mentioned, the accurate description of these delocalised systems is essential for predicting the optical properties of chromophores, such as their absorption maxima and vibrational frequencies. When using B3LYP, these properties can be modelled with sufficient accuracy to provide insights into their behaviour under experimental conditions, such as solvents like

water, guiding the interpretation of experimental results and the design of novel chromophores with enhanced properties.

The next consideration for QM simulations is the basis sets, which play a fundamental role in defining the mathematical representation of molecular orbitals. A basis set is a collection of functions, typically Gaussian-type orbitals, used to approximate the wavefunction of a system (Dunning and Hay 1977). These functions describe the spatial distribution of electrons in a molecule, and their quality directly affects the accuracy of the calculations. The choice of basis set is therefore a critical consideration in QM simulations, as it determines the balance between computational cost and the precision of the results. Basis sets can range from minimal to highly complex. Minimal basis sets, such as STO-3G, use the smallest number of functions required to describe each orbital, making them computationally efficient but less accurate for describing molecular properties, particularly for systems with electron delocalisation or polarisation (Hehre et al. 1969). Split-valence basis sets, like 3-21G and 6-31G, improve upon minimal basis sets by using more functions to describe valence orbitals, allowing for greater flexibility in representing electron density (Ditchfield et al. 1971; Binkley et al. 1980). For more accurate calculations, polarisation functions and diffuse functions can be added to a basis set. Polarisation functions (denoted by ‘d’ for heavy atoms and ‘p’ for hydrogen) allow orbitals to adopt more flexible shapes, improving the description of electron density in regions where bonding or lone pairs are distorted, such as in chromophores or during bond stretching (Hariharan and Pople 1973). Diffuse functions (denoted by ‘+’) are crucial for systems with electrons in extended or weakly bound states, such as anions or molecules with significant charge separation (Clark et al. 1983).

In this thesis, the 6-31G+(d,p) basis set is primarily used for the analysis due to its balance of computational efficiency and accuracy. The 6-31G component of this basis set is a split-valence set that provides a good description of core and valence electrons, while the inclusion of polarisation functions (d,p) improves the accuracy of calculations for molecules with complex geometries or electronic distributions. Additionally, the inclusion of diffuse functions (+) is critical for accurately describing electronic states in chromophores, especially with the inclusion of the deprotonated oxygen in the mCherry chromophore. While more advanced basis sets, such as triple-zeta, offer higher accuracy, they also come with significantly increased computational demands (Dyall 2016). For systems of the size and

complexity studied in this thesis, the 6-31G+(d,p) basis set provides the necessary precision without exceeding practical computational limits.

1.4.3.2 Application of QM calculations

The analysis of molecular systems using QM simulations involves a series of computational steps designed to refine the molecular geometry, calculate its energy, and predict its electronic and optical properties. Each step builds upon the previous one, providing increasingly detailed insights into the behaviour of the system under study. The process begins with a geometry optimisation, which is a critical for QM. During optimisation, the positions of the atoms in a molecule are iteratively adjusted to find the geometry that corresponds to the lowest energy state**.** This involves minimising the molecular energy with respect to atomic positions while ensuring that forces acting on the atoms approach an energy minimum. Geometry optimisation is essential for studying chromophores, as their electronic and optical properties are highly sensitive to large changes in their geometry. In this thesis, geometry optimisation was performed in three iterative steps to refine the molecular structure for subsequent analyses. The initial step involved the HF/3-21G method, a computationally inexpensive approach used to provide a rough initial geometry. This method establishes a reasonable starting point for further optimisation by accounting for basic electronic interactions. Following this, the structure was refined using the B3LYP functional with the 6-31G basis set, which improves the accuracy of the geometry by incorporating electron correlation effects into the calculations. Finally, the optimisation process was completed with the B3LYP/6-31G+(d,p) method, which includes polarisation and diffuse functions previously described. This stepwise approach ensures that the final optimised geometry is not only computationally efficient but also highly accurate, making it suitable for subsequent analyses.

Once the geometry is optimised, Density Functional Theory (DFT) calculations were performed on the molecules. This is a powerful computational method that focuses on electron density as the fundamental variable, providing a detailed picture of the electronic structure of molecules by calculating the Kohn-Sham equation (Kohn and Sham 1965). Using the B3LYP functional with the 6-31G+(d,p) basis set, DFT calculations in this thesis were utilised to explore the electronic properties of chromophores and their interactions with

their environments. One key application of DFT is the prediction of dipole moments, which reflect the distribution of charge within a molecule. Dipole moments are critical for understanding how chromophores interact with external electric fields or surrounding molecules, such as proteins or solvents (Gilmore and McKenzie 2006). DFT calculations also provide access to the frontier molecular orbitals, specifically the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) (Choudhary et al. 2019). The energy gap between these orbitals is crucial for understanding the electronic transitions of the chromophore. The gap between the HOMO and LUMO specifically corresponds to the energy of the $S_0 \rightarrow S_1$ transition, influencing the molecule's absorption properties. This information is particularly relevant for chromophores, as their conjugated systems and electronic structures dictate the wavelengths at which they absorb light, a fundamental property for applications in fluorescence microscopy and epr-RS. In this thesis, DFT was also applied to predict Raman spectra by calculating the vibrational modes and their relative intensities. DFT can then provide a theoretical Raman spectrum that can be compared to experimental results. This predictive capability can be a critical tool for interpreting Raman spectroscopic data, particularly in systems where experimental signals may be weak or obscured by background noise. The predicted Raman spectra offer insights into the vibrational properties of the chromophore and its interactions with the surrounding environment, supporting the design of chromophores with enhanced Raman activity.

The final calculation of QM simulations in this work will be the application of Time-Dependant Density Functional Theory (TD-DFT). This is an extension of DFT that allows for the study of excited-state properties of molecules by considering their response to time-dependent perturbations, such as light (Marques and Gross 2004; Herbert 2024). While DFT is primarily used to calculate ground-state properties, TD-DFT builds on this foundation to model electronic transitions between quantum states, enabling the prediction of how a molecule absorbs or emits light. This makes TD-DFT an essential tool for understanding the optical behaviour of chromophores, particularly their absorption spectra. In TD-DFT, the excitation energies of a molecule are calculated by solving the time-dependent Kohn-Sham equations, which describe the interaction of the molecule’s electron density with an oscillating external field, such as electromagnetic radiation (Ullrich 2011). These excitation energies correspond to the energy required to promote an electron from the HOMO to the LUMO or higher energy (LUMO+*n*) orbitals. A key outcome of TD-DFT calculations is the

prediction of the absorption spectrum, which shows the wavelengths of light absorbed by the molecule. For chromophores studied in this thesis, this provided a theoretical framework for understanding their experimental absorption spectra. Linking the calculated absorption wavelengths to experimental data, TD-DFT helps elucidate the relationship between structural features of the chromophore and its optical properties.

QM simulations form a crucial part of this thesis, providing detailed insights into the electronic structure and optical properties of chromophores. The integration of these computational techniques allows for a deeper exploration of how structural features influence chromophore behaviour, bridging the gap between theory and experiment. This integrative approach allows for a deeper interpretation of results, linking theoretical predictions with experimental findings to uncover insights into chromophore properties in complex biological environments. The theoretical foundations of these calculations, including the methodologies and equations, are described in detail from first principles in Methods section 2.7.

1.5 Aims and objectives

This thesis aims to advance the development and application of genetically encoded Raman-active probes by utilising a combination of experimental and computational approaches. The research is structured across four chapters, each addressing distinct but interconnected objectives to explore far-red and near-IR FPs, ncAA incorporation and Raman spectroscopy. These aims align with the overarching goal of understanding and enhancing the spectral and structural properties of chromophores for advanced imaging applications.

The initial focus, outlined in Chapter 3, is the characterisation and spectroscopic analysis of selected far-red and near-IR FPs to identify the most promising candidates for genetically encoded Raman-active probes. This chapter establishes a foundation for subsequent work by analysing the structural and spectral properties of these FPs, particularly their compatibility with Raman spectroscopy. Following this, the structural determination of mRhubarb720 via micro-focused crystallography will be performed, enabling the identification of key residues within the BV-binding pocket. This structural insight will inform future engineering strategies and aid in understanding how the protein environment influences the chromophore's behaviour.

Chapter 4 builds upon this groundwork by investigating the Raman enhancement properties of near-IR FPs. The objectives include evaluating the impact of Raman enhancement on chromophores and examining how the microscope and laser setup affect the Raman properties of these proteins. A comparative analysis of the Raman spectra from near-IR FPs, classical β-barrel FPs such as mCherry, and free-BV will help identify characteristic Raman peaks and their underlying electronic properties. The results from this chapter aim to refine the selection of FPs for *in situ* imaging in *E. coli* and mammalian cells, assessing their viability as genetically encoded Raman-active probes for real-time imaging.

Chapter 5 addresses a more targeted question of whether the incorporation of pCNPhe can induce through-space enhancement of Raman activity within the biologically silent region, when coupled to BV in mRhubarb720. Structural and dynamic consequences of pCNPhe incorporation will be analysed using MD simulations, focusing on how this modification impacts BV's structural behaviour and its interactions within the protein environment. In addition to MD, QM simulations will explore the electronic properties of BV in its protein-bound and free states, predicting optical and vibrational properties to complement experimental findings. These simulations will also identify alternative vibrational modes beyond the 1550–1850 cm^{-1} range, laying the groundwork for future epr-RS imaging applications using mRhubarb720.

Finally, Chapter 6 evaluates the impact of ncAA incorporation into the chromophore of mCherry, focusing on how these modifications alter spectral properties, such as absorbance and Raman activity. Computational methods, including MD simulations, will assess the structural stability of these chromophore variants, while QM simulations will predict their optical properties. This integrated approach seeks to elucidate how ncAA modifications influence the chromophore's functionality and to guide the design of novel variants with enhanced spectral features.

2 Materials, Methods, and Theory

2.1 Laboratory materials

2.1.1 Materials: chemicals, buffers, and antibiotics

The chemicals utilised for the described methodologies are shown below in Table 2.1. Composition of solutions used, along with the protocols for their preparation, are also outlined below. Stock solutions were produced by dissolving chemicals in deionised water (dH_2O), unless stated otherwise.

Table 2.1: Commonly used laboratory chemicals

Chemical	Specification/Preparation	Product Source
Acetic Acid	≥ 99.5% purity	Scientific Laboratory Supplies (SLS)
Acrylamide: Bis-Acrylamide	40% (w/v) Acrylamide: Bis-Acrylamide 29:1	Geneflow
Agarose	Molecular Biology Grade	Appleton Woods
Ammonium persulphate (APS)	Prepared as 10% (w/v) stock	Melford
L-Arabinose	≥ 98% purity	Sigma-Aldrich
β-mercaptoethanol	≥ 99% purity	Sigma-Aldrich
Bromophenol blue	Technical grade	Sigma-Aldrich
Bugbuster™	Technical grade	Novagen
Coomassie blue	Coomassie Brilliant Blue R250	Thermo Fisher Scientific
Deoxyadenosine triphosphate (dATP)	Prepared as a 10 mM stock	Promega
Deoxycytidine triphosphate (dCTP)	Prepared as a 10 mM stock	Promega
Deoxyguanosine triphosphate (dGTP)	Prepared as a 10 mM stock	Promega
Deoxythymidine triphosphate (dTTP)	Prepared as a 10 mM stock	Promega
DTT	Technical grade	Thermo Fisher Scientific
Ethanol	≥ 99.8% purity	VWR

Ethidium Bromide	Prepared as 10 mg/mL stock; 95% purity	Sigma-Aldrich
D-Glucose	Anhydrous	Melford
Glycerol	Analytical grade	Thermo Fisher Scientific
Glycine	Glycine free base	Melford
Imidazole	≥ 99% purity	Sigma-Aldrich
Isopropanol	≥ 99.5% purity	Thermo Fisher Scientific
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	Analytical grade	Roche
Lactose	≥ 98% purity	Sigma-Aldrich
Methanol	≥ 99.8% purity	Thermo Fisher Scientific
Sodium Chloride	≥ 99.5% purity	Thermo Fisher Scientific
Sodium dodecyl sulphate (SDS)	Prepared as 10% (w/v) stock	Melford
Sodium hydroxide	Analytical grade	Thermo Fisher Scientific
Sodium phosphate monobasic	Monohydrate; ≥ 98% purity	Melford
Sodium phosphate dibasic	Anhydrous; ≥ 98% purity	Melford
Tetramethylethylene-diamene (TEMED)	≥ 99% purity	Melford
Tris(hydroxymethyl) aminomethane hydrochloride (Tris-HCL)	Molecular Biology Grade Ultra Pure	Melford
Tris(hydroxymethyl) aminomethane base (Tris-base)	Molecular Biology Grade Ultra Pure	Melford

Various solutions and buffers were also prepared for the methodologies described in Table 2.2. The table outlines the composition, pH and sterilisation procedures required for each mixture.

Table 2.2: Composition of buffers and solutions commonly used in the laboratory.

Buffer / Solution	Composition
25x 18-amino acid mix (No Tyr/Cys) (5g each, Filter sterilisation)	Glutamic acid (Na salt), Aspartic Acid, Lysine-HCl, Arginine-HCl, Histidine-HCl-H_2O, Alanine, Proline, Glycine, Threonine, Serine, Glutamine, Asparagine-H_2O, Valine, Leucine, Isoleucine, Phenylalanine, Tryptophan, Methionine
Coomassie destain buffer	40% (v/v) methanol 10% (v/v) acetic acid
Coomassie stain buffer	40% (v/v) methanol 10% (v/v) acetic acid 0.1% (w/v) R250 Coomassie blue
CutSmart buffer 10X (New England Biolabs) (pH 7.9)	50 mM potassium acetate 20 mM Tris-acetate 10 mM magnesium acetate 100 µg/mL recombinant albumin
Gel Loading Dye no SDS 6x (New England Biolabs) (pH 8.0)	60 mM EDTA 19.8 mM Tris-HCl 15% Ficoll®-400 0.12% Dye 1 0.006% Dye 2
25X M-salts	125 mM sodium sulphate 625 mM sodium phosphate dibasic 625 mM potassium dihydrogen phosphate 1.25 M ammonium chloride
N3 buffer (Qiagen) (pH 4.8)	4.2 M guanidium hydrochloride 0.9 M potassium acetate
Ni affinity elution buffer (pH 8.0)	500 mM Imidazole 50 mM Tris-HCl
Ni affinity wash buffer (pH 8.0)	10 mM Imidazole 50 mM Tris-HCl
P1 buffer (Qiagen) (pH 8.0)	50 mM Tris-HCl 10 mM EDTA 100 µg/mL RNase A 0.1% (v/v) LyseBlue

P2 buffer (Qiagen)	200 mM sodium hydroxide 1% (w/v) SDS
PB buffer (Qiagen)	5 M guanidium hydrochloride 30% (v/v) isopropanol
PE buffer (Qiagen) (pH 7.5)	100 mM sodium chloride 10 mM Tris-HCl 80% (v/v) ethanol
Quick Ligase reaction buffer 2X (New England Biolabs) (pH 7.6)	66 mM Tris-HCl 10 mM magnesium chloride 1 mM dithiothreitol 1 mM adenosine triphosphate 7.5% (v/v) polyethylene glycol; PEG600
SDS-PAGE loading buffer	0.2% Tris-HCl pH 6.8 10% (v/v) β-mercaptoethanol 8% (w/v) glycerol 2% (w/v) SDS 0.04% (w/v) bromophenol blue
Sodium phosphate buffer (pH 7.0)	50 mM sodium phosphate dibasic/sodium phosphate monobasic
SDS-PAGE running buffer (pH 8.3)	192 mM glycine 25 mM Tris base 0.1% (w/v) SDS
Tris-acetate-EDTA (TAE) buffer (pH 8.8)	40 mM Tris-acetate 1 mM EDTA
Trace metals 5000X (Filter sterilisation)	50 mM iron (III) chloride 200 µM calcium chloride 100 µM manganese chloride 100 µM zinc sulphate 20 µM cobalt chloride 20 µM copper chloride 20 µM nickel chloride 20 µM sodium selenite 20 µM sodium molybdate 20 µM boric acid
Tris buffer (pH 8.0)	50 mM Tris-HCl/Tris-base

Antibiotics were prepared from 1000x stock solutions using ultra-pure MilliQ water and filtered via 0.22 μM syringe filters (ThermoFisher). Final working concentration of each antibiotic used were: 50 μg/mL Ampicillin (Melford), 50 μg/mL Chloramphenicol (Melford), 50 μg/mL Spectinomycin (Melford), 50 μg/mL Kanamycin (Melford) and 12.5 μg/mL Tetracycline (Melford).

2.1.2 Bacterial strains and growth media

Transformation and expression were performed using two *E.coli* strains: Top10^{TM} (Invitrogen, ThermoFisher) and BL21 (DE3) (NEB). The genotypes for both strains are listed in Table 2.3.

Table 2.3: *E.coli* cell strains and their respective genotypes and permitted expression vectors

Strain	Genotype	Expression Type	Proteins expressed
Top 10^{TM}	F^{-}*mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K λ^{-}*rps*L(Str^{R}) *end*A1 *nup*G	Arabinose-inducible promoter (pBAD)	mCherry
BL21 (DE3)	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*	T7 promoter (pET)	mRhubarb720 emiRFP670 miRFP670nano3

Bacterial growth media, including preprepared 2x Yeast Extract Tryptone (2xYT) medium (Melford), LB Broth (Melford) and LB Agar (Melford) was solubilised using dH_2O, as per the manufacturer's instructions, and sterilised by autoclaving at 121°C for 15 minutes. LB Agar was cooled to 50°C before selective antibiotic was supplemented, poured into plates, and sealed at 4°C.

Autoinduction media ZYM-5052 was used for pCNPhe and pCCPhe incorporated variants (Studier 2005). This included the following: 1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (v/v) glycerol, 0.05% (w/v) glucose, 0.2% lactose, 25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM $NaSO_4$, 2 mM $MgSO_4$ and 1x Trace Metal (4 μM $CaCl_2$, 2 μM MnCl2, 2 μM $ZnSO_4$, 0.4 μM $CoCl_2$, 0.4 μM $CuCl_2$, 0.4 μM $NiCl_2$, 0.4uM Na_2MoO_4, 0.4 μM H_3BO_3 and 10 μM $FeCl_3$ dissolved in ultra-pure MillQ water). Powdered tryptone and yeast extract was dissolved dH_2O and autoclaved

Minimal autoinduction media was attempted for the 3-NO_2-Tyr variants (Beyer et al. 2020). This used as similar composition: 0.25% (w/v) aspartate pH 7.5, 0.5% (v/v) glycerol, 25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM $NaSO_4$, 2 mM $MgSO_4$, 2 mM $MgSO_4$, 0.05% (w/v) glucose, 0.05% (w/v) arabinose (for mCherry) or 0.5 mM IPTG (for mRhubarb720), 0.2% (w/v) lactose, 1x Trace Metal (4 µM $CaCl_2$, 2 µM MnCl2, 2 µM $ZnSO_4$, 0.4 µM $CoCl_2$, 0.4 µM $CuCl_2$, 0.4 µM $NiCl_2$, 0.4uM Na_2MoO_4, 0.4 µM H_3BO_3 and 10 µM $FeCl_3$ dissolved in ultra-pure MillQ water), and 0.2 g/L of each of the following amino acid: Glutamic acid-Na salt, Aspartic acid, Lysine-HCl, Arginine-HCl, Histidine-HCl-H_2O, Alanine, Proline, Glycine, Threonine, Serine, Glutamine, Asparine-H_2O, Valine, Leucine, Isoleucine, Phenylalanine, Tryptophan and Methionine (note: no Cys/Tyr). All dissolved in autoclaved dH_2O.

2.2 Molecular Biology

2.2.1 Expression and ncAA incorporation plasmids

Various genetic constructs were used to express fluorescent proteins (FPs) and to facilitate the incorporation of non-canonical amino acids (ncAAs). WT mRhubarb720 and mCherry plasmids were acquired from Addgene whereas emiRFP670 and miRFP670nano3 were designed and produced via TwistBiosciences. The pET-Duet-mRhubarb720-HO1, pET28a(+)-Duet-emiRFP670-HO1 and pET28a(+)-Duet-miRFP670nano3-HO1 genes were expressed using an Isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible pET vector, while mCherry was expressed via the arabinose inducible pBAD vector (Figure 2.1). The near-IR FPs required a secondary protein, Heme Oxygenase 1 (HO-1), to be expressed to produce biliverdin through the cleavage of the α-methene bridge in Heme, within bacterial cells. The DNA and protein sequences for each construct are listed in the Supplementary Tables S1-2.

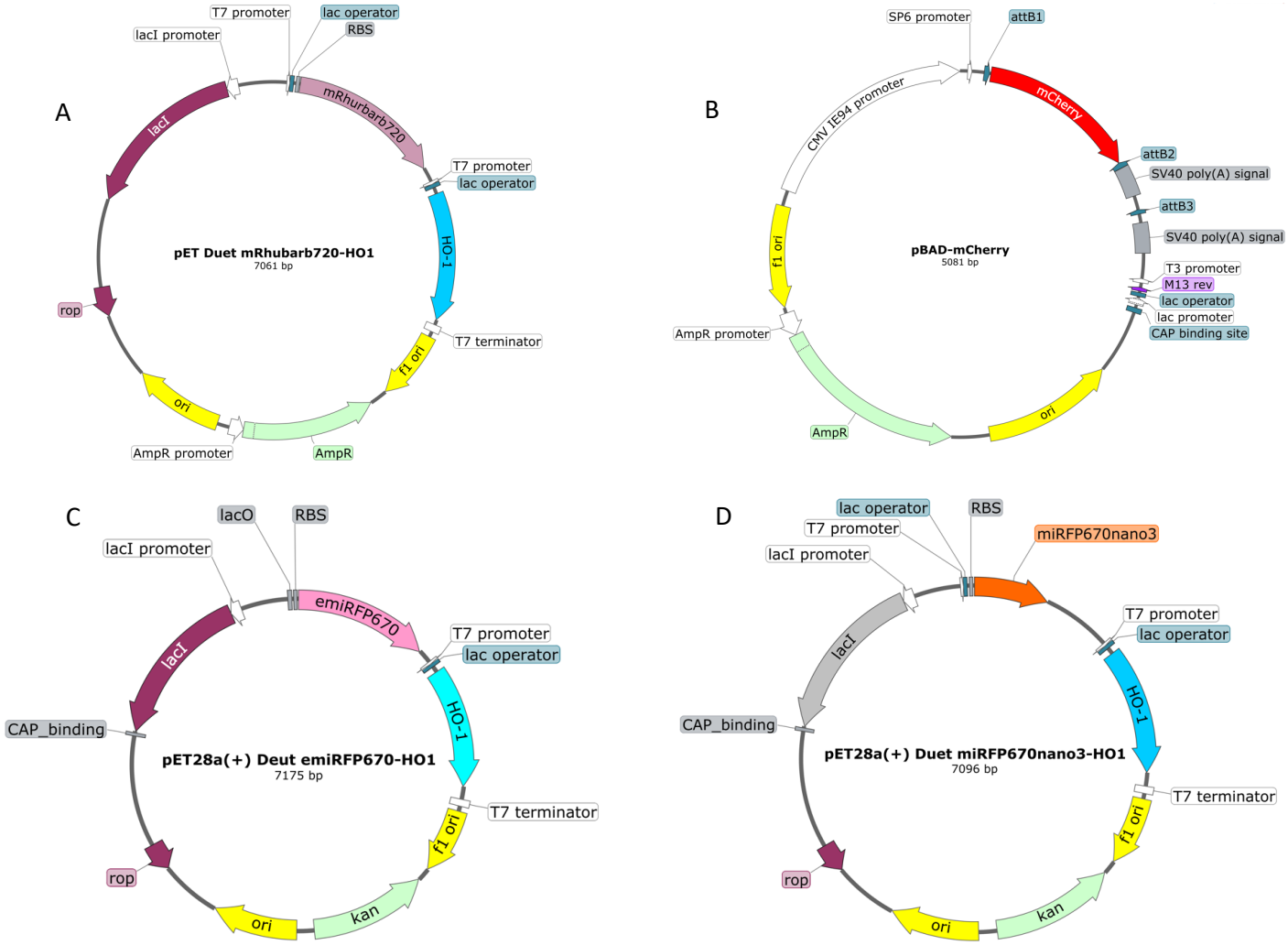


Figure 2.1: Graphical maps of the genetic constructs used to express FPs. (A) mRhubarb720 [Addgene: #141201]. (B) mCherryWT [Addgene: #176016]. (C) emiRFP670 [TwistBiosciences]. (D) miRFP670nano3 [TwistBiosciences]. Maps produced via SnapGene

The ncAA incorporation vectors were all derivatives of the pDULE system and acquired from Addgene. These plasmids contain an engineered orthogonal tRNA synthetases to recognise the amber stop codon (TAG). The pDULE2 was co-transformed to integrate 4-cyano-phenylalanine (pCNPhe) as well as 4-ethynyl-phenylalanine (pCCPhe) and both pDULE1 and pDULE3 was attempted for 3-Nitro-tyrosine (3-NO_2) incorporation (Figure 2.2). These orthogonal tRNA/aaRS pairs were derived from the *Methanococcus jannaschii* organism. Each plasmid constitutively expressed the tRNA/aaRS pair, pDULE1 and pDULE3 employed the Tetracycline resistance gene whereas pDULE2 worked with Spectinomycin

resistance. Each ncAA was solubilised in 1 M NaOH to a concentration of 1 mM for each culture.

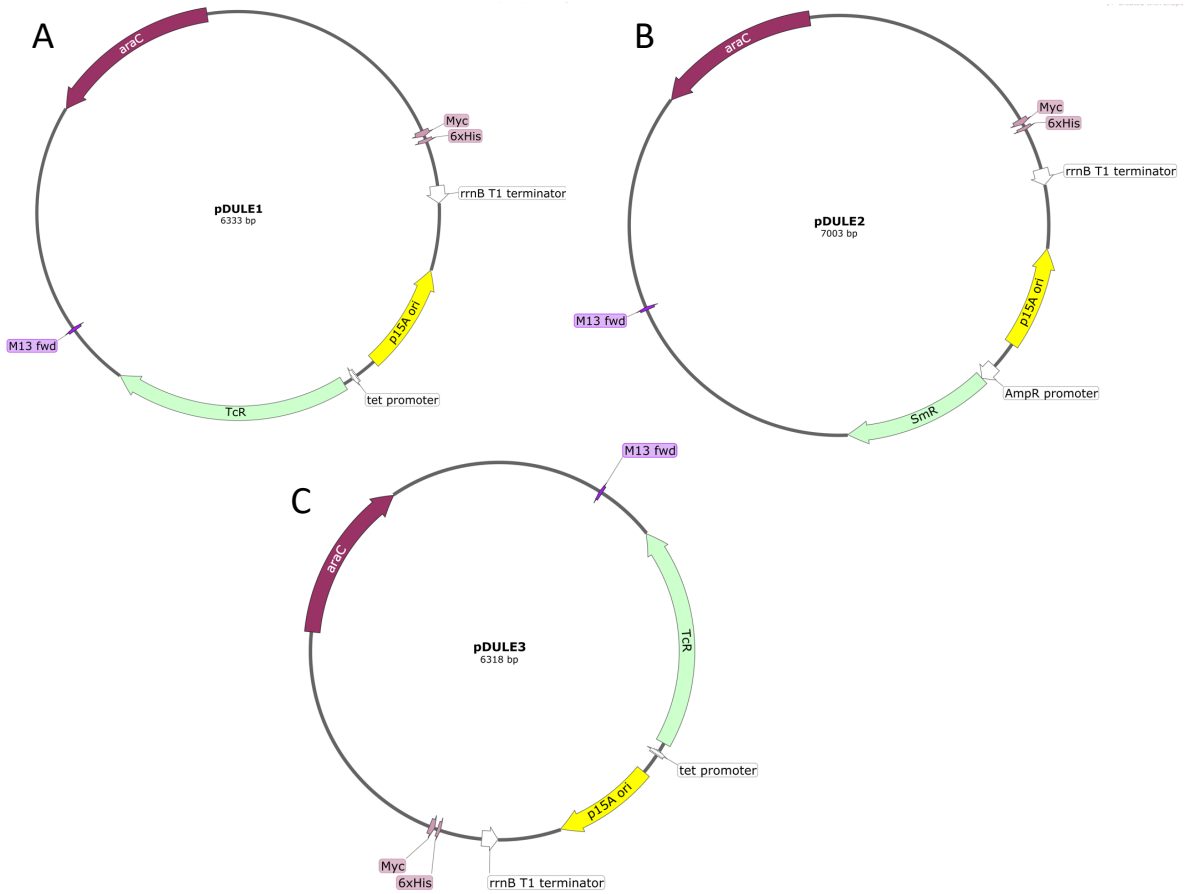


Figure 2.2: Graphical maps of the genetic constructs used to incorporate ncAAs. (A) pDULE1 [Addgene: #85494]. (B) pDULE2 [Addgene: #85495]. (C) pDULE3 [Addgene: #174078]. Maps produced via SnapGene

2.2.2 DNA Oligonucleotides

Mutants containing the amber stop codon for ncAA incorporation were generated using the mutagenesis primers below and were synthesised by Integrated DNA Technologies. Various tools such as Amplifx, NEBaseChanger and SnapGene were used to design the oligonucleotide primers (Table 2.2). Annealing temperatures (Tm) were determined using NEBaseChanger (https://nebasechanger.neb.com). Primers used to sequence gene constructs (T7_SeqP and pBAD_SeqP) are also shown in Table 2.4 below.

Table 2.4: Primer sequences and Tm's used in this thesis. Mutations have been highlighted in bold. T7 SeqP pair was donated by Rebecca Gwyther, and pBAD SeqP pair was designed by Samuel Reddington.

Primer	**Sequence (5' → 3')**	**Tm (°C)**
F_mRhubarb_D177TAG	CGAC**TAG**AGCGGCGAAG	66
R_mRhubarb_D177TAG	GAGGCGAAGCGATAGATCATC	
F_mRhubarb_Y198TAG	CTAGGCCAGCAC**TAG**CCTG	66
R_mRhubarb_Y198TAG	TTTTGACTCGACCTCGGC	
F_mRhubarb_Y211TAG	CGTCGGCTC**TAG**ACCATCAACC	67
R_mRhubarb_Y211TAG	GGCCTGCGCCGGAATATCT	
F_mRhubarb_N224TAG	CAAT**TAG**CGGCCGGTGC	67
R_mRhubarb_N224TAG	ATATCGGGAATGATCCGTACCG	
F_mRhubarb_F258TAG	CATCTGGAA**TAG**ATGCGCAAC	65
R_mRhubarb_F258TAG	GACGGGCGAGACGCT	
F_mRhubarb_S269TAG	GCACGGCACGATG**TAG**ATCTC	67
R_mRhubarb_S269TAG	ATGCCTATGTTGCGCATGAATTC	
F_mRhubarb_H285TAG	GATCGTTTGC**TAG**CACCGAAC	64
R_mRhubarb_H285TAG	AATCCCCACAGTCGCTCG	
F_mCherry_Y67TAG	CCTCAGTTCATG**TAG**GGCTCCAA	68
R_mCherry_Y67TAG	GGACAGGATGTCCCAGGCG	
F_T7_SeqP	TAATACGACTCACTATAGGG	Sequencing primers
R_T7_SeqP	CTAGTTATTGCTCAGCGGT	
F_pBAD_SeqP	ATGCCATAGCATTTTTATCC	
R_pBAD_SeqP	GATTTAATCTGTATCAGG	

2.2.3 Site Directed Mutagenesis

Whole plasmid polymerase chain reaction (PCR) was performed to amplify genes and to introduce site specific mutations using the primers in Table 2.2. The Q5® Site-Directed Mutagenesis Kit (NEB: E0554) was purchased to implement the mutations. The reaction mixture and PCR conditions are listed in Table 2.5. All reactions were set up in PCR tubes (Starlabs) and performed using a gradient PCR Thermo Cycler (MedGroup: NEUTC100).

Table 2.5: PCR composition and conditions for site-directed mutagenesis. Tm temperatures are in Table 2.4.

Materials		**Volume (in 50 μL reaction)**
Q5 Hot Start High-Fidelity 2X Master Mix		25 μL
Forward primer (10 μM)		2.5 μL
Reverse primer (10 μM)		2.5 μL
Template DNA (1 ng/μL)		1 μL
Nuclease-free water		19 μL
PCR conditions		
Step	**Temperature (°C)**	**Time**
Initial denaturation	98	30 seconds
30 cycles	98	10 seconds
	50-72 (Tm*)	30 seconds
	72	30 seconds/kb
Final extension	72	2 minutes
Hold	4	

The QIAquick® PCR purification kit (QIAGEN) was used to remove residual impurities. For the 50 μL reaction volume, 250 μL of Buffer PB was added to the sample and then applied to the QIAquick spin column. Centrifugation at 13000 rpm in a benchtop microfuge for 60 seconds followed, where the DNA binds to the silica gel in the spin column and impurities washed away in the flowthrough. PE buffer (750 μL) was added to the column and centrifuged for 60 seconds at 13000 rpm twice, discarding the flowthrough each time to wash the column. The spin column was placed into a new sterile 1.5 mL tube and 30 μL of PCR water added to the spin column and left to stand for 5 minutes. A final centrifugation step (13000 rpm for 60 seconds) was performed to elute the DNA.

The purified PCR product was then subject to phosphorylation and ligation steps. Purified PCR product (12 μL) was added to 12 μL of Quick Ligase buffer (NEB), 1 μL of 10 U/μL T4 polynucleotide kinase (NEB) and incubated at 37°C for 30 minutes. Ligation then followed by adding an additional 1 μL Quick Ligase (NEB) and incubating for 8 minutes at room temperature. To stop the reaction, another QIAquick® PCR purification step was

initiated using the above steps. DNA quantification then followed using the NanoDrop ND 1000 Spectrophotometer (Thermo Fisher) by measuring the absorbance (A_{280}) on the in-built nucleic acids setting with 1.5 μL of final ligated product.

2.2.4 DNA purification

To isolate plasmid DNA from bacterial cell cultures, the QIAprep Spin Miniprep Kit (QIAGEN) was used. Bacterial cell cultures (5-10 mL) were grown overnight at 37°C, then centrifuged at 8000 rpm for 20 minutes. Cell pellets were resuspended in 250 μL Buffer P1 and transferred to a 1.5 mL centrifuge tube. Buffer P2 (250 μL) was added and mixed thoroughly by inverting the tube until the solution became blue (via use of LyseBlue reagent). Buffer N3 (350 μL) was added and mixed thoroughly by inverting the tube until the solution became colourless. The microcentrifuge tube was then centrifuged for 10 minutes at 13000 rpm. The supernatant was added to a QIAprep spin column and centrifuged for 60 seconds at 13000 rpm. The spin column was washed with 750 μL Buffer PE and centrifuged twice for 60 seconds at 13000 rpm with flowthrough discarded after each cycle. The spin column was transferred to a clean 1.5 mL microcentrifuge tube and 50 μL of PCR water was added to the centre of the column and allowed to stand for up to 10 minutes. The tubes were finally centrifuged for 60 seconds at 13000 rpm to elute the pure plasmids. DNA quantification then followed using the NanoDrop ND 1000 Spectrophotometer (Thermo Fisher).

2.2.5 Agarose gel-electrophoresis

Agarose gels (1% w/v) were made by dissolving 0.5 g agarose (Melford) in 50 mL 1x Tris-acetate-EDTA (TAE) buffer with the addition of 0.25 mg/mL ethidium bromide (Sigma Aldrich). Running buffer for the gel-electrophoresis was also TAE buffer. DNA samples (5 μL) were mixed with 1 μL of 6x DNA loading dye (bromophenol blue and xylene cyanol FF; New England Biolabs). DNA samples were loaded into the wells in the gel; a 1 Kb plus DNA ladder (NEB) was also added to estimate the molecular weight of linear, double stranded DNA fragments. Gel-electrophoresis was performed by applying 120 V for 30-45 minutes, with DNA products imaged by illuminating gels under UV transilluminator (GelDoc-It; UV products Ltd.).

2.2.6 Bacterial transformation and protein expression

Near-IR constructs were transformed into chemically competent *E.coli* BL21(DE3). mCherry construct was transformed into chemically competent *E.coli* TOP10™. A heat shock transformation protocol was performed, whereby 1 ng of purified plasmid DNA was added to 50 μL of competent cells and incubated on ice for 30 minutes. The competent cell/DNA mixture was then incubated in a 42°C water bath for 45 seconds and then subsequently placed on ice for an extra 2 minutes. Super Optimal Broth (SOB) media (450 μL; without antibiotic) was added and the cells allowed to recover for 60 minutes at 37°C in a shaking incubator. Each transformation were plated on LB agar with appropriate antibiotics and incubated at 37°C overnight. Selected colonies were then picked and incubated in 5 mL 2xYT cultures overnight at 37°C. The cultures were then used to inoculate 0.5 L 2xYT medium supplemented with the appropriate antibiotics.

To recombinantly produce near-IR FPs, cultures were incubated at 37°C until an OD_{600} of 0.4-0.8 was reached, upon which 0.5 mM IPTG was added. The cultures were then incubated for a further 24 hours at 22°C with shaking. For mCherry production, cultures were incubated at 37°C until an OD_{600} of 0.4-0.8 was reached, then 0.05% (w/v) arabinose was added, with cultures incubated for 18 hours at 25°C with shaking.

To incorporate ncAA into recombinant FPs, the plasmids containing the FPs (mCherry-pBAD [Addgene: #54630], pET-DUET-1-mRhubarb720-HO [Addgene: #141201]), were co-transformed with their respective pDULE plasmid (pDULE1 [Addgene: #85494], pDULE2 [Addgene: #85495], pDULE3 [Addgene: #174078]) and transformed using the same heat shock protocol outlined above into their corresponding *E.coli* cell type listed in Table 2.1. Double transformed constructs were then plated on LB agar plates with the appropriate FP selected antibiotic and a secondary antibiotic dependant on the pDULE used (pDULE1 and pDULE3 - 12.5 μg/mL tetracycline; pDULE2 – 50 μg/mL spectinomycin). Plates were then incubated at 37°C overnight. For pCNPhe and pCCPhe incorporation, colonies picked for a second overnight incubation at 37°C in 5 mL of 2xYT media containing 50 μg/mL spectinomycin and 50 μg/mL ampicillin. These 5 mL cultures were used to inoculate 0.5 L autoinduction media containing the same concentration of antibiotics (please see Section 2.1.2). For the 3-NO_2-Tyr incorporation, colonies were picked for overnight incubation at 37°C in 5 mL of non-inducing minimal media (without arabinose) containing 12.5 μg/mL

tetracycline and 50 μg/mL ampicillin. These overnights were used to inoculate 0.5 L inducing minimal media (Section 2.1.2). These cultures were then incubated at 37°C for 1 hour, the prepared ncAA was then added to the culture (See Methods 2.2.1). The cultures were incubated further at 37°C until an OD_{600} of 0.4-0.8 was reached. Then mRhubarb720 cell expression was incubated for 24 hours at 22°C with shaking and mCherry cell expression was incubated for 18 hours at 25°C with shaking.

2.2.7 Protein purification

Cells producing the required recombinant protein were harvested by centrifugation at 5000 rpm at 4°C for 20 minutes using a Thermo Fisher Fiberlite F9-6 x 1000 LEX Fixed Angle Rotor and resuspended in 50 mM Tris buffer (pH 8.0) and lysed using a French press; cell lysate was collected and then centrifuged at 25000 rpm for 1 hour at 4°C using a Thermo Fisher Fiberlite F21-8 x 50y Fixed Angle Rotor, where supernatant was collected. The FPs were purified using a His-trap HP nickel affinity column (Cytiva) linked to an ATKA Purifier FPLC. The His-trap column was equilibrated and washed with 50 mM Tris (pH 8.0), 300 mM NaCl and 10 mM Imidazole and eluted with 50 mM Tris (pH 8.0), 300 mM NaCl with 500 mM Imidazole. Pooled FPs samples were then subject size exclusion chromatography (SEC) using a HiLoad 26/600 Superdex 200 column, where the column was equilibrated with 50 mM Tris (pH 8.0). Specific monitoring absorbance wavelengths were used to detect protein elution, with the wavelength used dependent on the λ_{max} of the FP. Protein purity was assessed using SDS-PAGE analysis (Section 2.3.1).

2.3 Protein analysis

2.3.1 Sodium dodecyl sulphate polyacrylamide gel-electrophoresis (SDS-PAGE)

SDS-PAGE gels were produced in-house consisting of two components: resolving gel and stacking gel. The resolving gel comprised of 0.375 mM Tris buffer (pH 8.8), 31.3% bis-acrylamide solution, 0.1% w/v SDS, 0.05% w/v APS and 0.02% v/v TEMED. The stacking gel comprised of 65 mM Tris buffer (pH 6.8), 13% bis-acrylamide solution, 0.2% w/v SDS, 0.1% w/v APS and 0.02% v/v TEMED. The resolving gel was loaded and solidified into the casket

first and then the stacking gel loaded and solidified on top of the resolving gel. The gel caskets were loaded into a Mini-PROTEAN Tetra cell tank. Samples were subject to 5x SDS-PAGE reducing loading buffer (10% (w/v) SDS, 25% (v/v) ß-mercaptoethonal, 0.05% (w/v) bromophenol blue, 30% (v/v) glycerol, and 250 mM Tris-HCl pH 6.8) followed a denaturation step at 98°C for 10 minutes. Each sample was loaded into a separate well within the casket, alongside a pre-stained protein molecular weight ladder (New England Biolabs) loaded into its own designated well. Gel electrophoresis was performed at 175 V for 45-60 minutes. Gels were stained with a Coomassie stain solution (0.1% (w/w) Coomassie Blue R250, 40% (v/v) methanol and 10% (v/v) acetic acid) and then destained in the corresponding destain solution (40% (v/v) methanol and 10% (v/v) acetic acid). Imaging of gels was carried out using the GelDoc-It system (UV products Ltd.).

2.3.2 Calculation of protein concentrations and extinction coefficients

The concentration of purified proteins was initially determined using the DC Protein Assay (BioRAD) using the manufactures guidelines. Either BSA or mRhubarb720 (0.2-1.5 mg/ml) was used as the concentration standard. The extinction coefficient (ε) of each protein at 280 nm and its absorbance maximum (λ_{max}) were determined using the Beer-Lambert law ($A = \varepsilon cl$). Once ε was calculated, protein concentrations could be directly determined from absorbance measurements at the appropriate wavelengths, as absorbance (in a.u.) is proportional to protein concentration.

2.3.3 Spectral analysis

Absorbances was determined using a Cary 60 UV-vis spectrophotometer (Agilent Technologies). These measurements required blanks for baselining spectra, usually 50 mM Tris (pH 8.0) but varied depending on the buffers used. Absorbance was measured between 200-800 nm, with a path length of 1 cm. Fluorescence emission was determined using Varian Cary Eclipse Fluorescence spectrophotometer (Varian). Spectra were measured using the absorbance maxima value as the excitation wavelength. Fluorescence excitation spectra were measured using the emission maxima value as the detection emission wavelength. The wavelengths measured during fluorescence analysis were dependent on the specific protein being observed. For absorbance measurements, 10 μM protein samples were used and 5

μM samples for fluorescence measurements. Quartz cuvettes were used in all spectral analysis.

2.3.4 Quantum yield (QY) determination

QY calculations were performed using a dye with a known QY as a standard that excited and emitted in a range similar to that of the unknown samples. For mCherry Rhodamine 101 was used (QY = 1.00, λmax = 590 nm). However, no appropriate standard could be found for mRhubarb720, therefore the WT and its already reported value was used (QY = 0.064). The paper that generated mRhubarb720 used another near-IR FP (iRFP713) as a reference to calculate its QY (Rogers et al. 2019). Purified FP samples (and Rhodamine 101) were diluted to a series of concentrations at absorbances of 0.1, 0.08, 0.06, 0.04, 0.02 and 0 a.u. (blank). Emission spectra were measured for both the test sample and the standard. The integrated fluorescence intensities were plotted against the corresponding absorbance values for each, producing straight-line graphs (intersecting the orgin (0,0)) for the test sample and the standard. The gradients of these graphs, representing the relationship between fluorescence intensity and absorbance, were then used in Equation 1.

Equation 1

$$QY_x = QY_{ST}\left(\frac{Gradient_x}{Gradient_{ST}}\right)\left[\frac{n_x}{n_{ST}}\right]$$

Where the subscript x denotes the test sample and subscript ST represents the standard sample. The refractive index is represented by the squared bracket and is negligible and therefore, approximated as 1.

2.3.5 Mass spectrometry

Purified samples were diluted to 10 μM and analysed at the Mass Spectrometry facility in the School of Chemistry, Cardiff University. The samples were subject to liquid chromatography-mass spectrometry (LC-MS) where 1 μL of sample is injected into a Waters Acquity H-Class UPLC system coupled to a Waters Synapt G2-Si quadrupole time of flight mass spectrometer. The column used a Waters Acquity UPLC Protein C4 BEH column 300 Å,

1.7 μm (2.1 x 100 mm) and held at 60°C. Flow rate was set to 0.2 mL/min and mass spectrometry data was collected in positive electrospray ionisation mode and the data was analysed using Waters MassLynx 4.1 with the deconvoluted mass spectra generated using the MaxEnt 1 software. Mass prediction was carried out using the ExPASy ProtParam tool. The covalent addition of biliverdin (BV) was taken into account by adding 583 Da (BV's molecular mass) and subtracting one hydrogen (1 Da; lost during covalent attachment) for WT. For the F258pCNPhe mutant, the predicted mass was calculated by substituting pCNPhe with phenylalanine, adding the mass of the nitrile group (26 Da) and the BV molecule (583 Da), and then subtracting the mass of two hydrogens (2 Da) to account for the BV covalent linkage and the replacement of a hydrogen atom by the nitrile group.

2.4 Protein crystallography

2.4.1 Crystal formation

Purified mRhubarb720 WT samples were concentrated to 10 mg/mL in 50 mM Tris (pH 8.0) buffer. Crystallisation was implemented using various conditions that comprise the PACT *premier*™ HT-96 / FX-96 crystallisation screen (Molecular Dimensions). Equal volumes protein (0.2 μL) and crystallisation buffer (0.2 μL) were placed into each 96-Well Triple Drop Crystallisation Plates (Molecular Dimensions). The solutions were dispensed using the Mosquito® LV Crystal robot (sptlabtech) in the Medicines Discovery Institute at Cardiff University with the help of Dr Magdalena Lipka-Lloyd. The crystallisation plates were incubated at 25°C and monitored weekly for crystal growth. After 3 weeks, only microcrystals formed with the most promising condition being Well D2 containing 0.1 M DL-Malic acid, MES monohydrate (MMT), 25% w/v polyethylene glycol (PEG) 1500 at pH 5.0. The tray was transported at 25°C to Diamond Light Source (DLS), Harwell, UK for Micro focussing X-ray diffraction experiments.

2.4.2 Micro-focussing and refinement

Data was collected by Dr Anna Warren at DLS, where D2 of the tray was used. Well/buffer solution (1.5 μL) was pipetted on into a crystal drop and aspirated to isolate microcrystals. Fresh 0.1 M MMT, 25% w/v PEG 1500 (pH 5.0) buffer (1.5 μL) was added to the crystals and blotted for 6 seconds on to a cryo-TEM grid (Cu 200 Quantifoil 2/2 mesh)

before plunged into liquid ethane (Leice EM GP2 plunge freezer). The cryo-TEM grid was then mounted onto the VMXm beamline for the diffraction experiments. Initial refinement was undertaken at DLS using Alphafold2 with subsequent refinement procedures using REFMAC5 via CCP4 suite (Murshudov et al. 2011).

2.5 Cell imaging

2.5.1 Mammalian cell constructs

Three mammalian cell genetic constructs were designed to fuse mRhubarb720 WT to cellular proteins for production in HeLa cells. The fusion partners were Histone H2B (H2B), Lysosomal-associated membrane protein 1 (LAMP1) and pLifeAct (F-actin binding peptide) (DNA sequences can be found in Table S3, protein sequences can be found in Table S4). These genetic constructs were synthesised by TwistBiosciences and placed within their pTwist CMV vector. The plasmid maps are shown in Figure 2.3. The constructs contain a linker between the mRhubarb720 probe and the target biomolecule (cyan in construct maps).

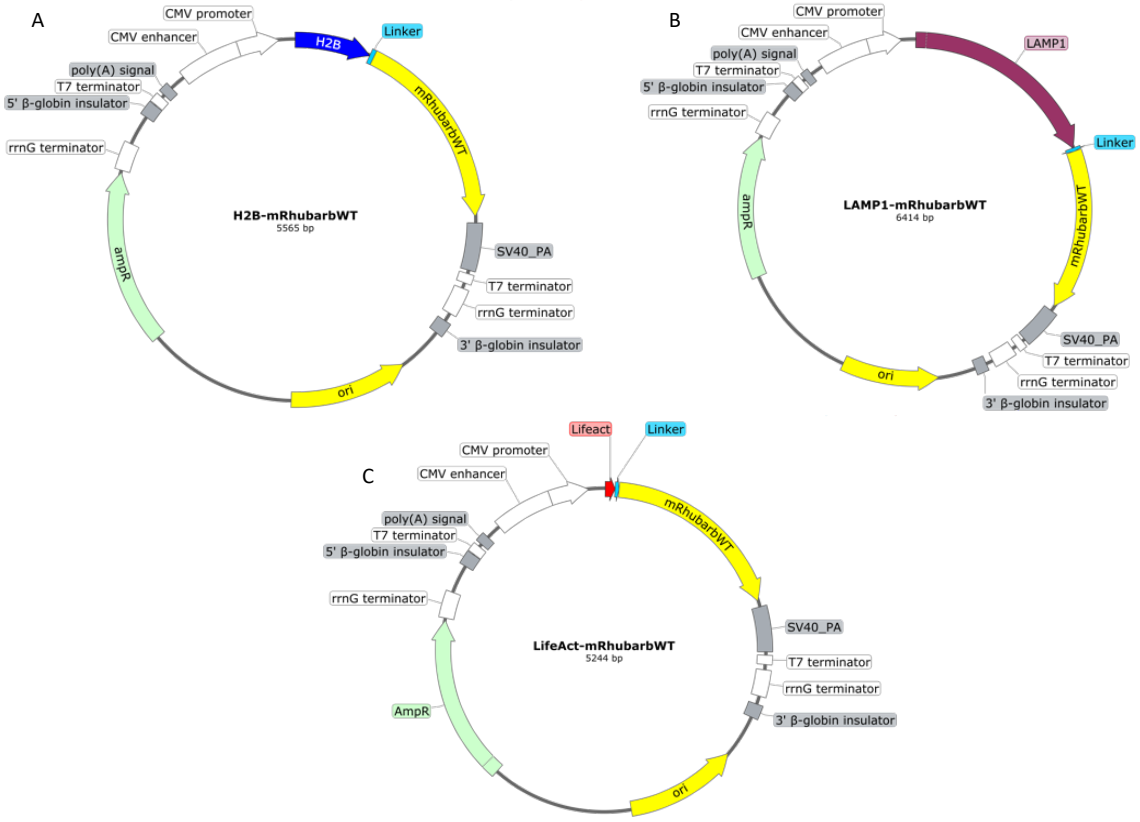


Figure 2.3: Plasmid maps of the genetic constructs used for mammalian cell expression. (A) H2B, (B) LAMP1, (C) LifeAct. Maps produced via SnapGene

2.5.2 Transfection of mammalian constructs

To produce high yields of purified plasmid DNA for transfections, the QIAGEN Maxiprep Plasmid Purification Kit was used. Overnight cultures *E.coli* TOP10™ cells (100 mL), each transformed with a single mammalian construct, were harvested by centrifugation at 5000 rpm at 4°C for 20 minutes using a Thermo Fisher Fiberlite F9-6 x 1000 LEX Fixed Angle Rotor. The supernatant was discarded, and the bacterial pellet was resuspended in 10 mL of Buffer P1. Subsequently, 10 mL of Buffer P2 was added, and the solution was gently mixed by inversion until the mixture turned blue, using LyseBlue reagent as an indicator. This suspension was incubated at room temperature for 3 minutes. Next, 10 mL of Buffer S3 was added, and the tube was inverted until the mixture became colourless. The lysate was then centrifuged at 10000 rpm at 4°C for 30 minutes using a Thermo Fisher Fiberlite F14-6 x 250y Fixed Angle Rotor. The resulting supernatant was transferred to a fresh tube and centrifuged again at 10,000 rpm for 15 minutes to ensure complete removal of the pellet. A QIAGEN column was equilibrated with 10 mL of Buffer QBT, allowing the buffer to flow through by gravity. The supernatant was then loaded onto the column and also allowed to pass through by gravity. The column was washed twice with 30 mL of Buffer QC, with the washes flowing through the column by gravity. DNA was eluted from the column using 15 mL of Buffer QF, collected in a sterile Falcon tube. To precipitate the eluted DNA, 10.5 mL of isopropanol was added, and the mixture was centrifuged at 8,000 rpm for 30 minutes at 4°C using a Thermo Fisher Fiberlite F14-6 x 250y Fixed Angle Rotor. The supernatant was carefully decanted, and the DNA pellet was washed with 5 mL of 70% ethanol. This was followed by another centrifugation at 8,000 rpm for 10 minutes. After removing the ethanol, the DNA pellet was air-dried for 10 minutes and resuspended in 1 mL of TE buffer (pH 8.0). The concentration and purity of the DNA were quantified using a NanoDrop ND 1000 Spectrophotometer (Thermo Fisher).

The following transfection and cell culture steps were performed by Prof Pete Watson. The human cervical cell line HeLa, obtained from the American Type Culture Collection (ATCC) and Mycoplasma tested, was cultured at 5% CO_2 and 37°C in a humidified incubator. Cells were maintained in Minimal Essential Medium (MEM) containing GlutaMAX™ (Thermofisher), supplemented with 10% Foetal Bovine Serum (FBS) (Thermofisher), 1X Non-Essential Amino Acids (NEAA), and 1 mM Sodium Pyruvate. Cells

were grown in the absence of antibiotics. DNA transfection was performed using Fugene6 (Promega) for either 6-well plates (Corning) or MatTek dishes with a 35 mm^2 growth area (Mattek Corporation). For the transfection mix, 3 µL of Fugene6 was added to 100 µL of OptiMEM (Thermofisher), mixed, and incubated for 5 minutes at room temperature. Then, 1000 ng of DNA was added, mixed, and incubated for 20 minutes. After incubation, 94 µL of the transfection mix was pipetted into the cell culture media of the cells.

Cells were passaged every 3-4 days upon reaching 80% confluency. All solutions were pre-warmed to 37°C in a water bath, and passaging was performed in a sterile cell culture hood disinfected with 70% Industrial Methylated Spirit (IMS) before and after use. For passaging, cells were washed once with 1 mL of 0.25% Trypsin/EDTA (Thermofisher), followed by a 5-minute incubation at 37°C with 1 mL of 0.25% Trypsin/EDTA. Cell detachment was confirmed using brightfield microscopy. Detached cells were resuspended in 9 mL of culture media, and new 10 cm culture plates were prepared using cell dilutions of 1:5, 1:10, and 1:20 in a final volume of 10 mL.

2.5.3 Cell preparation for imaging

For bacterial cell imaging, overnight cultures of *E. coli* BL21(DE3) cells expressing mRhubarb720 were grown at 22°C (see Section 2.2.6). Overnight culture (500 µl) was centrifuged for 3 minutes at 10000 rpm at room temperature on a table top centrifuge and supernatant was removed. Cell pellet was resuspended in 50 µl of 50 mM Tris buffer. CultureWell™ silicone gaskets (Grace Bio-Labs) were placed onto 10-12 mm Menzel Glaser glass microscope slides (ThermoFisher) where 50 µl of 1% agarose solution was added to the gaskets. After 20 seconds of cooling, 50 µl of bacteria was added. The slides were then stored at 4°C until imaging commenced (usually within 2 days).

The mammalian cell preparation was performed by Dr Pete Watson and Dr David Regan. For live cell analysis, glass bottomed Mattek dishes were used where the media is replaced with Fluorobrite DMEM with the same supplements as the original media and then imaged on the microscope. The coverslips would be washed in PBS and then fixed with 4% paraformaldehyde in PBS for 20 minutes. They would then be washed in PBS and mounted on a cut down microscope slide using a 0.12mm imaging gasket (Grace Biolabs) with PBS buffer in the resulting chamber.

2.5.4 Imaging techniques

Cell imaging was performed by Dr David Regan. Fluorescence imaging was achieved using a Nikon Ti-U inverted microscope with a 20x 0.75 NA dry objective and a 1.5x tube lens, with a 100 W halogen lamp used for transillumination. The images were taken with a 1 second exposure time and detected by a CCD camera (Hamamatsu Orca 285, having 1344 × 1024 pixels of 6.45 µm size, 18 ke full well capacity, 7 e read noise). For quantitative differential interference contrast (qDIC), images were taken at opposite phase offsets, with the polariser angle set at +/- 30 degrees, taken at a centre wavelength of 550 nm (Nikon GIF plus Schott BG40 filter), with an exposure time of 5.0 ms, averaging over 128 frames per image.

For the coherent anti-Stokes Raman spectroscopy (CARS) and stimulated Raman spectroscopy (SRS) imaging, experiments were conducted using two distinct setups: one at Cardiff University and one at Konstanz University. At Cardiff University, a purpose-built inverted microscope was used with a femtosecond laser system (Figure 2.4). The Stokes beam was generated by a Spectra-Physics InSight X3 laser at 1040 nm with a pulse width of ~120 fs and a repetition rate of 80 MHz. The Pump beam was tuneable between 680 nm and 1300 nm, with a pulse width of ~150 fs. Beam powers were set to 40 mW for the Stokes beam and 13.3 mW for the Pump beam. Detection components included a Hamamatsu H7422-40 photomultiplier tube for forward-CARS signals and a Zurich Instruments HF2LI lock-in amplifier for SRS signals, operating at a modulation frequency of 2.5 MHz. Filter cube F1 (Thorlabs FGS1000) was positioned before the PMT to block excitation light and collect only the blue-shifted anti-Stokes signal in CARS, while Filter cube F2 (Thorlabs FGS1300) was placed before the SRS photodiode to isolate the transmitted Pump beam and suppress residual Stokes light. Dichroic beam splitters (Thorlabs DMSP1000) were used for wavelength separation. Beam alignment and calibration were performed using a polystyrene reference sample before imaging sessions. Data was taken in the order of fluorescence, qDIC then CARS/SRS simultaneously, then repeated for multiple images. For further technical specifications, refer to Regan et al. (2019). (Regan et al. 2019)

At Konstanz University, experiments employed a commercial inverted microscope platform integrated with a femtosecond dual-output laser system. The Stokes beam was tuned to near-infrared wavelengths while the Pump beam operated in the visible spectrum,

both with pulse widths of approximately 100–150 fs. Beam modulation systems enabled precise control, ensuring high signal-to-noise ratios for SRS detection. Lock-in amplifiers and custom optical filters optimized the system for Raman-based imaging. Additional technical specifications and configurations are detailed in (Riek et al. 2016).

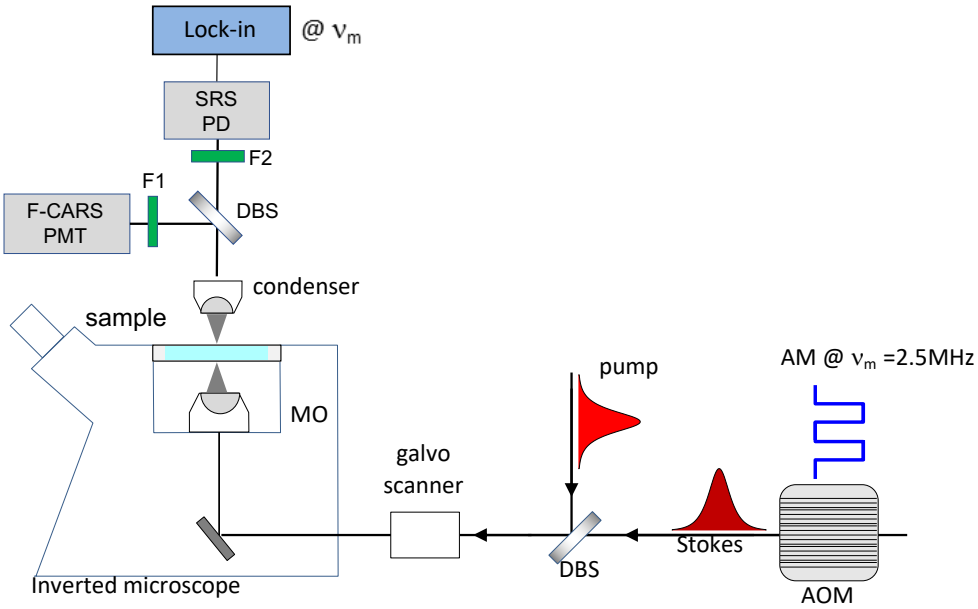


Figure 2.4: Schematic of CARS/SRS microscope used for imaging in Cardiff University. SRS Photodiode (PD) detects the SRS signal and coverts the modulated light to electrical signals. Lock-in amplifier takes these electrical signals at specific modulation frequencies (ν_m) and filters out noise. F-CARS Photomultiplier Tube (PMT) detects the CARS signal. F1 and F2 are optical filter cubes to selectively isolate certain wavelengths. The Dichroic Beam Splitter (DBS) near the optical filter cubes reflects specific wavelengths while transmitting other frequencies for the separation of CARS/SRS signal, whereas the DBS near the laser excitation sources combines the frequencies. The inverted microscope contains the microscope objective (MO) to focus light onto the sample. Galvo scanner moves the Pump and Stokes beams across the sample to scan for imaging. Acousto-Optic Modulator (AOM) modulates the intensity of the Stokes beam at a frequency of ν_m = 2.5 MHz. The amplitude modulation (AM) is critical for the lock-in detection in SRS measurements.

For the measurements, the lock-in amplifier time constant was set to auto, and the phase of the SRS signal was verified using the lock-in reference phase, adjusted to zero degrees (initially 63 degrees). The SRS signal (modulated signal), SRS_DC (unmodulated transmitted Pump), and Fwd CARS (anti-Stokes emission) channels were used for signal

acquisition. SRS was demodulated via lock-in detection at 2.5 MHz, while SRS_DC served as a reference. The Fwd CARS signal was collected via PMT with a gain of 4.13. Before starting the scan, the pump and Stokes beam shutters were closed, and offsets were applied to all channels. The scan was then initiated, with the SRS signal monitored at dim 0 of the lock-in amplifier. All SRS/CARS images were taken using 40 mW Stokes power (100%) and 13.3 mW Pump power (33%) unless stated otherwise. Scan centres were set to 1740 cm^{-1} for fingerprint region scans and 2100 cm^{-1} for biological silent window scans. There were 20 repetitions for each sample, unless stated otherwise. Images were processed and quantitatively measured and analysed using ImageJ software (Schneider et al. 2012).

2.6 Molecular dynamics

2.6.1 Computational requirements

Molecular dynamics (MD) and quantum mechanical (QM) simulations required high performance computation (HPC) resources, therefore access to Supercomputing Wales HAWK server was used, containing 280 nodes, 12,736 cores and 69 TB of total memory. Both CPU (AMD(R) Epyc 7502 2.5GHz with 32 cores each) and GPU (Nvidia V100 GPUs with 16GB of RAM on 15 nodes) was implemented in this research. The GROMACS software package was used for MD simulations (Van der Spoel et al. 2005; Pronk et al. 2013; Abraham et al. 2015). Gaussian09 software package was used for all QM simulations (Frisch et al. 2016).

2.6.2 Molecular models for MD simulations

For MD simulations of mRhubarb720, input structures were taken from the crystal structure (see Chapter 3; PDB 9F03). For mCherry (PDB: 2H5Q) the crystal structure was retrieved from the Protein Data Bank (Shu et al. 2006). Molecular analysis software, PyMOL, was used to prepare the structures for input into the simulations, with pCNPhe residue mutagenesis performed using the in-built mutagenesis function (Schrödinger 2015). The pCNPhe residue was designed computationally by Dr David Gfeller and incorporated as part of the PyMOL mutagenesis options through the SwissSideChain plugin (Gfeller et al. 2012; https://www.swisssidechain.ch/). The chromophore mutations were built manually within PyMOL using its in-built builder function.

2.6.3 Parameterisation of novel compounds

The CHARMM27 force field was employed to parameterise BV bound to mRhubarb, free-BV, and variants of the mCherry chromophore (Mackerell et al. 2004; Bjelkmar et al. 2010). Parameterisation involved generating the force field parameters, including atomic charges, bond lengths, angles, dihedrals, impropers and non-bonded interactions, to accurately describe the chemical properties of each compound. Initial geometries were obtained from X-ray crystallographic data, with atomic charges derived using Atomic Charge Calculator 2 (Raček et al. 2020). The molecular mechanics for each parameterisation is expressed by the CHARMM27 forcefield function (Equation 2).

Equation 2

$$v = \sum_{bonds} \frac{1}{2} k_b (b - b_o)^2 + \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_o)^2 + \sum_{dihedrals} k_\phi [1 + cos(n\phi - \delta)] + \sum_{impropers} \frac{1}{2} k_\omega (\omega - \omega_o)^2 + \sum_{Urey-Bradley} \frac{1}{2} k_u (u - u_o)^2 + \sum_{nonbonded} \epsilon_{LJ} \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - \left(\frac{R_{min,ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\epsilon r_{ij}}$$

The potential energy (v) of a molecular system is a summation of internal and non-bonded interactions. Internal interactions are defined by bonds (b), angles (θ), Urey-Bradley (u), dihedral angles (ϕ), and improper dihedral angles (ω). Each internal interaction is associated with a force constant (k_b, k_θ, k_u, k_ϕ, and k_ω), which defines the energy contribution (kcal/mol) per unit deviation (Å^2, radian^2) from equilibrium values. The potential energy contributions from bond, angle, Urey-Bradley, and improper terms are modelled as quadratic functions, reflecting the deviation from equilibrium values (b_o, θ_o, u_o, and ω_o). For bond lengths, the distance between two bonded atoms is compared against its equilibrium distance. Similarly, the angle term describes the angle between two bonds connecting three atoms, measured against its equilibrium value. The Urey-Bradley function accounts for the distance between the outer atoms in a three-atom sequence, and its equilibrium distance. Improper dihedrals maintain planarity by measuring the angle

between the planes formed by four atoms (i, j, k, l), where three peripheral atoms (j, k, l) are bonded to a central atom (i), compared to its equilibrium value. Dihedral terms involve a trigonometric function that accounts for rotational barriers, with n representing the periodicity and δ the phase shift, to capture torsional flexibility around bonds. The dihedral energy term depends on the angle between planes defined by four consecutively bonded atoms, describing the rotational freedom.

The final summation representing the non-bonded interactions are modelled using a combination of Lennard-Jones (LJ) and Coulombic functions. The Lennard-Jones potential describes van der Waals interactions, where ϵ_{LJ} defines the depth of the potential well, $R_{min,ij}$ is the distance at which the potential is minimised (between atom i and j), and r_{ij} is the distance between atoms. Electrostatic interactions are calculated using Coulomb's law, where q_i and q_j are the partial charges of interacting atoms, and ϵ is the dielectric constant, scaling the interaction over the distance r_{ij}.

Parameter values, including force constants and bonded parameters, were calculated using ACPYPE (AnteChamber Python Parser interfacE), which automates the generation of topology and parameter files compatible with CHARMM27 (Sousa da Silva and Vranken 2012). All parameters for BV-bound to mRhubarb720, free-BV and mCherry chromophore variants are listed in the Supplementary Material P1-3.

2.6.4 Molecular dynamics simulations

Input models in PDB format were converted to GROMACS-compatible Gromos87 (.gro) files, where hydrogens were added. CHARMM27 forcefield parameters modified to include the new components outlined in Section 2.6.3 were applied. The parameterisation ensured accurate representation of the molecular structures within the simulations. Each model was placed into a cubic simulation box with dimensions set to three times the largest distance between atoms in the structure to prevent boundary effects. The boxes were then solvated using a CHARMM-implemented TIP3P water model, with LJ-potentials assigned for water interactions. The overall charge of each system was calculated, and counter-ions (either Na^+ or Cl^-) were added to neutralise the systems.

Energy minimisation was performed using the steepest descent algorithm with a step size of 0.01 nm and a maximum force threshold of 1000 kJ/mol/nm to optimise the potential

energy. During minimisation, interactions between neighbouring atoms were calculated using a Verlet cut-off scheme, with a neighbour list updated at each step. Electrostatics were treated with Particle Mesh Ewald (PME) using a 1 nm cut-off for Coulomb interactions, while van der Waals interactions were considered up to a 1 nm cut-off. Periodic boundary conditions (PBC) were applied in all three dimensions (xyz) to simulate an infinite system.

Temperature, pressure, and density were controlled through a two-step equilibration process using a leap-frog integrator. Each equilibration step was conducted for 100 ps with a 2 fs time step. The LINCS algorithm was used to apply holonomic constraints, ensuring accurate bond length correction during rotational motions. Constraints were applied to all bonds involving hydrogen atoms. Neighbour searching was handled using the Verlet cut-off scheme, and PME was used for electrostatics with a Fourier-space grid spacing of 0.16 nm. Velocity rescaling with a stochastic distribution was used for temperature coupling. The system was divided into separate atom groups for the protein and solvent during equilibration. Temperature coupling was applied with a time constant of 0.1 ps, maintaining a reference temperature of 311 K (37.5°C) to mimic biological conditions. In the second equilibration step, isotropic pressure coupling was included using the Parrinello-Rahman barostat, allowing uniform scaling of the simulation box dimensions. The pressure coupling time constant was set to 2 ps, with a reference pressure of 1 bar. Following equilibration, density values were verified before proceeding to the production phase. The equilibrated systems served as inputs for the production runs, maintaining the same parameters as used in equilibration. Each production simulation was conducted for between 200 ps and 500 ns with a 2 fs time step, repeated in triplicate. Trajectory files of each simulation were converted so that the proteins were centred in the water box for the following analysis.

2.6.5 Calculating Root Mean Square Deviations (RMSD)

The RMSD provides a measure of how much a structure deviates from a reference structure over time by using the ‘gmx rms’ command and is calculated using Equation 3. The reference structure in each case is the first frame of the trajectory. For whole protein RMSD calculations, the deviation is measured using the backbone atoms. For the chromophore RMSD calculations, all atoms were calculated.

Equation 3

$$RMSD(t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left|r_i(t) - r_i^{ref}\right|^2}$$

At each given time point (t), each atom has a position of r_i, in respect to the reference coordinate of the same atom r_i^{ref}, where N is equal to the number of atoms measured.

2.6.6 Calculating Root Mean Square Fluctuations (RMSF)

The RMSF is a measure of the fluctuation of individual atoms or residues with respect to their reference structure using the 'gmx rmsf' command (Equation 4). Similarly to the RMSD calculation, for whole protein RMSF calculations, the C_α was selected to measure the fluctuation of the residues whereas all heavy atoms (no hydrogens), were selected for the chromophore analysis.

Equation 4

$$RMSF(i) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left|r_i(t) - r_i^{ref}\right|^2}$$

Each atom (i) has a position of r_i relative to the reference position of the measured atom r_i^{ref}, at any given time point (t). The total number of time points (T) is used to average and normalise all given time points over the entire simulation.

2.6.7 Calculating the Radius of Gyration (Rg)

The Rg is a measure of whether the protein or chromophore is expanding or contracting throughout the simulation, using the 'gmx gyrate' command. For all Rg measurements, each atom within either the protein or the chromophore was used in Equation 5.

Equation 5

$$R_g = \sqrt{\frac{\sum_{i=1}^{N} m_i \left| r_i - \left(\frac{\sum_{i=1}^{N} m_i r_i}{\sum_{i=1}^{N} m_i} \right) \right|^2}{\sum_{i=1}^{N} m_i}}$$

The bracketed term represents the position of the centre of mass for the entire system, where the position of each atom r_i and the mass of each atom m_i weighted against the average of the measured structure. The squared distance of each atom from the centre of a mass is calculated, indicating how far each atom is away from this central point, with the summation of all atoms (N) to give a mass-weighted sum of these squared distances.

2.6.8 Calculating the number of hydrogen bonds

Using the 'gmx hbond' command, the number of hydrogen bonds were calculated on specified cut-offs for both distance and angle. In this analysis, hydrogen bonds were defined by a cut-off distance of 0.35 nm between the donor and acceptor atoms and an angle cut-off of 30° for the hydrogen-donor-acceptor configuration. The hydrogen atoms of OH and NH groups were treated as hydrogen bond donors, with oxygen (O) atoms and nitrogen (N) atoms were regarded as acceptors. For each mRhubarb720 simulation, hydrogen bonds were analysed between two specified groups: BV as one group, and the combined protein and solvent as the second group. For the mCherry simulations, separate analyses were performed for each variant: the deprotonated oxygen in WT, the nitrile group in CRO-CN, and the ethynyl group in CRO-CC (all positioned at the para-position on the phenolic ring) were each defined as one group, with the combined protein and solvent as the second group.

2.6.9 Calculating the Solvent Accessible Surface Area (SASA)

The SASA measures the surface area of atoms accessible to the solvent in a molecular system via the 'gmx sasa' command. For each analysis, the solvent probe radius was set to 0.14 nm, representing an approximation of the effective radius of a water molecule as a single entity. Each SASA simulation was computed for the whole protein and the chromophores individually.

2.6.10 Kruskal-Wallis statistical test

The Kruskal-Wallis test is a non-parametric statistical method used to compare the medians of three or more independent groups (Kruskal and Wallis 1952). Unlike parametric tests (such as ANOVA), the Kruskal-Wallis test does not assume that the data follow a normal distribution, making it suitable for non-normally distributed or skewed data, which are common in MD simulations. The Kruskal-Wallis test is calculated in Equation 6.

Equation 6

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} n_i \left(R_i - \frac{N+1}{2} \right)^2$$

The test statistic, H, is calculated with the total number of observations across all groups, N. The number of groups (variants) is represented by k, and the number of observations in the i-th group is n_i, with R_i representing the rank for each i-th group. The resulting H-value follows a chi-square distribution with $k - 1$ degrees of freedom, which determines the resulting p-value.

In this study, the Kruskal-Wallis test was applied to assess whether there were statistically significant differences in RMSD, RMSF, SASA, Rg and hydrogen bonding across WT and mutant simulations. Due to the high number of data points in MD simulations, a strict p-values cut-off was selected ($p < 0.0001$). With large datasets, a more stringent threshold for significance is often necessary to ensure that differences reflect meaningful structural changes rather than mere statistical artifacts. The tests for the MD simulations yielded very low p-values, suggesting that even subtle differences between groups were statistically significant. Each simulation was tested against all three WT simulations. Mutant simulations with non-significant p-values against any of the WT simulations, were treated as such and were not flagged as having significant differences.

2.7 Quantum mechanical approaches and theory from first principles

2.7.1 Molecular structures for QM simulations

The initial QM simulation structures used were retrieved from the MD simulations after the equilibration steps, as detailed in Section 2.6.4. The MD simulation systems contained the whole protein, solvent molecules and ions that are not directly involved in the QM analysis. Only the core region containing the chromophore and selected interacting residues were retained (BV: Chapter 5, Figure 5.28; mCherry: Chapter 6, Figure 6.11-12), whereas the rest of the protein and solvent were removed using Gaussview5 (Dennington et al. 2016). To balance computational efficiency, non-essential portions of interacting residues and parts of the mCherry chromophore were truncated (BV was not truncated for these simulations). Regions that did not contribute to the chromophores electronic environment were removed and replaced with methyl groups to simplify the QM simulations. The atomic positions of the interacting residues, and the added methyl groups, were frozen throughout the QM analysis to preserve their interactions with the chromophore. For simulations involving chromophore mutations, modifications were made using the builder function in Gaussview5. All input structures and Gaussian09 scripts can be found in the Supplementary Section 9.4.

2.7.2 Simulation parameters

The following settings were used for each geometry optimisation and subsequent frequency and TD-DFT calculations. The Polarisable Continuum Model (PCM) was used, via the keyword 'scrf=(solvent=water)', to simulate the aqueous environment around the chromophore and interacting residues. The solvent, water, was modelled with a static dielectric constant (ϵ) of 78.3553. Gaussian09 constructs the solvent cavity using a scaled van der Waals cavity (scaling factor = 1.100) via the GePol algorithm with Universal Force Field (UFF) atomic radii. Polarisation charges were calculated via matrix inversion, using spherical Gaussians and a Karplus/York smoothing algorithm to ensure continuity across the molecular surface (Pascual-ahuir et al. 1994; Steindal et al. 2011; Mennucci 2012).

The ‘int=acc2e=11’ keyword was used throughout each calculation to reduce computation times. This setting adjusts the precision with convergence at 10^{-11} (default is 10^{-12}) which two-electron integrals are calculated (Welker et al. 2015). Two-electron integrals describe the interactions between pairs of electrons, accounting for their repulsion and its impact on the overall energy of the system. In preliminary tests, it was found that without this setting, the simulation times increased substantially, making this adjustment necessary to achieve a balance between speed and accuracy.

For the first optimisation, ‘geom=connectivity’ keyword was used to maintain the atomic connectivity derived from the MD structures. For all subsequent optimisations, ‘geom=checkpoint’ and ‘guess=read’ were used to read the optimised geometry and wavefunction from previous steps, ensuring continuity and faster convergence. For the frequency and TD-DFT calculations, geometries were retrieved from the final geometry optimisation step and computed with ‘geom=connectivity’ keyword again.

2.7.3 Functional and Basis set selection

The concept behind each functional and basis set is described in Chapter 1.4.3.1. For the geometry optimisation step of each structure, a three-step approach was initiated: HF/3-21G, B3LYP/6-31G and B3LYP/6-31G+(d,p). This strategy was chosen to balance computational efficiency with the accuracy of the resulting optimised geometries. The initial optimisation was performed using the Hartree-Fock (HF) method with the 3-21G basis set to provide a computationally efficient starting point to relax the MD minimised structures. This basis set includes a smaller set of atomic orbitals, allowing for fast convergence and to establish a reasonable starting geometry.

The second geometry optimisation employed the B3LYP functional with the 6-31G basis set. B3LYP was chosen for its ability to accurately describe electron correlation effects, which are critical for capturing interactions within the delocalised structures of the chromophore. The 6-31G basis set offers a more detailed representation of molecular orbitals compared to the initial 3-21G calculation, enhancing the accuracy of the optimised structures while maintaining reasonable computational costs.

The final step utilised the B3LYP/6-31G+(d,p) level of theory. This basis set includes a diffuse function (+) to better describe electron densities in regions far from the nuclei,

particularly important for anionic states. The inclusion of polarisation functions (d,p) adds flexibility in the shape that orbitals can adopt, allowing the basis set to better describe bonding within a molecule. The frequency and TD-DFT calculations were also conducted using this level of theory to ensure a consistent and accurate evaluation of the electronic properties of the chromophores. This functional and basis set pair were used in previous literature on similar structures to ones in this thesis (Topol et al. 2011; Tachibana et al. 2021).

2.7.4 Geometry optimisation

The objective of geometry optimisation is to identify the most stable molecular structure, where the potential energy of the system is minimised and the forces acting on each atom are balanced. This process is central to understanding the stable conformations of molecular systems, such as chromophores, in QM simulations. The underlying concept is the exploration of the potential energy surface (PES), which represents the energy landscape of a molecule as a function of its atomic positions. Over many iterations in adjusting atomic coordinates, the PES should reach a local energy minima, representing a stable structure. The process involves calculating the potential gradient of the energy with respect to atomic positions, as represented in Equation 7.

Equation 7

$$F_i = -\frac{\partial E_i}{\partial r_i}$$

The forces acting on each atom (F_i) are derived from the gradient of energy (E), with respect to the position of atom r_i. The negative gradient indicates to Gaussian09 that it should adjust the atomic positions in the direction of where the energy decreases. A gradient-based algorithm known as the Berny optimisation method (Gaussian09 default), calculates the energy and energy gradient for the current geometry and then adjusts the atomic positions accordingly to reduce the energy (Li and Frisch 2006).

To describe the curvature of the PES, the software uses a Hessian matrix (H_{ij}) to determine how atoms should be adjusted as the optimisation approaches a local minimum

(Equation 8). The matrix derives how the energy (E) changes with respect to small changes in the coordinates r_i and r_j (which represent positions of atoms i and j).

Equation 8

$$H_{ij} = \frac{\partial^2 E}{\partial r_i \partial r_j}$$

The partial derivative indicates how the energy responds to small changes in atomic positions. For example, the first derivative $\frac{\partial E}{\partial r_i}$ measures how the energy changes as the position i is varied along a particular direction (x, y, z) while keeping all other atoms fixed, which indicates the sensitivity of the energy to the direction of movements of a singular atom. When incorporating the second derivative, $\frac{\partial^2 E}{\partial r_i \partial r_j}$, this describes how the energy changes when both atom i and j are moved, these analyses are commonly used to test whether the energy is on a saddle point or in fact in a local minima (Li and Frisch 2006).

The Hessian matrix effectively encodes for the force constants for bond stretching, angle bending and torsions when optimising the molecule, whereas the atomic charges influence the gradient of energy with respect to the forces acting on the atom. The iterative optimisation process continues until specific convergence criteria are met, signifying that the structure has reached a stable minimum on the PES. In this study, the criteria included a maximum force on an individual atom is less than 4.5×10^{-4} Hartree/Bohr and a maximum root mean square (RMS) force across all atoms is less than 3×10^{-4} Hartree/Bohr. As mentioned above, this is step is repeated three times for each structure, using progressively complex functionals and basis sets until the B3LYP/6-31G+(d,p) structure is optimised.

2.7.5 Calculating classical dipole moments

The distribution of discrete electronic charge in the molecule directly influences how it interacts with the external electric fields, which is determined by calculating the classical dipole moment (Equation 9).

Equation 9

$$\vec{\mu} = \sum_i q_i \vec{r}_i$$

The dipole moment ($\vec{\mu}$) is the summation of the charges (q_i) of each atom (i) with $\vec{r}_i$ representing the direction of the vector (Housecroft and Sharpe 2018). This is achieved by performing a single-point energy on Gaussian09. Transition dipole moments describe the transition between ground and excited states, this can be observed in Equation 18 (Chapter 2.7.8).

2.7.6 Density Functional Theory (DFT)

Before describing the DFT methodology, it is important to note that all electronic structure calculations herein are performed under the Born–Oppenheimer approximation (Born and Oppenheimer 1927). This fundamental approximation decouples the electronic and nuclear motions by assuming that the nuclei, being much heavier and moving on much slower timescales, remain essentially stationary during electronic transitions. As a result, the electronic wavefunctions and energies are determined for fixed nuclear positions, while the nuclear kinetic energy is neglected in the solution of the electronic Schrödinger equation.

Under the Born-Oppenheimer approximation, the standard electronic Schrödinger equation takes the form, $\hat{H}\psi = E\psi$, where $\hat{H}$ is the electronic Hamiltonian operator, ψ is the wavefunction and E is the corresponding energy eigenvalue (Schrödinger 1926). This equation formally describes how the energy and spatial distribution of electrons are determined for a static arrangement of nuclei. However, solving this exactly for real systems is difficult, but DFT offers an alternative by focusing on the electron density rather than the full many-electron wavefunction. The Kohn-Sham equations, derived within DFT, replace the many-body problem with a set of effective single-particle equations that retain a similar mathematical structure to the Schrödinger equation.

The energies of the HOMO (Highest Occupied Molecular Orbital) and LUMO (Lowest Unoccupied Molecular Orbital) were determined using Density Functional Theory (DFT) via frequency calculations, as implemented in Gaussian09. Schrödinger's equation, which is fundamental to quantum mechanics, underlies these calculations and provides the basis from which the Kohn-Sham equation is adapted (Sahni 2016). The Kohn-Sham equation

serves as the foundation of DFT, providing a framework to the many-body problem by balancing the kinetic energy of electrons, their interactions with nuclei, and electron-electron interactions. This process involves a series of calculations that are systematically combined to determine the energy levels and wavefunctions of the molecular orbitals.

The quantum state of an electron is fundamental to this concept, and it is represented by the wavefunction (ψ_r) which provides information about the probability amplitude of finding an electron at particular position r in space. By squaring the wavefunctions absolute value ($|\psi_r|^2$), this determines the probability density, that describes where the electron is most likely to be found. To ensure that the wavefunction represents a full probability distribution, it must be normalised to equal 1, meaning that the total probability of finding the electron inherently satisfied. To do this, the integral is taken over all space, expressed as Equation 10.

Equation 10

$$\int |\psi_r|^2 \, dr = 1$$

This equation then can be extrapolated into calculating the expectation values $\langle A \rangle$, such as position, energy (Hamiltonian operator, $\widehat{H}$), or momentum of a given system using the following the integral in Equation 11.

Equation 11

$$\langle A \rangle = \int \psi_r^* A \psi_r \, dr$$

This term contains a complex conjugate ψ_r^*, which is a fundamental requirement of quantum theory. It ensures that expectation values, such as those for probability densities and energy, are real quantities. Known as the operator A, represents the quantity measured where observations are taken. The integral sums up all the possible contributions across a given space of an electron, weighted by the probability of finding the electron at any given point, given by the product of $\psi_r^* \psi_r$ (Fleming 2024).

Building on this foundation, the Laplacian operator ∇^2, is applied to the wavefunction to generate the curvature in 3D space by deriving the Cartesian coordinates (x, y, z) via partial derivatives in Equation 12.

Equation 12

$$\nabla^2\psi_r = \frac{\partial^2\psi_r}{\partial x^2} + \frac{\partial^2\psi_r}{\partial y^2} + \frac{\partial^2\psi_r}{\partial z^2}$$

By combining Laplacian operator with the mass of an electron (m) and reduced Plancks constant ($\hbar$), the first expression of the Kohn-Sham equation is determined, where the kinetic energy contribution of a particular electron within a particular molecular orbital is given by Equation 13.

Equation 13

$$-\frac{\hbar^2}{2m}\nabla^2\psi_r$$

Since the kinetic energy and the positions of an electron can be described by the equations stated above, the next component of the Kohn-Sham equation describes the various interactions that affect electrons within an orbital. Termed as the effective potential $V_{eff}(r)$, which is equal to the sum of three potentials $V_{ext}(r) + V_H(r) + V_{XC}(r)$, given by Equation 14.

Equation 14

$$V_{eff}(r) = -\sum_{\alpha}\frac{Z_\alpha}{|r - r_\alpha|} + \int\frac{\rho(r')}{|r - r'|}dr' + V_{XC}(r)$$

The first term is the external potential $V_{ext}(r)$ accounts for the attraction between electrons (r) and the nuclei (α) with the charge of the nucleus (Z_α) and its position (r_α) being accountable to the calculation. The second term is the Hartree potential $V_H(r)$ that represents the Coulomb repulsion between electrons calculated directly from the electron

density $\rho(r')$ of all given positions. This accounts for the repulsion force at position r, based on their spatial distribution. And finally, the exchange-correlation potential $V_{XC}(r)$, which is the most complex component of the Kohn-Sham equation. This accounts for the effects of electron exchange, governed by the Pauli exclusion principle, as well as correlation effects that describes how electrons avoid each other due to their interactions. Since the exact form of V_{XC} is still unknown, functionals like B3LYP are crucial. These provide approximations to the value by combining HF theory with empirical correlation terms, in order to accurately simulate electron-electron interactions, balanced with computational efficiency. Combining Equations 13 and 14, the complete Kohn-Sham equation, used as the basis of DFT, can be stated in Equation 15.

Equation 15

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V_{eff}(r)\right)\psi_r = \epsilon_i\psi_r$$

This equation is iteratively solved by Gaussian09 for each molecular orbital ψ_r, calculating orbital energies ϵ_i that describe the energy levels of the electrons. The highest of these occupied energy orbitals is the HOMO and the lowest unoccupied energy orbitals is the LUMO. The energy differences between these occupied and virtual/unoccupied orbitals are critical for understanding the electronic properties of the molecule, including optical and transition characteristics (Hohenberg and Kohn 1964; Kohn and Sham 1965; Yu et al. 2016).

2.7.7 Time-Dependent Density Functional Theory (TD-DFT)

To study the excited states of the chromophore and its interaction with light, TD-DFT was implemented to extend the principles of ground state DFT to include time-dependent perturbations. This allows the simulation of electronic transitions when a molecule absorbs a photon. This is achieved by adapting the Kohn-Sham equation to account for the behaviour of electrons for time-dependant fields, such as light, changing the overall electron density (Marques and Gross 2004). The theoretical foundation for these time-dependent approaches originates from the time-dependent Schrödinger equation, expressed as in Equation 16.

Equation 16

$$\widehat{H}\psi(r,t) = j\hbar\frac{\partial}{\partial t}\psi(r,t)$$

The $\widehat{H}$ is the Hamiltonian operator, $\psi(r,t)$ is the time-dependent electronic wavefunction, j is the imaginary unit and the $\hbar$ is the reduced Planck constant. This equation describes how the quantum state of a system evolves over time in response to external perturbations, such as oscillating electric fields from incident light. Within TD-DFT, this formalism is adapted by using a time-dependent version of the Kohn-Sham equation, where non-interacting electrons move under an effective potential that evolves with the time-dependent electron density. The time-dependant Kohn-Sham equation takes a similar form and is expressed as Equation 17 (Schrödinger 1926; Kohn and Sham 1965).

Equation 17

$$j\hbar\frac{\partial}{\partial t}\psi_i(r,t) = \left(-\frac{\hbar^2}{2m}\nabla^2 + V_{eff}[\rho(r,t)]\right)\psi_i(r,t)$$

In this equation, j is the imaginary unit that corrects how the electron density changes in response to the oscillating nature of the electric field from light. The term $\frac{\partial}{\partial t}$ represents the time derivative of the wavefunction $\psi_i(r,t)$, that defines how a particular i-th molecular orbital changes in respect to time at position r. The $\rho(r,t)$ is the time-dependant electron density, where the effective potential $V_{eff}[\rho(r,t)]$ now depends on the evolving changes in electron density to calculate electron-nuclear, electron-electron interactions (Zwiebach 2022).

2.7.8 UV-Vis plot transformations from TD-DFT excited state calculations

The TD-DFT calculations output energies for the ground state (S_0) and the specified number of excited states (S_1…S_i). The calculations in this thesis used the keyword 'nstates=10' to calculate the first 10 excited states (S_1…S_{10}). The HOMO-LUMO (S_0-S_1) gap is calculated for each of these transitions equalling the excitation frequencies (v_i) for the UV-Vis spectrum. The transition dipoles are concurrently calculated by deriving Equation 16 to

describe the electron density changes in response to light between two states and is defined by Equation 18.

Equation 18

$$\vec{\mu}_{ti} = -e \int \psi_0 \vec{r} \psi_i dr$$

This integral essentially sums the products of the position vectors $\vec{r}$ and the overlap between ground state wavefunctions ψ_0 and excited state wavefunctions ψ_i, weighted by the electron charge $-e$, capturing how the electron cloud shifts during transition, giving the transition dipole ($\vec{\mu}_t$) for the i-th transition. Gaussian09 uses Dipole strength D_i to calculate oscillator strength which is given by the absolute square of the transition dipole, $D_i = \left|\vec{\mu}_t\right|^2$ (Williams 1999). The relationship between D_i and oscillator strength f_i is given by Equation 19.

Equation 19

$$f_i = \frac{8\pi^2 v_i m_e c}{3he^2} D_i$$

The formula contains a number of constants: mass of the electron (m_e), charge of an electron (e), speed of light (c) and Planck's constant (h). Combined with excitation frequencies (v_i) calculated from TD-DFT will give f_i. The final step to generating an absorption is to apply a Gaussian broadening parameter on all states calculated which is described in Equation 20.

Equation 20

$$A(\lambda) \propto \sum_{i=1}^{n} f_i \cdot \frac{1}{\sigma} exp\left[-\left(\frac{\lambda - v_i}{\sigma}\right)^2\right]$$

The calculated absorbance $A(\lambda)$ is proportional to the expression with relative intensities at wavelength λ scaled. Each transition i with respective frequency v_i and oscillator strength f_i is broadened by the Gaussian term σ which can be defined as the

standard deviation in wavenumbers, related to the width of the simulated band (in GaussView5, σ = 0.4 eV as default). The final absorbance $A(\lambda)$ is the sum of all these contributions (nstates = 10 in this study), which provides a smooth absorbance spectrum.

Equation 19 and 20 is provided in the Gaussian09 documentation for generating UV-Vis plots (https://gaussian.com/uvvisplot/). Please note, Equation 20 was derived as this study presented the values in arbitrary units (a.u.) to emphasise the relative peak intensities and positions rather than the absolute absorbance values. The reason for this was not all interacting residues were present in the simulations which skewed the absolute absorbance intensities, therefore not comparable to experimental data. As a result, unit-specific scaling factors were removed from the derivation listed in the documentation.

2.7.9 Determining Raman activity and intensity

The 'freq=raman' keyword was used for all QM models to perform a vibrational analysis that calculates how the polarisability of a molecule changes as it undergoes different vibrational motions. Polarisability describes how the electron cloud around a molecule can be distorted by an external electric field, such as incident light. This is why the polarisation functions in the 6-31G+(d,p) basis set was critical as it captures all the electron distortions within the biomolecules.

To calculate the Raman activity/intensity of the simulation, each individual mode k requires determining the polarisability tensor, which describes how the electron cloud of the molecule is distorted when subject to an electric field during vibrational motion. There are six in total: $R_{kxx}, R_{kyy}, R_{kzz}, R_{kxy}, R_{kxz}, R_{kyz}$, which uses a principle axes system (xyz) that are unique for any given Raman band, and essentially describes how the dipole moment of the molecule changes when responding to light (Tsuboi et al. 2009). For example, if the electric field is applied in the y direction, and the change in dipole is in the x direction, the off-diagonal component R_{kxy} tensor is captured (units of Å^3).

The components of polarisability tensors will be used to calculate the isotropic polarisability (a), which is the average change in polarisability across three spatial dimensions (xx, yy, zz) as the molecule vibrates (Equation 21). The anisotropic polarisability (γ^2), which quantifies how the shape of the electron cloud changes directionally (xy, xz, yz), offers insights into the asymmetric distortion during the vibrational mode (Equation 22).

Equation 21

$$a = \sqrt[3]{\left(R_{kxx} + R_{kyy} + R_{kzz}\right)}$$

Equation 22

$$\gamma^2 = \sqrt{\left(R_{kxx} - R_{kyy}\right)^2 + (R_{kxx} - R_{kzz})^2 + \left(R_{kyy} - R_{kzz}\right)^2 + 6\left(R_{kxy} + R_{kxz} + R_{kyz}\right)}$$

These equations are well documented in literature and are critical for calculating the Raman activity of a vibrational mode and the Raman intensity (Porezag and Pederson 1996; Long 2002). To calculate the Raman activity R_A for each individual vibrational mode, is given by Equation 23.

Equation 23

$$R_A = 45a^2 + 7\gamma^2$$

The constants 45 and 7 derive from the theory of Raman scattering for randomly oriented molecules that are essentially weighting factors that balance the contributions from all molecular orientations (Bogaard and Haines 1980). The R_A can then be converted into the differential Raman scattering cross-section ($\frac{d\sigma_v}{d\Omega}$), which provides a measure of the scattered lights intensity for a given vibrational mode as a function of the scattering angle, stated in Equation 24.

Equation 24

$$\frac{d\sigma_v}{d\Omega} = \frac{\omega_S^4 \hbar(n+1)}{2\omega_k c^4} \cdot \frac{R_A}{45}$$

This equation incorporates many constants such as, reduced Planck's constant ($\hbar$), speed of light (c), Bose-Einstein thermal population term $(n+1)$ and 45^{-1}. These factors account for temperature variations and the dependence of the scattered light's intensity on the incident photon's energy (Bagheri and Komsa 2023). Such factors are essential for precise measurements in experimental Raman spectroscopy, where the environment can

influence the observed intensity. However, for computational simulations, these constants do not change across vibrational modes and can be simplified to focus on the relative Raman intensities, given by Equation 25.

Equation 25

$$I_k \propto \frac{\omega_S^4}{\omega_k} \cdot R_A$$

In this equation, the relative Raman intensity (I_k) is proportional to the scattered lights frequency (ω_s) and inversely proportional to the vibrational modes frequency (ω_k). The power of 4 indicates that the intensity of the scatter light increases substantially as the scattered frequency increases. Angular light frequency can then be converted to linear light frequency ($v_s = \omega_s/2\pi$), therefore, calculating the Raman shift (Δv) will be possible by $\Delta v = v_0 - v_s$. Where the incident light frequency (v_0) is set to default by Gaussian09 (1064 nm). The value of v_0 will primarily influence the calculated Raman intensities rather than the actual positions of the vibrational peaks (frequencies) on the Raman spectrum. The Raman shift Δv depends on the difference between the incident and scattered frequencies, making it independent of the absolute value of v_0. As a result, Δv directly corresponds to the energy difference of absorption/emission of the vibrational modes of the molecule, allowing the spectrum to represent the molecular vibrations accurately.

Once the Raman shifts and corresponding intensities are calculated, Gaussian broadening is applied to produce a realistic spectrum, similar to the UV-Vis spectrum. The broadened intensity profile for each vibrational mode is given in Equation 26.

Equation 26

$$I(\tilde{v}) \propto \sum_k I_k \cdot \exp\left[-\left(\frac{\tilde{v} - \Delta v}{\sigma}\right)^2\right]$$

The intensity (I) at each wavenumber, $\tilde{v}$, is being calculated on the spectrum for each vibrational mode k. The σ is the standard deviation of the Gaussian function, set at default to 10 cm^{-1} (0.00124 eV), which determines the width of each peak. This value is much smaller than the absorbance spectra calculation because of the sharper spectral peaks of

vibrational modes (He et al. 2023). Finally for each predicted Raman spectra, a scaling factor of 0.97 was applied to the wavenumbers as the B3LYP method has been reported to overestimate the raw frequencies by 3% (Sinha et al. 2004).

3 Characterisation of near-IR FPs and structure determination of mRhubarb720.

3.1 Introduction

Using the in-house set up to implement electronically pre-resonant Raman scattering (epr-RS), far-red/near-IR chromophores are required to exploit the molecular electronic transitions by tuning the Stokes laser to match the molecules excitation energy. This is due to the nature of the 820 nm Pump laser used to excite the molecules in pre-resonant excitation conditions, where excitation is required in the far red/near-IR spectrum in order to be detected. Further requirements are that the chromophore needs to be incorporated into a monomeric genetically tagged probe, with a low quantum yield and high extinction coefficient so that the fluorescence does not overpower the Raman detectors but can attenuate photons with high efficiency. The furthest red-shifted monomeric near-IR FP that has been published is mRhubarb720, with reported characteristics of 701 nm λmax, 6.46% QY and 94,941 $M^{-1}cm^{-1}$ (Rogers et al. 2019). This gives us the best possibility at observing of epr-RS with the current microscope set-up available to us at Cardiff University. One of the most widespread red fluorescing proteins is mCherry, with a 587 nm λmax, 22% QY and a reported 52,000 $M^{-1}cm^{-1}$ (Shaner et al. 2004). The absorption maxima and QY are sub-optimal for epr-RS using our set-up, however due to its popularity in cell biology, it would be important to include this FP in this study as a comparison to the near-IR FPs.

This microscopy method was demonstrated with commercial dyes that absorb in the near-IR, particularly ATTO740, with photostable detection and super-multiplexing attributable to the sharp spectral peaks observed – producing up to 24 resolvable probes (Wei et al. 2017). The limitation with the commercial dyes is synonymous with fluorescence microscopy, where genetic incorporation with targets is not possible with dyes. Therefore, the need to employ genetically encoded proteins would mitigate this restriction, with the aim of accurate spatial and temporal detection within live-cells, which is currently unachievable with Raman-active dyes.

Near-IR FPs are a subclass of fluorescent probes engineered primarily for deep tissue imaging. They offer enhanced tissue penetration due to reduced absorption and scattering,

coupled with minimal phototoxicity, as the lower energy and longer wavelengths required for excitation causes less damage to the cells. The majority of these FPs utilise an extended conjugated double bonded cofactor, biliverdin XI-alpha (BV), which is reported to be autocatalytically incorporated after translation of protein. Therefore, this protein structure-function relationship will be critical in understanding the optimal parameters of epr-RS on what bonds to vibrationally detect. Since, the BV compound is responsible for the molecular electronic transitions, the protein scaffold surrounding the cofactor will play a crucial role in modulating the spectral properties of the FP. Modifying the protein complex will alter the spectroscopic properties of the FP itself and it will have an impact on the enhanced Raman scattering upon coherent nonlinear Raman excitation using near-IR lasers. Therefore, it is critical to understand and determine the BV binding cavity environment so that the rationale in what mutations to utilise is guided and structurally viable.

Most near-IR FPs, were engineered from a subset of phytochromes, which are crucial environmental sensors in various organisms. Phytochromes typically form a three-domain structure with the PAS (per-ARNT-Sim), GAF (cGMP phosphodiesterase/adenylate cyclase/FhIA) and PHY (phytochrome-specific) (Auldridge and Forest 2011). While phytochromes found in most bacteria use BV as a chromophore, with an absorption maximum of ~700 nm, cyanobacteria and plants utilise alternative tetrapyrroles such as phycocyanobilin or phytochromobilin, which absorb at ~630 nm (Hughes et al. 1997; Franklin and Whitelam 2005). Consequently, BV-binding FPs are the ideal focus when developing near-IR Raman probes due to their relatively red-shifted absorption.

Native bacterial phytochromes (BphPs) have a distinct structural feature where BV attachment site is located in the PAS domain, facilitated by a cysteine residue that forms a thioester bond. As previously mentioned, BV incorporation is an autocatalytic, post-translational modification, that follows the catalysis of heme via the bilin biosynthesis pathway (Wegele et al. 2004; Dammeyer and Frankenberg-Dinkel 2008). The enzyme responsible for the catalysis, heme oxygenase, is essential for protein maturation as it catalyses the regioselective cleavage of heme to produce BV (Wegele et al. 2004).

In this study, various BphPs-dervived proteins were considered; however, the focus was on the furthest red-shifted, monomeric FPs for their suitability in Raman applications. These included mRhubarb720, derived from RpBphP2 family of phytochromes, as well as enhanced-miRFP670 (emiRFP670) and miRFP670nano3, both derived from RpBphP1

phytochromes. emiRFP670 was selected for spectroscopic analysis due to its close structural similarity to miRFP670, one of the first near-IR FPs. In miRFP670, the N-terminal domain was found to promote protein degradation via a reported signalling pathway; removing this domain led to the creation of an emiRFP670 (Matlashov et al. 2020). The miRFP670nano3 was another candidate, notable for being half the size of both mRhubarb720 and emiRFP670, with a molecular weight at ~17 kDa (Oliinyk et al. 2022). Smaller detectable probes are generally more advantageous due to the reduced likelihood of steric clashes when tagging biomolecular targets.

One of the target proteins, mRhubarb720, had no published crystal structure available and given that the chromophore is a non-proteinous component, modelling approaches such as AlphaFold will not provide us with such critical information. Thus, to understand the function of this near-IR FP, it was required to elucidate the molecular structure to identify the underlying mechanisms of the protein. There are several methods available for determining molecular structures. However, due to the small size of crystals that were produced, it was decided to employ a fairly novel technique known as micro-focusing. As the name suggests, the method uses highly focused X-ray beams to analyse very small crystal samples, which has been used to determine protein structures that could not be crystallised to sufficient sizes for the traditional techniques (Ramakrishnan 2015). By concentrating the X-ray beam to a micro-scale spot size, researchers can achieve higher resolution data from crystals that are too small or imperfect for conventional X-ray diffraction methods (Mosselmans et al. 2009). There is also an ability to use multiple crystals as long as they arise from the same conditions, which is a constraint in standard X-ray crystallography. The data interpretation from the micro-focus beamline can be challenging due to the heterogenous and disordered samples that are submitted, therefore, refinement will be a critical step in determining the structures. For this reason, the objective was to integrate two refinement strategies: Alphafold2 and REFMAC5.

Alphafold2, a well-documented protein modelling predictor, is trained via the EMBL-EBI protein structure database that contains over 214 million predicted protein structures (Tunyasuvunakool et al. 2021; Varadi et al. 2024). This tool is now embedded into the downstream processing pipelines, after collecting the diffraction data from the VMXm beamline at Diamond Lightsource. After conversion of diffraction data into electron density maps; the regions where atoms are probabilistically located, the amino acid sequence is

supplied into the Alphafold2 and an initial protein sequence is predicted (Xu et al. 2023). Once this data is processed, another refinement process, REFMAC5, is applied to the initial protein prediction in order to refine the fit of the Alphafold2 model to the crystallographic data. As part of the CCP4 suite, the refinement tool uses maximum likelihood and Bayesian statistics to ensure chemical and structural integrity to the models (Murshudov et al. 2011). Combining both strategies will allow to overcome the challenges in interpretating the data generated from heterogenous samples via the micro focussing beamline.

This chapter aims to characterise and spectroscopically analyse the selected far-red/near-IR FPs to identify the most promising genetically encoded Raman-active probe for cell imaging applications. Subsequent structure determination of mRhubarb720 will follow via micro focussing crystallography to identify key residues within the BV-binding pocket.

3.2 Results and discussion

3.2.1 Biochemical analysis of WT near-IR FPs and mCherry.

The mRhubarb720 plasmid (ID: #141201) and mCherryWT (ID: #176016) was acquired via Addgene. However, the emiRFP670 and miRFP670nano3 plasmids had to be designed with Heme oxygenase within the same vector in order to produce the FP with BV catalysis and incorporation (Figure 3.1.a). These plasmids constructed by TwistBiosciences, using their pET28+ expression vector for bacterial overexpression in *E.coli* BL21's. All FPs contained a N-terminal hexahistidine-tag for purification as described in the Methods (Section 2.2.6).

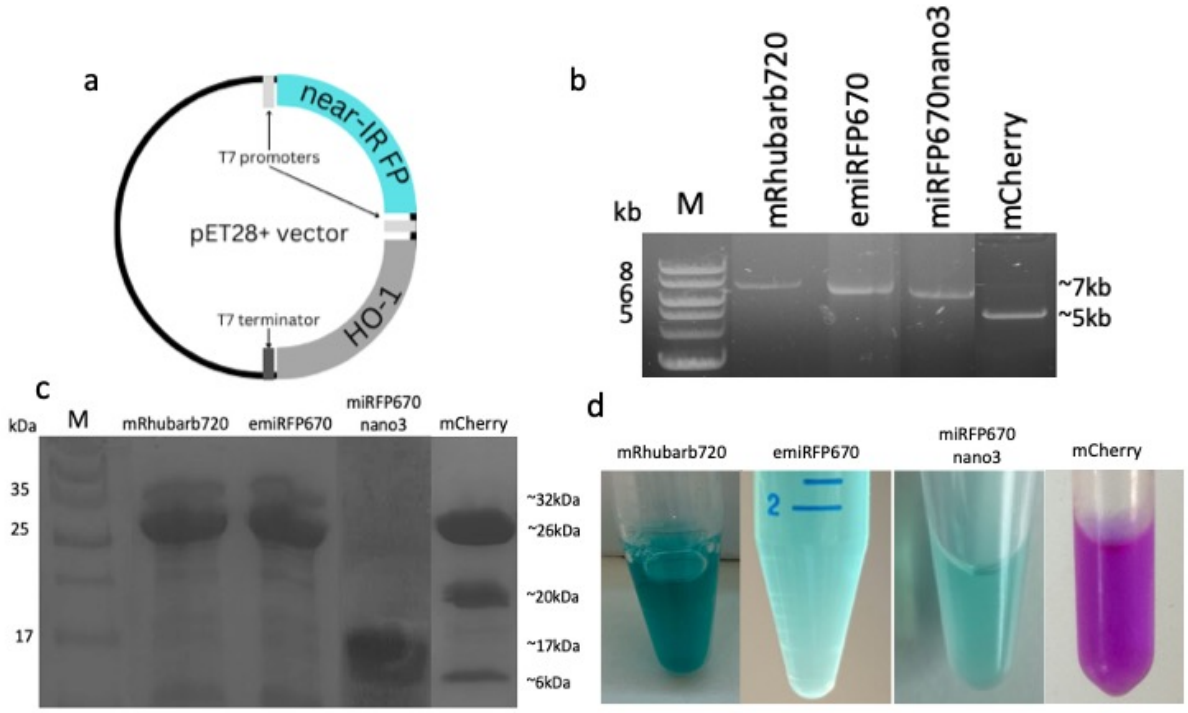


Figure 3.1: Design, expression, and purification of near-IR FPs and mCherry. (a) Plasmid design schematic of all near-IR FPs. (b) Agarose gel-electrophoresis. (c) SDS-PAGE after purification. (d) Post-purification image of near-IR FPs displaying the colour of BV-binding proteins as well as mCherryWT.

The template plasmids ordered from Addgene/TwistBiosciences, were transformed into a bacterial cloning strain, *E.coli* TOP10. The agarose gel-electrophoresis confirmed that each of the near-IR FP vector sizes was ~7kb, which was the approximate size of the ordered plasmids (Figure 3.1.b). The mCherry vector was also confirmed to be the correct size at ~5kb. The proteins were successfully produced in significant quantities via the overexpression in *E.coli*, induced using 500 µM IPTG for the pET vectors whereas the mCherry pBAD system was induced via 1% arabinose. After 18 hours of incubation at 22-25°C (near-IR FPs/mCherry) and cell lysis, the recombinant protein was purified by nickel affinity chromatography and size exclusion chromatography (Methods section 2.2.7). This two-step process ensured the elimination of all other proteins present in the *E.coli* as well as the residual contaminants, including imidazole and NaCl.

The purified protein was subject to an SDS-PAGE to confirm purity of the samples (Figure 3.1c). As anticipated, the mRhubarb720 and emiRFP670 displayed bands around 32kDa and miRFP670nano3 was roughly half the size, measuring at around 17kDa. However,

mCherry displayed three bands, one at the expected size of 26 kDa and two bands at 20 kDa and 6 kDa respectively, the two smaller bands are indicative of a cleavage at the chromophore when undergoing denaturation, which is common in beta-barrel FPs (Gross et al. 2000). The concentrated proteins were pooled where a notable turquoise-blue colour is observed for the near-IR FPs, characteristic of all BV-binding proteins as there is a gap in absorption that reflects blue light.

3.2.2 Spectral characterisation of WT near-IR FPs

To understand the suitability of the FPs for developing a Raman-active tag, the spectral properties needed to be analysed. There is a direct correlation in absorption maximums and the intensity of electronic excitation for the Raman spectroscopy due to the pre-resonant excitation conditions previously described (Chapter 1.1.3).

It is also essential to analyse the spectral properties of free-BV as this cofactor will be present within eukaryotic cells. This is important as the aim to genetically-encode the Raman-active probe via a protein scaffold and therefore, if free-BV were to be detected, then the probes will be ineffective as background signal will overpower the desired targets. To determine whether free-BV would have an impact on cell imaging, the UV-Vis absorption spectra was measured along with the three near-IR FPs (Figure 3.2.a)

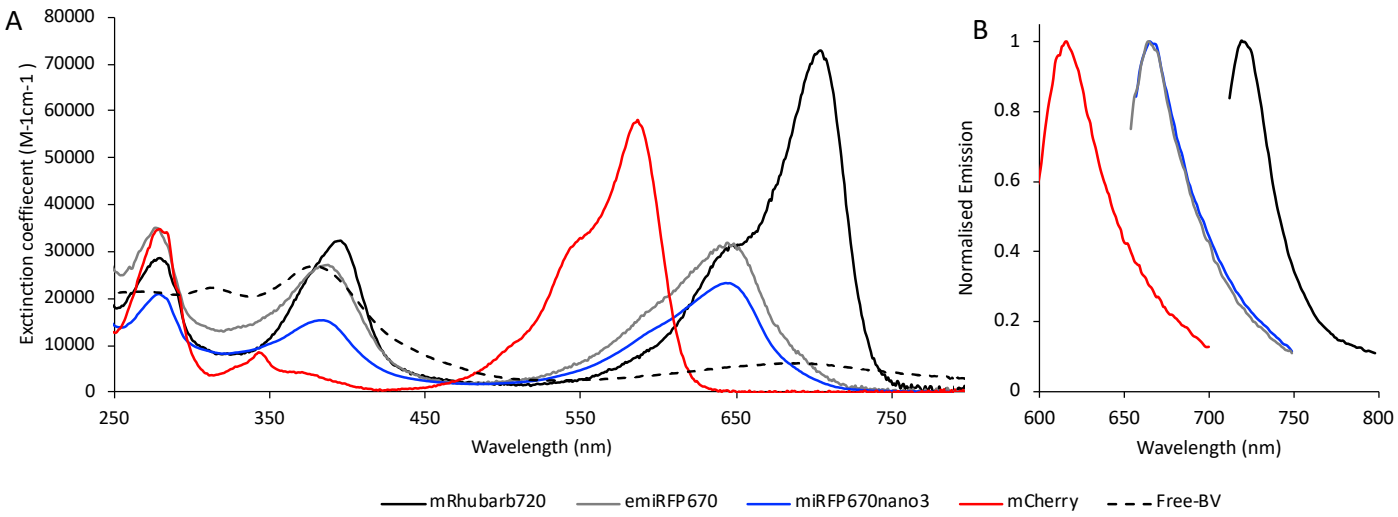


Figure 3.2: Spectral properties of the mCherry, near-IR FPs and free-BV. (A) Absorbance spectra of each FP with the comparison to free-BV for the near-IR FP. (B) Emission spectrum profiles of the FPs, free-BV did not show fluorescent properties. The excitation wavelengths was corresponded to the absorbance maximums. mRhubarb720 (black line, excited at 703nm), emiRFP670 (blue line, excited

at 644nm), miRFP670nano3 (grey line, excited at 645nm), mCherry (red line, excited at 587nm), free-BV (black dotted-line).

For free BV, there is no significant absorbance, including at the higher near-IR wavelengths. This is particularly important as no absorbance in the near-IR will equate to no fluorescence or excitation in the pre-resonant excitation window, due to the Pump laser wavelength being at 820 nm. Therefore, free-BV will not be detected in the *in vitro* Raman spectroscopy or the cell imaging. mRhubarb720 displayed the most promising spectral characteristics due to its high molar extinction coefficient and red-shifted absorbance and fluorescence. The emiRFP670 and miRFP670nano3 was expected to emit at 670nm and to absorb at 644nm and 645nm, respectively. One characteristic observed from the absorption is the peak residing at ~380-400nm which can be described as the BV peak, only present within our measurements when BV is present. The mCherry was the most blue-shifted FP in our collection, with an absorption maxima of 587nm and emission at 610nm. However, the previously reported molar extinction coefficients for each FP were inflated compared to our findings (Table 3.1).

Table 3.1: Spectral properties determined in this study compared to previously reported spectral properties. (a) Values reported in FPbase (https://www.fpbase.org), (b) Reported by Rogers *et al.* (2019), (c) Matlashov *et al* (2020), (d) Oliinyk e*t al.* (2022), (e) Shaner *et al* (2004).

Near-IR FP	Current findings		Previously reported [a]		Calculated	Measured
	λ_{max} (nm)	ε ($mM^{-1}cm^{-1}$)	λ_{max} (nm)	ε ($mM^{-1}cm^{-1}$)	ε at 280nm ($mM^{-1}cm^{-1}$)	ε at 280nm ($mM^{-1}cm^{-1}$)
mRhubarb720	703	72.5	701[b]	94.9[b]	21.19	21.51
emIRFP670	644	29.1	642[c]	87.4[c]	30.94	30.95
miRFP670nano3	645	31.9	645[d]	129.0[d]	17.42	17.44
mCherry	587	58.1	587[e]	72.0[e]	34.38	34.77

In this study, protein concentrations were independently validated using BioRAD DC Assay, and the molar extinction coefficients were then determined using the measured protein concentrations. To confirm the accuracy of our molar extinction coefficient, theoretical molar extinction coefficients at 280 nm were calculated and compared to the measured values, as shown in Table 3.1. The previously reported values were confirmed to be overinflated because, if they were accurate, the 280 nm molar extinction coefficients would be significantly higher. This conclusion is based on the expectation that the ratio

between the 280 nm and the absorption maximum extinction coefficients would remain consistent with that of a purified protein, which was not observed in this case.

Regardless of the previously reported values, these findings support the rationale for selecting mRhubarb720 over the other FPs. Its relatively red-shifted absorbance and significantly higher molar extinction coefficient make it the most suitable choice as a genetically encoded Raman active probe for the in-house epr-RS microscope set-up at Cardiff University.

3.2.3 Crystal formation, microfocusing and refinement to determine the structure of mRhubarb720

The literature surrounding RpBphP2-derived near-IR FPs is limited when discussing the structure of BV-binding proteins, as the only solved crystal structure for this lineage is the original RpBphP2 phytochrome (PDB: 4E04) (Bellini and Papiz 2012) For example, the most red-shifted BV-binding FP, mRhubarb720, there is no structural information available, despite there being 16 mutations compared to its nearest structurally determined homologue, RpBphP2. Most of these mutations reside in dimer interface of RpBphP2 with the sole function of monomerising the protein. Given that we propose to use this protein as a Raman-active probe, there is a need to understand the molecular configuration that defines functions, especially around the BV-binding pocket, in order to understand what the key mutations that give rise to the spectral properties. There are other near-IR BV-binding FPs (derived from RpBphP1) with a determined structure, namely miRFP670 (PDB: 5VIV) and miRFP670nano3 (PDB: 7LSC) (Baloban et al. 2017). Most of the solved structures of near-IR BV-binding FPs are derived from the RpBphP1 bacterial phytochrome photoreceptors.

Crystallography requires substantial quantities of purified protein. To achieve this, we expressed, purified, and concentrated the mRhubarb720 to 10 mg/mL. As previously mentioned, the purification process involved two chromatography steps: nickel affinity, followed by size-exclusion, as detailed in Methods section 2.2.7. The size-exclusion chromatography also provided valuable quantitative data such as accurate molecular weight determination when using calibration curves (Figure 3.3). Furthermore, there was a singular peak that highlights the fact that mRhubarb720 is truly monomeric.

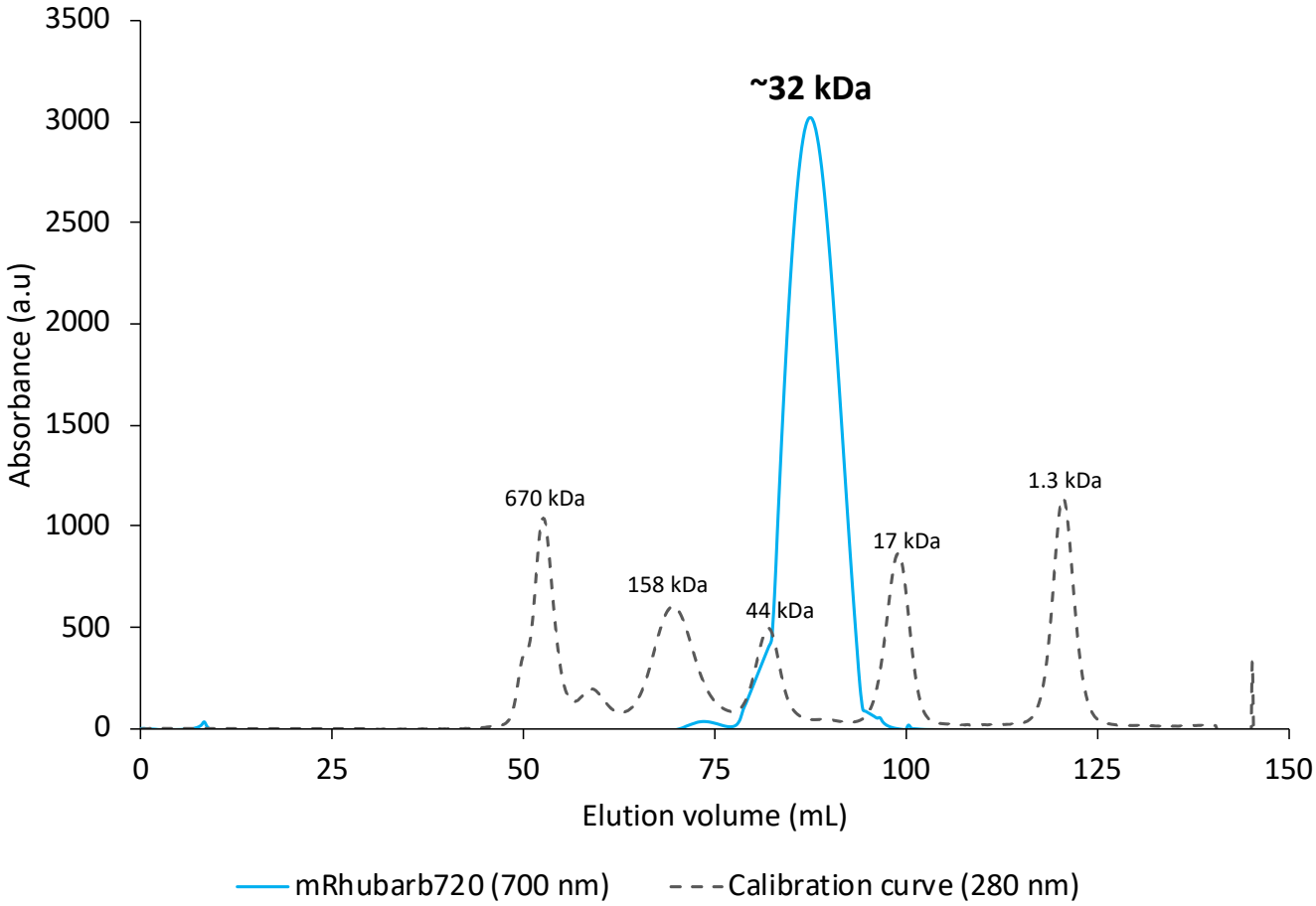
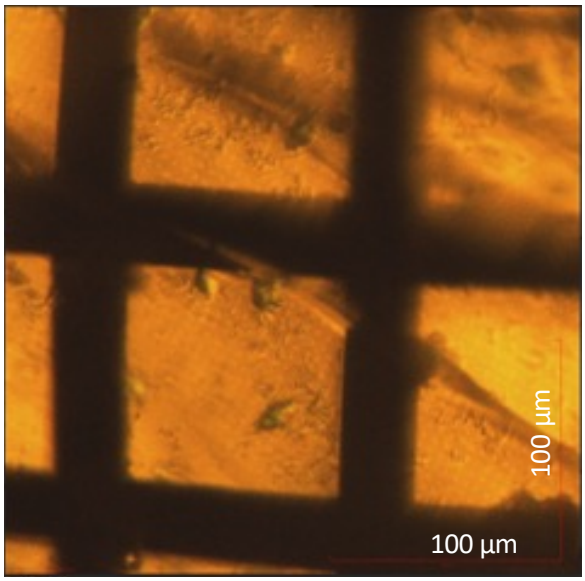
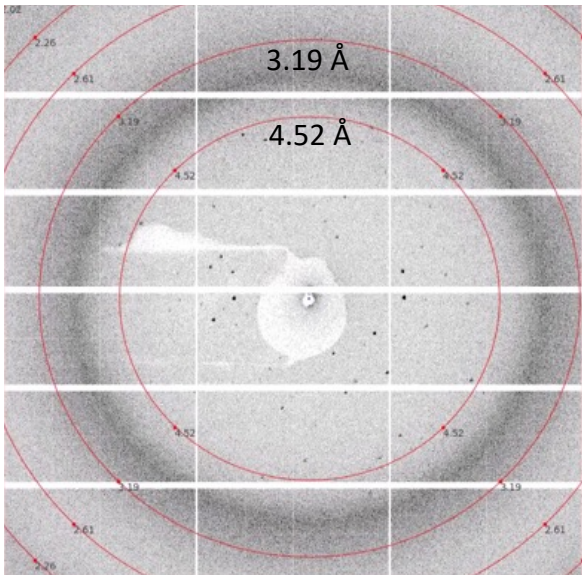


Figure 3.3: Size-exclusion chromatography of mRhubarb720. Elution from nickel affinity purification was subject to SEC. The mRhubarb720 (blue line) was measured at 700 nm absorbance, whereas the calibration curve (grey dotted line) was measured at 280 nm. The S200 HiLoad 16-60 calibration curve was purchased from BioRAD for the Superdex 200 SEC column.

After the purification process, the protein was pooled from 76-95 mL and concentrated, so that mRhubarb720 was subjected to multiple crystallisation trails. However, the PACT premier screen was the only plate that produced crystals of any size (Figure 3.4). The crystals we used appeared in the well containing 0.1M MMT, 25% w/v PEG 1500 at pH 5.0, with 21 days of incubation at 20°C. Using these crystal conditions, the size of the crystals was limited to ~20 µm, therefore, was it was decided to use micro-focussing crystallography using the VMXm beamline at Diamond Lightsource. There was a decision made to not use standard beamlines or macromolecular crystallography techniques as this would obliterate the crystals due to the larger beam size. Micro-focussing yielded diffraction data with a resolution of ~3.3 Å, also observed in Figure 3.4.



VMXm beamline statistics	
Energy/ wavelength	21.326 keV, 0.5814 Å
Exposure time	0.2 s
Oscillation width	0.1°
Transmission	100%
Beam size	3 x 3 µm



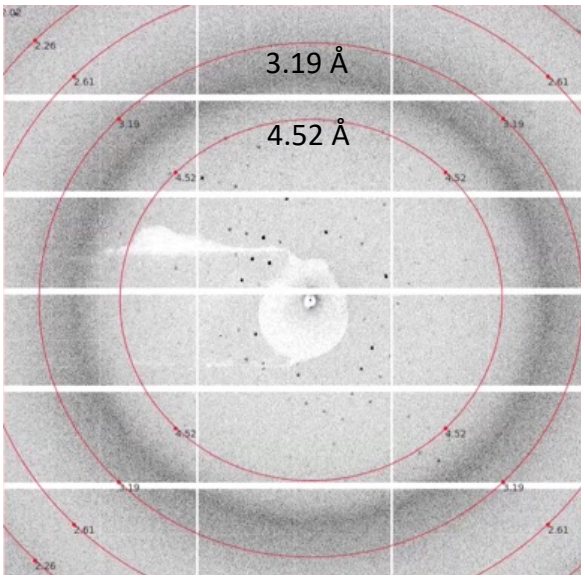


Figure 3.4: Crystal size, beamline statistics and typical diffraction images. The crystals submitted to micro-focusing can be observed in the top-left image (scale bar 100 x 100 µm). Small segments of data collected from each crystal using the VMXm beamline stats (top-right), which produced a number of diffraction patterns, typical diffraction images can be observed at the bottom.

Although diffraction patterns were observed, resolution was not optimal as diffraction spots were only visible up to 3.3 Å, between the resolution rings of 3.19-4.52 Å. Full diffraction statistics from the micro-focusing data collection can be observed in Table 3.2. There were 21,465 total reflections which gave 4.2x coverage of the crystal, denoted by the multiplicity. There was 95.1% completeness and the correlation coefficient (CC1/2) was equal to 0.971, suggesting that the data is just about adequate for refinement.

Table 3.2: Statistics from Micro-focusing X-ray diffraction structure determination of mRhubarb720. Three crystals were merged to produce this data set. Coordinate Estimated Standard Uncertainty in (Å), calculated based on maximum likelihood statistics. Figures in brackets refer to outer resolution shell.

Crystal Data – PDB: 9F03 **Crystallisation Conditions**	
Space group	P 4_1 2_1 2
a (Å)	58.85
b (Å)	58.85
c (Å)	185.55
α,β,γ (°)	90.0, 90.0, 90.0
Resolution (Å)	3.30 – 27.41
Outer shell	3.30 – 3.57
R-merge (%)	33.4 (164.5)
R-pim (%)	20.2 (106.7)
R-meas (%)	37.2 (179.3)
CC1/2	0.971 (0.315)
I / σ(I)	4.8 (0.7)
Completeness (%)	95.1 (90.7)
Multiplicity	4.2 (2.9)
Total Measurements	21,465 (2,786)
Unique Reflections	5,072 (953)
Wilson B-factor(Å^2)	77.6
Refinement Statistics	
Refined atoms	2,382
Protein atoms	2,339
Non-protein atoms	43
Water molecules	0
R-work reflections	4,779
R-free reflections	261
R-work/R-free (%)	22.5 / 30.9
rms deviations (ML target in brackets)	
Bond lengths (Å)	0.008 (0.013)
Bond Angles (°)	1.448 (1.648)
[1]Coordinate error	0.892
Mean B value (Å^2)	86.0
Ramachandran Statistics (PDB Validation)	
Favoured/allowed/Outliers	230 / 46 / 28
%	75.7 / 15.1 / 9.2

The crystal structure was defined by the space group $P4_12_12$, and was handled at Diamond Light Source using Alphafold2 in their embedded downstream processing pipeline which takes the protein sequence and applies homology modelling in order to improve structural integrity of the protein. Subsequent refinement using REFMAC5 was implemented

via CCP4 suite to allow various restraints on atomic displacement parameters to optimise the molecular structure. The R factor and R-free were at 22.5% and 30.9%, respectively, which infers that our data is reliable to a low resolution of 3.3 Å. When trying to improve our resolution with further refinement, these R values were not agreeable with proposed structure.

To improve the resolution, further crystal trails would need to be deployed. To get the microcrystals used for the structure determination via micro-focussing, there were six crystal screens with 96 different conditions within each well. Only three wells produced any form of crystal which suggests the optimisations plate would need to be implemented on with conditions used to produce micro-crystals. Therefore, varying concentrations and pH's would need to be applied to MMT and PEG 1500. The goal would be to generate large crystals so that high resolution X-ray crystallography could be used to determine the structure.

3.2.4 Structural insights of mRhubarb720 and BV cofactor binding environment

The aims of determining the mRhubarb720 structure were two-fold: first, to understand and map the BV-environment, and second, to explore how residues surrounding the cofactor modulated the BV, and consequently, the spectral properties of the FP. The structure is monomeric and comprises of two subdomains, PAS and GAF (Figure 3.5.a-b). Previous literature indicates that the BV is covalently bound to Cys15, forming a thioester bond, an attachment suggested to occur through an autocatalytic mechanism (Rogers et al. 2019). In other RpBphP2-derived near-IR FPs, this covalent attachment typically originates from the PAS domain. However, protein engineering allows for attachments to occur in the GAF domain, provided that the sulphur atom of the cysteine is within 1.8 Å of the BV's covalent attachment site. Our crystallography data confirms the presence of this thioester bond, as observed between Cys15 to a carbon atom bound to the A-ring of BV, with a bond length measured at 1.67 Å (Figure 3.5.c).

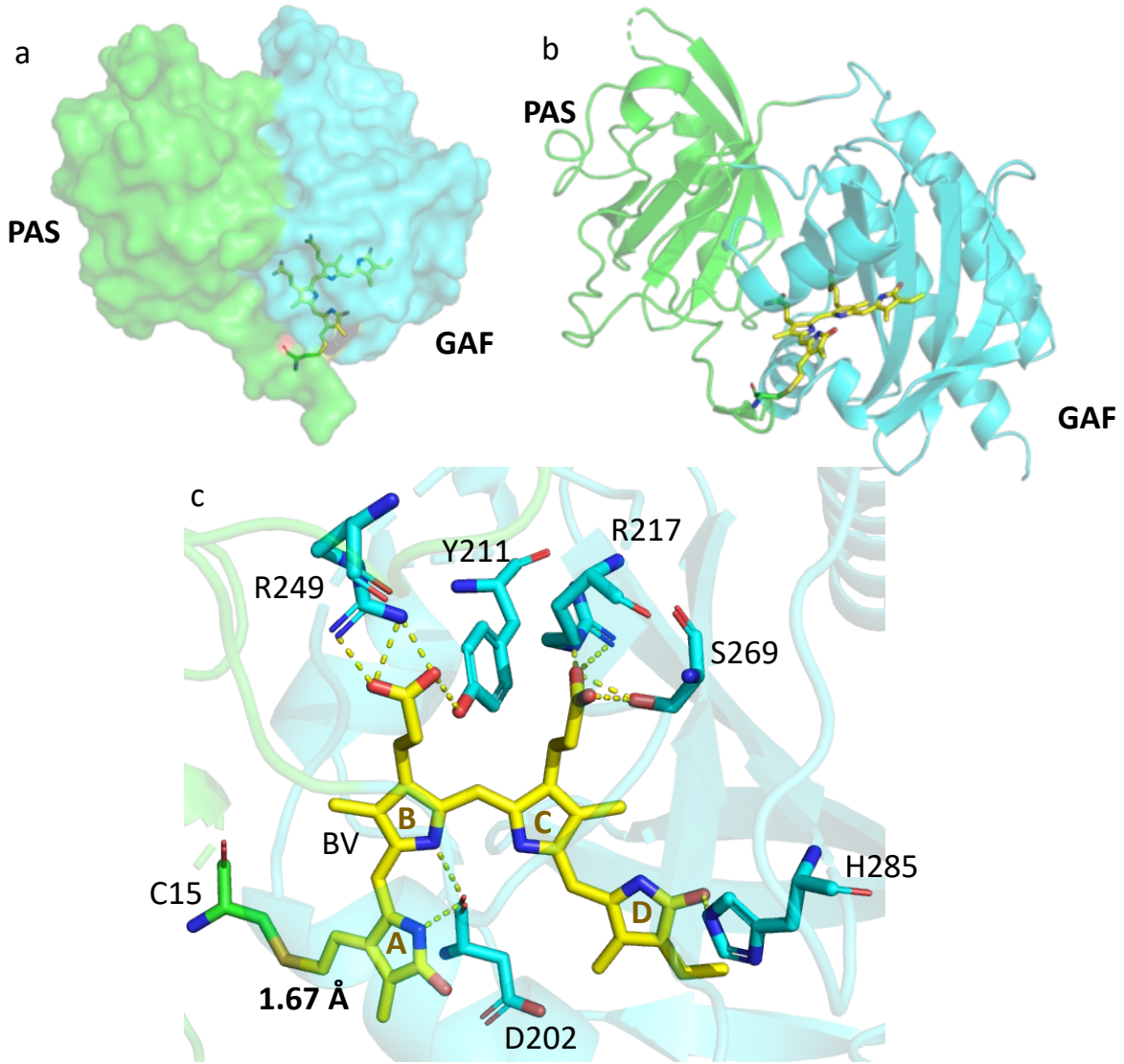


Figure 3.5: Final mRhubarb720 model. (a) Surface configuration in respect to BV (yellow) and (b) secondary structure of mRhubarb720. Green represents the PAS domain whereas cyan shows the GAF domain. (c) BV cofactor environment with measurement highlighted between Cys15 and the carbon binding site on BV, with annotated residues that show possible polar contacts between BV and protein scaffold. The yellow dashed lines highlight where the possible polar networks would be, determined via PyMOL find function. The rings in the BV have also been annotated to specify the nomenclature of each ring.

There are two distinct domains in mRhubarb, as highlighted in Figure 3.5.a, with Cys15 residing within the PAS domain. The first 12 residues of mRhubarb720 remain unresolved due to the inherent flexibility of the N-terminus. Additionally, there are several potential polar contacts between the BV chromophore and the protein scaffold, with the involved residues and bonds annotated and highlighted in Figure 3.5.c. These polar contacts

were identified using PyMOL's "find" function, which identifies contacts between polar atoms within a 3.5 Å distance cut-off; thus, the accuracy of these findings may be limited.

The crystallography also provided crucial insights into the conformation of the BV, which is critical to the spectral properties of the FP. This confirmation was inferred from the electron density observed around the BV (Figure 3.6.a). The D-ring of BV is known to flip due to photo-isomerisation and can accommodate two states (Pr and Pfr) (Yang et al. 2009). The crystallography data shows that the D-ring assumes a Pr state, suggesting that mRhubarb720 has a Pr ground state configuration, due to H285 in Figure 3.5.c, similar to the other near-IR FPs discussed in Chapter 1.2.2.2. Additionally, a notable observation was the lack of electron density around the supposed thioester bond. This could indicate that the bond is flexible and prone to movement, or it may suggest the presence of some unbound BV in the population of mRhubarb720 crystals.

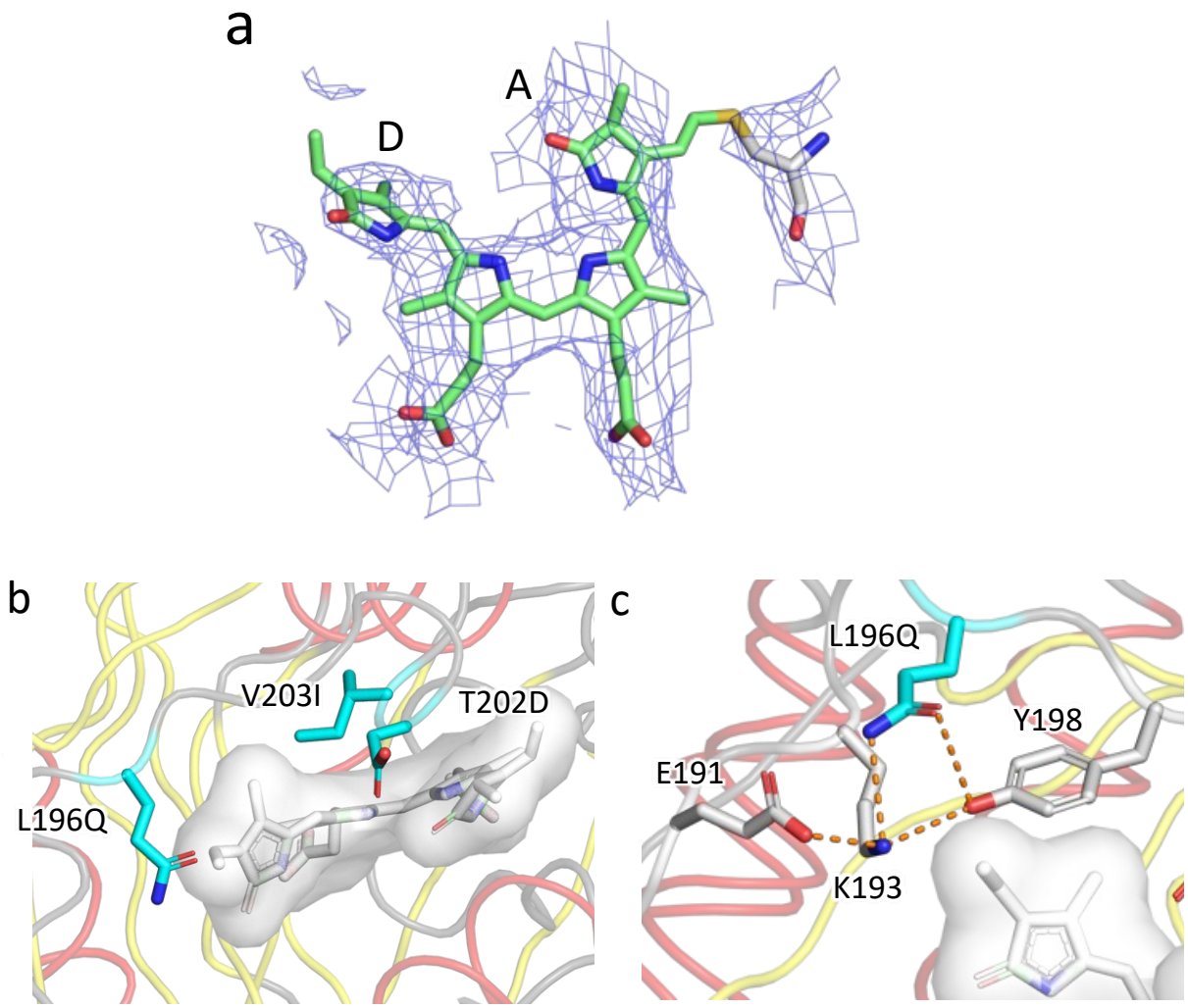


Figure 3.6: BV configuration and environment. (a) Electron density of BV cofactor and bound Cys15. (b) Mutations responsible for generating redshifted spectral properties. (c) Potential polar interaction network between L196Q and other residues.

The positioning of key mutations in generating mRhubarb720 from iRFP have been highlighted in Figure 3.6.b. The placement of these mutations highlights the close structure-function relationship within mRhubarb720 as these three mutations are responsible for generating the most red-shifted monomeric near-IR FP. These key mutations give rise to the spectral properties of mRhubarb720, and these will be the targets for inciting further spectral changes in respect to improving the Raman efficiency. Mutations T202D and V203I have been reversed to their original residues in RpBphP2, rather than what was contained in the iRFP sequences, due to their reported redshift (Rogers et al. 2019). The L196Q mutation is positioned so a polar network could hypothetically give rise to the shifts in absorption and

emission, especially as these mutations are placed close to the D-ring (Figure 3.6.c). This could be through H-bonding, electrostatic interactions (specifically with Lys193) or general solvent effects if H_2O is present, therefore, Q196 will be a key target to incur spectral shifts.

3.2.5 Comparing mRhubarb720 with other BV-binding FPs

As previously stated, the most closely related published BV-binding protein crystal structure currently available to mRhubarb720 is RpBphP2 (Figure 3.7.a). This protein's original function involves regulating the biosynthesis of light harvesting complex LH4, with photoconversion behaviour that adopts the dark-adapted Pr state that photoconverts to Pfr state upon excitation (Kumarapperuma et al. 2023). RpBphP2 is homodimeric, and many mutations used to engineer near-IR FPs focussed on generating a stable monomeric variant. mRhubarb720 and RpBphP2 share 95% sequence identity, with 16 differing residues. Structural alignment using PyMOL resulted in an RSMD of 0.418 Å, indicating that both structures are significantly comparable (Figure 3.7.b).

Within the BV-binding pocket, only three mutations are present: Y193K, L196Q and F198Y, which are positioned near the D-ring (Figure 3.7.c). These mutations likely contribute to the formation of a polar interaction network, which can significantly influence the spectral properties of the protein. For instance, Y193K could introduce a positive charge that interacts with the D-ring, which could stabilise the Pr state and/or alter the energy barriers associated with the photoconversion process. L196Q may present H-bonding capabilities within the binding pocket, affecting the conformation of the BV. Finally, F198Y introduces a hydroxyl group, which could form another H-bond around the D-ring of BV. These combined effects may contribute to the spectral properties observed in mRhubarb720 compared to its non-fluorescent parent protein, RpBphP2.

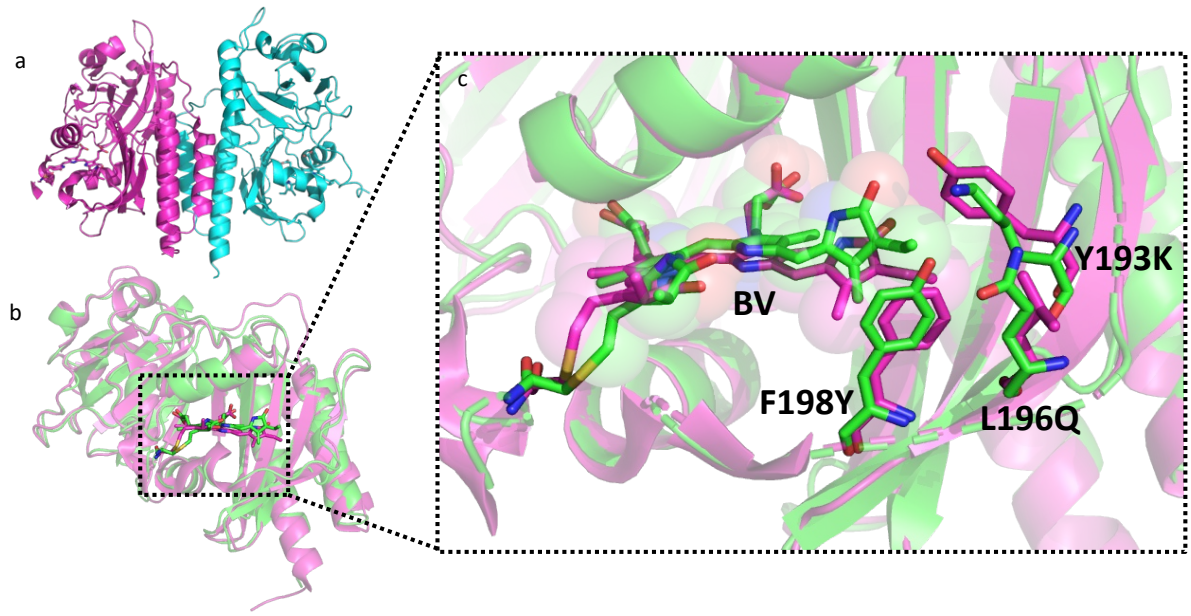


Figure 3.7: Structural comparison of mRhubarb720 (green) and RpBphP2 (magenta and cyan). (a) Cartoon representation of the crystal structure of RpBphP2 (PDB: 4E04). (b) Alignment of mRhubarb720 to one subunit of RpBphP2. (c) Structural comparisons of the BV-binding pocket with relevant mutations.

The D-ring in mRhubarb720's BV is observed to exhibit a great degree of rotation compared to the BV in RpBphP2. This conformation is suboptimal because more planar chromophores promote symmetrical electron distribution, resulting in a higher dipole moment. This, in turn, leads to more efficient photonic transfer, which correlates with higher molar extinction coefficients and enhanced emission.

Due to the limited availability of solved crystal structures within the RpBphP2 family of near-IR FPs, the focus of structural comparison shifted to RpBphP1-derived FPs (Figure 3.8). Specifically, emiRFP670 and miRFP670nano3, which have previously undergone spectral analysis (Section 3.2.2). The reasons for choosing these proteins were two-fold; firstly, emiRFP670 was very photostable in comparison to a host of other near-IR FPs, with a half-life time of 450 seconds, due to the removal of the N-termini (Zhang et al. 2023). For reference, the predecessor miRFP670 has a half-life of 155 seconds. Secondly, there was a FP that reportedly produced very similar emission profiles to emiRFP670 but was roughly half the size, hence the name, miRFP670nano3 (Oliinyk et al. 2022).

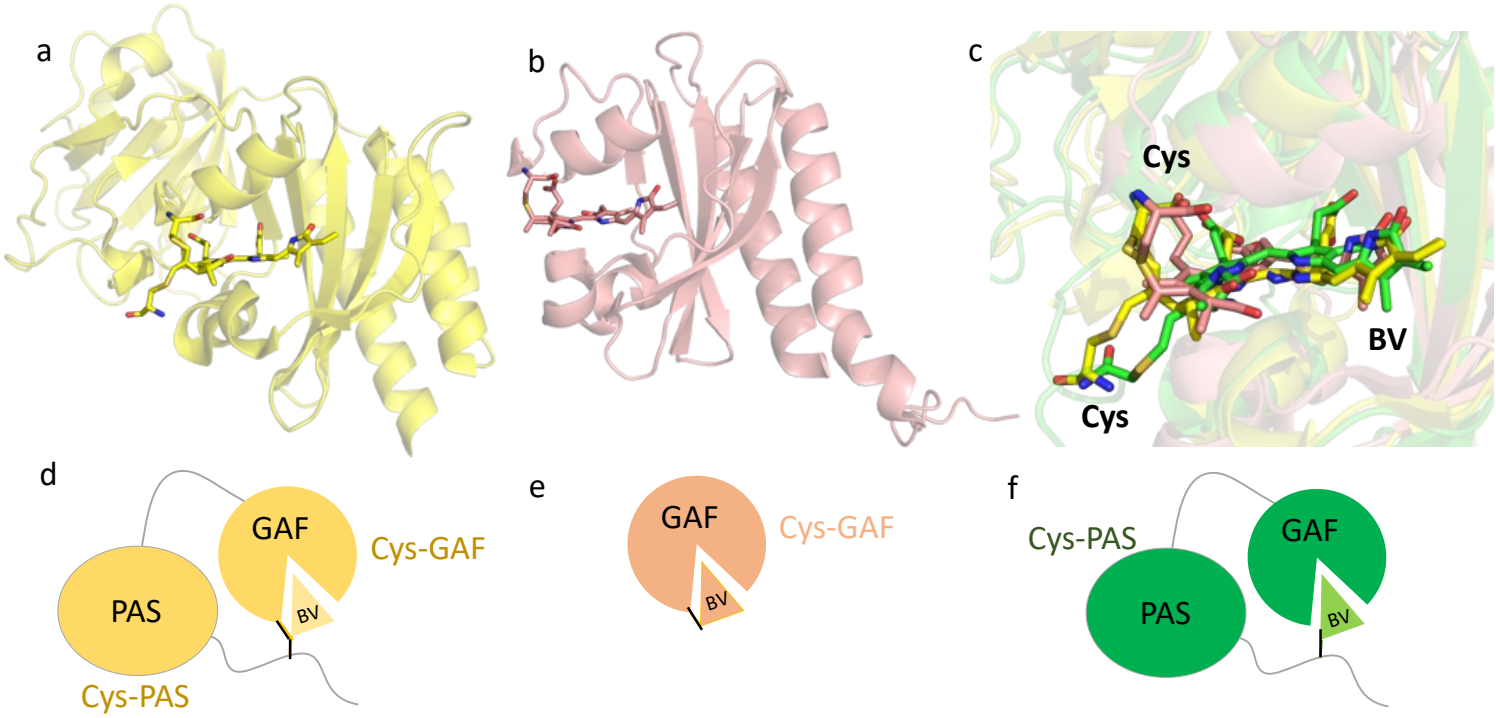


Figure 3.8: Structural comparison of near-IR FPs with BV attachment configuration schematics. (a) miRFP670 (yellow, PDB: 5VIV). (b) miRFP670nano3 (pink, PDB: 7LSC). (c) BV binding pocket comparisons of mRhubarb720 (green), miRFP670 and miRFP670nano3. (d) emiRFP670, (e) miRFP670nano3 and (f) mRhubarb720 BV-binding complex diagrams. Black lines show possible covalent attachments.

The emiRFP670 has a sequence similarity of 38% compared to mRhubarb720, as well as a RMSD value of 1.061 Å when aligned. This suggests a notable structural similarity despite the low sequence identity, indicating functional conservation of key structural elements, particularly within the BV-binding pocket. The RSMD value reflects that, although there are differences between the proteins, they still adopt a similar 3D structure, which is critical for their near-IR spectral properties. emiRFP670 contains two cysteine residues capable of forming the thioester bond that binds the BV. This covalent linkage can occur either via the PAS domain (Cys20) or the GAF domain (Cys253) (Figure 3.8.d). The Cys20 thioester bond is analogous to the BV-binding configuration observed in mRhubarb720. It is unlikely that both linkages can occur simultaneously, suggesting that emiRFP670 may exist as a mixed population of protein with either Cys20 or Cys253 covalent linked to BV. Only one key residue is near the D-ring of BV differs in either protein, K193 in mRhubarb720 is a

tyrosine in emiRFP670, which is analogous to the RpBphP2 configuration. This suggests that this lysine may contribute to the red-shifted properties of mRhubarb720.

In contrast, the sequence similarity between mRhubarb720 and miRFP670nano3 is lower, at 22%, with a significantly higher RMSD of 11.0 Å. These values suggest poor alignment and substantial differences between the two proteins, which is expected given that miRFP670nano3 contains only the GAF domain. The thioester bond is facilitated by Cys86 in the GAF domain on the BV, which is analogous to the GAF domain linkage observed in emiRFP670. When comparing the RpBphP1-derived near-IR FPs, there is a 25% sequence similarity, with an RMSD of 1.564 Å. The low sequence similarity is expected, as miRFP670nano3 lacks the PAS domain. However, the RMSD indicates that their GAF domains are relatively well conserved, which aligns with their nearly identical spectral properties.

3.3 Conclusion

Given the spectral properties of each FP, mRhubarb720 emerged as the optimal candidate for cell imaging, due to its significantly higher molar extinction coefficient, as well as its red-shifted absorption and emission. While all FPs will undergo Raman spectroscopic analysis using SRS/CARS, those with the most red-shifted and the highest light attenuation are most likely to exhibit Raman intensity amplification. This is due to the nature of the microscope set up and the pre-resonant excitation conditions set by the Pump beam's wavelength at 820 nm. This characteristic is crucial for effective detection, as pre-resonant coherent Raman spectroscopy relies on coherent non-linear enhancement, where the Pump and Stokes laser fields constructively interfere with absorbed photons within the molecule.

When applying pre-resonant coherent Raman spectroscopy to the BV-binding probes, free-BV is unlikely to be detected due to the lack of absorptive properties. This suggests that during imaging, only the genetically-encoded protein scaffold containing BV will be subject to Raman amplification, enabling a specific detection of bonds within the BV chromophore.

A key result in this chapter was the determination of the mRhubarb720 structure, addressing the previous lack of RpBphP2-derived crystal structures for engineered near-IR FPs. Despite the resolution not being optimal, we successfully determined the structural environment of the BV cofactor, which is highly valuable for our objectives. This paves the

way for identifying and applying mutational targets to enhance the probes suitability for Raman spectroscopy, such as reducing BV flexibility and increasing the dipole moment to amplify the Raman scattering cross-section. Furthermore, we now have the capability to introduce high-depth *in-silico* analysis to further achieve our aims and objectives.

4 Electronic pre-resonant Raman spectroscopy of near-IR FPs and cell imaging of mRhubarb720.

4.1 Introduction

The combination of fluorescence microscopy and genetically encoded probes (i.e. fluorescent proteins; FPs) has revolutionised cell biology as it opens up the ability to observe spatiotemporal events such as location, production and interactions of specific protein targets *in situ*. First discovered in the 1960's, green fluorescent protein (GFP) has allowed for this visualisation and tracking of dynamic processes, crucial for advancements in cell biology, neuroscience, and medical diagnostics (Lichtman and Conchello 2005; Chudakov et al. 2010).

The contribution of FPs to the biological community has been vitally important, however, there are limitations to using FPs and fluorescent microscopy. Continued excitation can permanently and chemically damage FPs through photobleaching (Grigorenko et al. 2015). For example, FPs can be photobleached due to the presence of molecular oxygen and reactive oxygen species near the chromophore, leading to oxidative damage and degradation of the chromophore (Wiens et al. 2018). This can introduce significant challenges across various experimental contexts. For instance, in gene expression reporter assays, fluorescence intensity will not be proportional to expression if FPs are damaged. This degradation can also obstruct live-cell imaging, where continuous fluorescence is essential for monitoring dynamic processes. Additionally, high-content screening and quantitative imaging studies are similarly affected, as photobleaching can lead to inaccurate quantification and skewed data interpretation. Furthermore, there is reported cytotoxicity from reactive oxygen species, generated via over-excitation, that cause oxidative stress and induction of apoptosis (Liu et al. 1999; Ansari et al. 2016). There are also limitations with regards to the number of FPs (and thus the number of targets) that can be monitored at any one time as FP emission peaks are quite broad in respect to the visible spectrum. The visible spectrum is restricted to the 300-800 nm wavelength range and each emission peak will span roughly ~50-100 nm. Consequently, it is feasible to apply 5-7 different FPs to a sample before their emission spectra significantly overlap, making the colours indistinguishable (Orth et al. 2018). Increasing the number of FPs within a sample may also inadvertently

induce a fluorescence resonance energy transfer (FRET) phenomenon, where energy transfer between closely positioned fluorophores occurs, resulting in unintended interactions between signals (Sekar and Periasamy 2003). This can further blur the distinction between emission peaks, making it challenging to accurately resolve and interpret fluorescence data.

To overcome the above limitations, integrating new technologies that utilise FPs beyond the visible spectrum, such as near-IR FPs which operate outside the traditional UV-Vis range, can be beneficial. This approach can help reduce spectral overlap and minimises excitation energy, thereby reducing photobleaching. Raman microscopy is a potentially new approach which is based on Raman scattering, the inelastic scattering of photons when light interacts with vibrational excitations of matter. Due to this interaction, a small fraction of the incident photons is scattered with a shift in frequency corresponding to the vibrational frequency of specific bonds. The resulting peaks in the Raman scattering spectra provide detailed information about molecular vibrations and chemical compositions (Zhang et al. 2010). Raman scattering can be spatially resolved using a confocal excitation and detection, forming a Raman microscope (Chaichi et al. 2018).

The main limitation of Raman spectroscopy is the signal intensity, making it virtually impossible to detect individual bonds within a cellular environment without altering the methodology (Ember et al. 2017). While SERS setups can locally enhance fields to the point of single-molecule detection, they are generally unsuitable for 3D cell imaging applications. There are two variables that can be adjusted: the Raman scattering cross-section, that is limited to 10^{-29} cm^2 per vibrational mode, and/or the power of incident lasers. Caution is needed when increasing the laser power, as continuous excitation could potentially damage the sample. Therefore, the primary focus should be on increasing the Raman scattering cross-section. For perspective, assuming a 100 mW excitation light source and a photon energy of 2 eV, with a cross-section of 10^{-29} cm^2 and a focus size of 1.8 µm, a single molecule to emit 0.0001 scattered photons/second (Kneipp and Kneipp 2006). This would require an excitation time of over two hours to observe one photon per molecule, which can be considered an extremely low signal strength and would be damaging to the sample.

Electronic pre-resonant Raman Scattering (epr-RS) combines two concepts that should increase the Raman scattering cross-section by enhancing the Raman polarisability of the molecule, ideally with a dependency on the inverse square of the detuning factor,

(1/|Excitation energy – Electronic transition energy|)^2. The cross-section can be increased by multiple orders of magnitude by applying the electronic resonant Raman effect to the frequency of the incident lasers (Efremov et al. 2008). This requires an electronically active chemical group (e.g. a chromophore), containing a vibrational target/bond to amplify the signal. This coupling effect significantly promotes molecular vibrations observed in the electronic resonant Raman effect (Wei and Min 2018). The second concept, coherent nonlinear enhancement, involves two incident lasers (pump and Stokes), excited to specific frequencies that can cause constructive interference of the molecular vibrations, therefore, exponentially increasing the signal intensity for long range electronic interactions (Vinegoni et al. 2004).

The two primary methods for implementing epr-RS are Stimulated Raman Scattering (SRS) and Coherent anti-Stokes Raman Scattering (CARS), which enhance Raman signals to improve the sensitivity for detecting specific chemical bonds. Both methods use two lasers: a Pump beam and a Stokes beam. The interaction between the Pump and Stokes beams creates a vibrational coherence in the sample, resulting in coherently scattered signals. In SRS, this coherence produces either stimulated Raman gain (in the Stokes beam) or stimulated Raman loss (in the Pump beam; SRL), enhancing the Raman response (Evans and Xie 2008). These concepts have been demonstrated using a near-IR chemical dyes (ATTO700-740), where 24 resolvable probes were characterised within a sample, allowing for super-multiplexing properties (Wei et al. 2017). However, these probes were not genetically-encoded and targeted metabolic products instead of proteins. The utilisation of the super-multiplexing technique with FPs would allow for the specific targeting of proteins. This has the potential to revolutionise biological cell imaging by Raman microscopy in the same way as FPs greatly advanced fluorescence microscopy.

Due to the nature of the microscope setup, FPs need to be carefully selected so that the chemical bonds that are being detected are not overwhelmed by the fluorescence emitted from the electronically active chromophores. As mentioned, the Pump beam is fixed at 820 nm, and its purpose is to excite the chromophore within the FPs pre-resonant excitation condition (explained in Chapter 1.3.3). Therefore, the setup would be more responsive to the near-IR FPs that contain BV. These FPs, such as mRhubarb720, have low QY with the aim that fluorescence will not overpower the Raman signal detected in CARS. Additionally, they have comparatively higher extinction coefficients, which means they have

a larger molecular dipole moment and thus higher Raman cross-sections, making them ideal for epr-RS. These near-IR FPs contain a number of C=C, C=N and C=O bonds within the electronically enhanced linear tetrapyrrole chromophore, which can be vibrationally detected between 1630-1660 cm^{-1}, falling within the fingerprint region of the Raman spectrum. Given the abundance of these bonds in biological systems, significant signal amplification via coherent nonlinear enhancement is necessary to detect them effectively free of background within this region.

This chapter aims to conduct an analysis of the enhanced Raman properties of selected near-IR FPs. The objectives include developing a rationale for FP selection based on the Raman enhancement of the chromophore, investigating the impact of the microscope/laser setup on the Raman properties of these proteins. Furthermore, it is important to identify the characteristic Raman peaks specific to the near-IR FPs and compare these spectra with those of classical β-barrel FPs (e.g. mCherry) and free-BV (the chromophore for many IRFPs). Based on the analysis, the selection of the most promising of FPs will be designated for *in situ* imaging in *E.coli* and then as fusions in mammalian cells to examine the suitability of genetically encoding Raman-active probes.

4.2 Results and discussion

4.2.1 Raman spectral analysis of FPs

The relationship between near-IR FP probes to the Raman signal, in particular SRS, is crucial to understanding what wavenumber to use for detection in cell imaging. In this section, the focus will be on Raman spectroscopy on selected near-IR FP, a classical β-barrel far red FP (mCherry) as well as free BV, using SRS to outline the vibrational frequencies within the fingerprint region.

The SRS intensities vary across each of the FPs measured within the fingerprint region (Figure 4.1). The FP with the highest SRS intensity is mRhubarb720, which confirms our hypothesis that the further red-shifted FPs, will present greater enhancements with the current set-up. The SRS measurements in Figure 4.1 are non-background subtracted, therefore, we observe three peaks within the fingerprint region – the broadest peak is present between 1700-1850 cm^{-1} is caused via a two-photon absorption effect. However, this effect is not present in the free-BV sample.

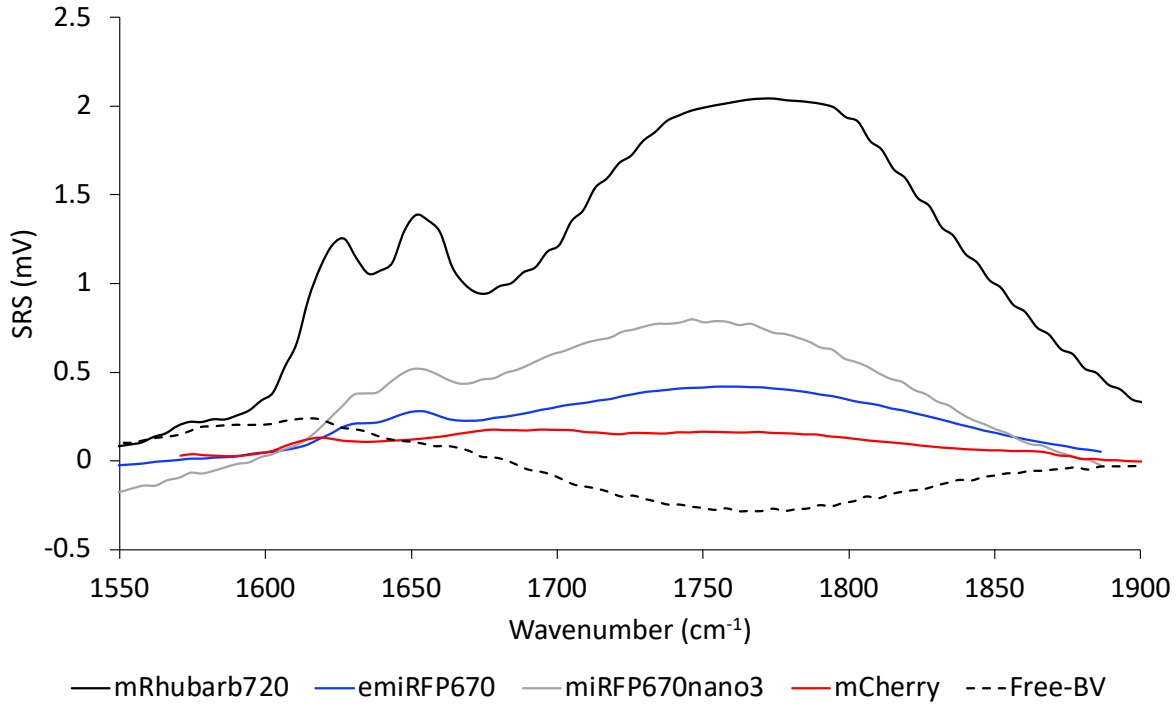


Figure 4.1: Non-background subtracted SRS measurements of FPs and Free-BV within the fingerprint region. mRhubarb720 (black), emiRFP670 (blue), miRFP670nano3 (grey), mCherry (red) and free-BV (black dashed line). All measurements taken at 1 mM sample concentration with the FPs in 50 mM Tris buffer, and the Free-BV in 50 mM Tris and 50 mM NaOH buffer.

When pre-resonantly exciting chromophores, this two-photon absorption (TPA) phenomenon is observed due to the simultaneous absorption of the pump and Stokes lasers when they overlap temporally. Since the spectral tuning in the setup used is using spectral focussing, the spectral overlap varies with the wavenumber addressed. TPA is resulting in two-photon fluorescence and is enabled by a missing inversion symmetry along the molecular dipole. The molecule is excited, absorbing two photons at the same time to create an electronic excitation, which subsequently relaxes within the excited state manifold similar to after one-photon absorption, which can fluoresce. It is a background without vibrational information and is fitted with a suited broad peak function and removed.

As free-BV has a much broader and weaker absorbance, likely due to the reduced coherence of the dipole excitation when surrounded by solvent, this phenomenon is not observed. The solvent for the *in vitro* absorbance reading contains NaOH to solubilise the cofactor which will create a more basic pH and lead to deprotonation of the BV, that could

lead to alternative spectral results to the BV-bound to mRhubarb720. The SRL measurements reported here use high frequency modulation of the Stokes (2.5 MHz) and phase-sensitive detection, which allows to separate instantaneous TPA and SRL to long-lived photothermal effects which are observed in quadrature (i.e. with 90 degree phase shift). The in-quadrature signal measured separately in the dual channel lock-in used was insignificant for all measurements reported (apart from localised absorbers in the samples, unrelated to the fluorophore).

Since mRhubarb720 exhibits the highest Raman enhancement, the background-subtracted curves for the other FPs and free-BV are compared to those of mRhubarb720 in Figure 4.2. The SRL intensity for each mRhubarb720 peak is 0.90 mV at 1630 cm^{-1} and 0.82 mV at 1660 cm^{-1}. There is no significant SRL for free-BV, indicating that free BV will not create a significant background in cell imaging (Figure 4.2.A). mRhubarb720 is found to have an 11-fold larger SRL signal than BV at 1627-1652 cm^{-1}. Only protein-BV complexes are found to have vibrational features in the SRL spectrum, providing a pathway for genetically encoding Raman-active probes to be used in live-cell imaging.

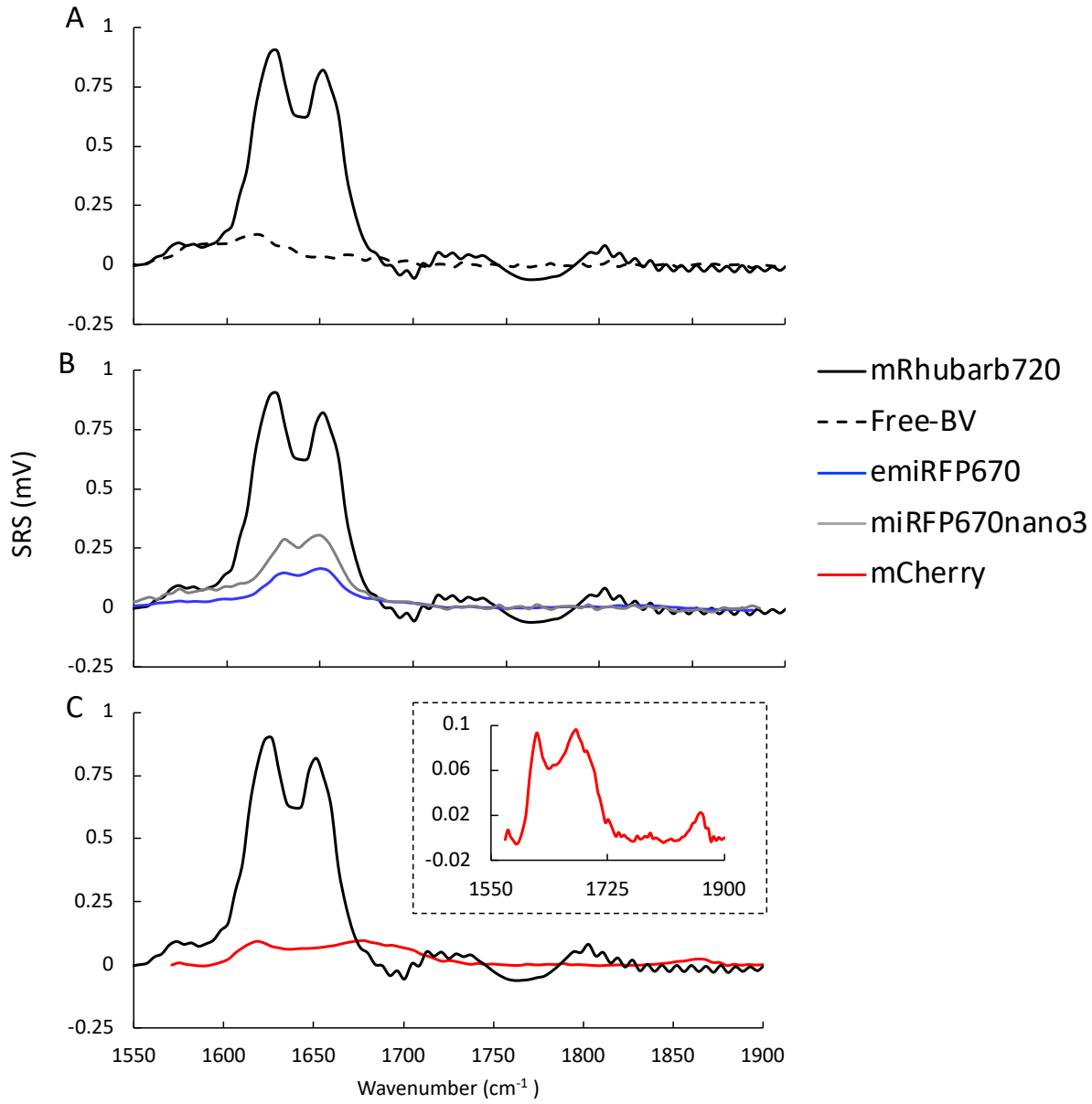


Figure 4.2: Background subtracted SRL measurements of FPs and free-BV in the fingerprint region. (A) mRhubarb720 (black line) compared to free-BV (black dotted-line). (B) emIRFP670 (blue line) and mIRFP670nano3 (grey line) compared to mRhubarb720. (C) mCherry (red) compared to mRhubarb720, with excerpt showing mCherry only. All measurements were taken with 1 mM sample concentration.

For emiRFP670 and miRFP670nano3, there is also a double peak present at 1627-1652 cm^{-1} which can be more clearly observed when the two-photon absorption background is removed (Figure 4.2.B). The SRL intensities for miRFP670nano3 with 0.29 mV and 0.30 mV, whereas emiRFP670 was lower again at 0.15 mV and 0.16 mV, respectively. The SRL intensity is increasing closer to resonance, and therefore, mRhubarb720 was the most enhanced

near-IR FP due to the red-shifted nature of its absorbance and increased molecular excitation coefficient compared to the others (see Chapter 3.2.2).

mCherry exhibited the lowest Raman enhancement among all the FPs, with intensities of 0.093 mV at 1619 cm^{-1} and 0.096 mV at 1678 cm^{-1} (Figure 4.2.C). These wavenumbers differ from those observed in the BV-binding FPs, which is expected due to the difference in the number and type of vibrational bonds within the respective chromophores. The specific configurations of the chromophores dictate the precise vibrational frequencies of the bonds (C=C, C=N and C=O). Since the structure of the mCherry chromophore is significantly different from that of BV, the enhancement peaks vary, even though the same vibrational bonds are being detected.

The SRS data highlights the differences among the selected FPs and supports the selection of mRhubarb720 for further cell imaging studies. Its superior SRS intensity, along with the distinct double peak observed, makes it the most promising candidate for developing a genetically encoded Raman-active probe for cellular imaging. The strong Raman enhancement provided by mRhubarb720 ensures higher sensitivity, allowing for the detection of lower concentrations within cells, in respect to the other FPs in this study.

4.2.2 The effect of the Pump/Stokes laser power on mRhubarb720.

As previously mentioned, both SRS and CARS uses two lasers that are tuned to create vibrational coherence in the sample. In SRS, this coherence leads to either a stimulated Raman gain in the Stokes beam or a SRL in the pump beam. In CARS, the same vibrational coherence results in anti-Stokes scattering at a frequency higher than the Pump. Both methods enhance the Raman signal, allowing for the sensitive detection of specific molecular bonds. Therefore, it is crucial to understand how laser power influences the SRS and CARS intensity output to accurately determine the optimal laser power for maximising signal strength. If the laser power is too high, the molecules can be photobleached and rendered unusable. Conversely, when the laser power is too low, the inelastic scattering will be undetectable. The mRhubarb720 sample was measured with varying Pump and Stokes laser power and the SRS and CARS signal was detected (Figure 4.3).

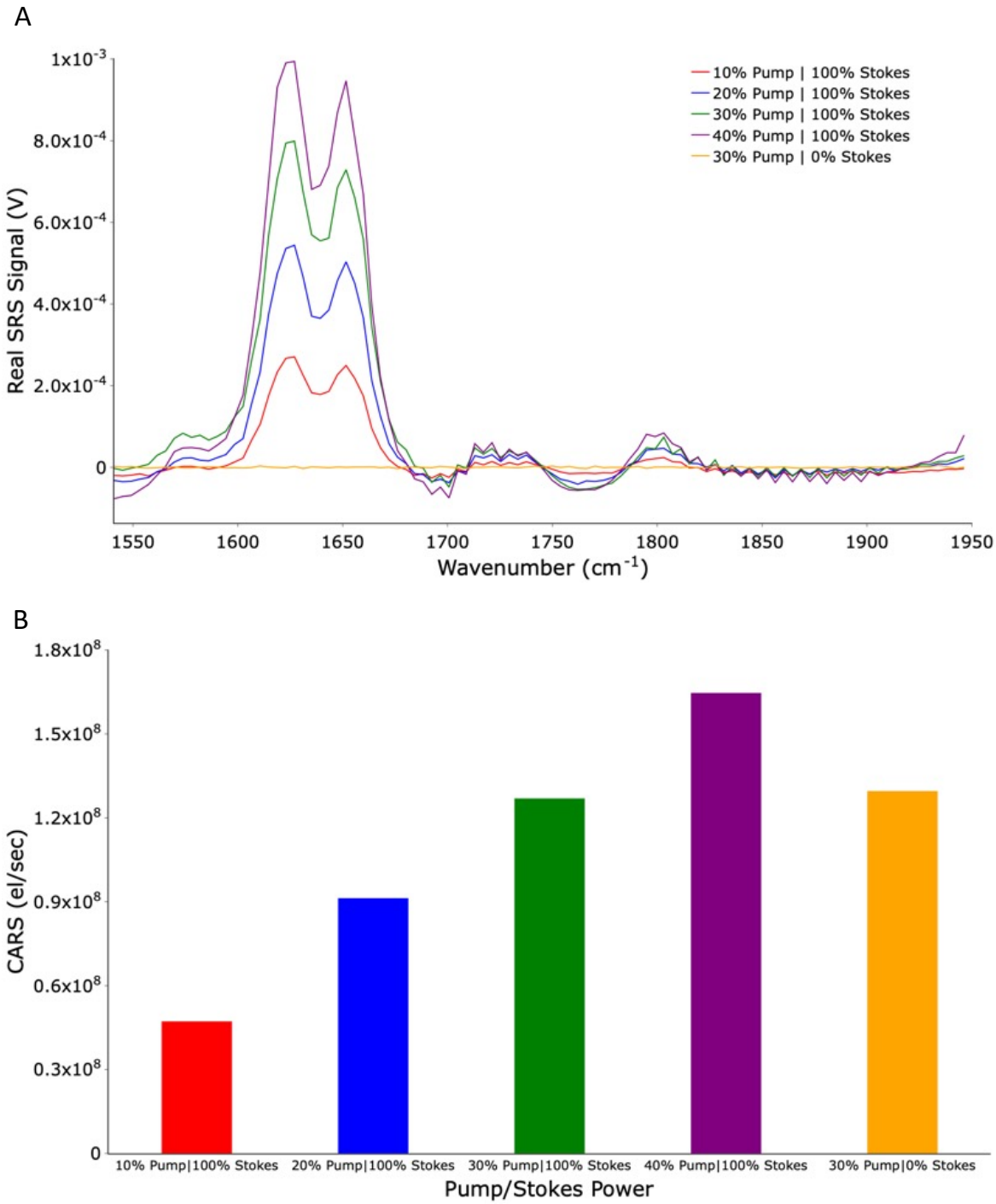


Figure 4.3: Measured SRS and CARS signal of mRhubarb720 versus Pump/Stokes power. (A) SRS signal and (B) CARS signal measurements. All measurements were taken with 1mM mRhubarb720 concentration. 10% Pump | 100% Stokes (red), 20% Pump | 100% Stokes (blue), 30% Pump | 100% Stokes (green), 40% Pump | 100% Stokes (purple), 30% Pump | 0% Stokes (orange).

The pump and Stokes laser power is denoted as percentages of the overall power illuminating the sample (40 mW at 100%). The in-phase SRL signal is found to be proportional to the pump power, as expected from the SRL process with a signal scaling proportional to both pump and stokes power. When the Stokes beam is set to zero, no SRL signal is found, as expected. The detected CARS signal was rather independent of the wavenumbers measured (1550-1950 cm^{-1}), i.e. the pump-Stokes delay, and was found to be dominated by the pump only, changing only slightly without the Stokes. CARS intensity is expected to scale linearly with Stokes power and quadratically with pump power. The signal detected is not CARS but rather mRhubarb720 fluorescence after one or two-photon excitation, which occurs at the same wavelength as CARS and is transmitted by the detection filter. Therefore, the microscopy method of choice will be SRS to observe the pre-resonantly enhanced coherent Raman spectra arising from electronic transitions within the fluorescent proteins.

The Pump and Stokes pulse power ratio is set to about 1:3, using 13.3 mW Pump and 40 mW Stokes, to achieve a strong SRS signal while limiting the damage of the molecules targeted for enhancement (Váczi et al. 2023). The pump power is chosen lower than the Stokes power as it prevents more bleaching due to the TPA previously mentioned.

4.2.3 *In situ* imaging and analysis of mRhubarb720: *E.coli*

For near-IR FPs to act as probes in Raman-based cell imaging approaches they need to be detectable in a cellular environment. Initial imaging experiments used over-expressed mRhubarb720 in *E. coli*. If mRhubarb720 functions effectively as a genetically encoded SRS probe, it is expected to be observable using the in-house SRS microscopy setup.

To demonstrate this concept, quantitative Differential Interference Contrast (qDIC) and fluorescence imaging were first performed to identify the location of the bacteria and to show that mRhubarb720 is being expressed due to its inherent fluorescence. These imaging experiments also served as controls to verify that bacteria were being imaged and to identify which are producing mRhubarb720. qDIC uses gradients in the optical path length to introduce contrast to the samples, whereas the fluorescence will show the mRhubarb720-expressing bacteria (Figure 4.4).

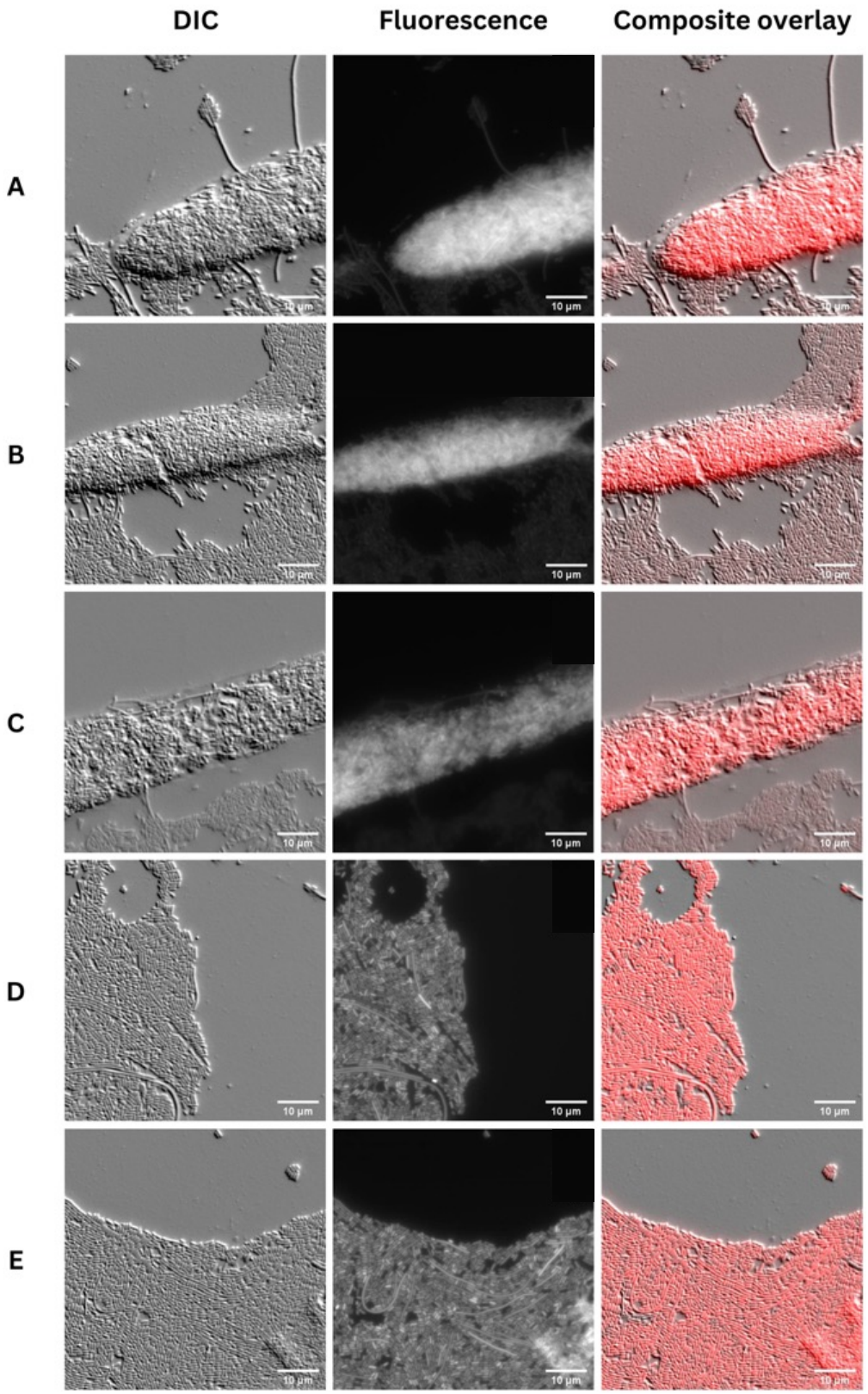


Figure 4.4: Fluorescence and qDIC imagining of mRhubarb720-expressing *E.coli* cells. Five sample images labelled A-E with qDIC parameters using a range from black (-0.14 rad) to white (0.14 rad). The fluorescence ranges from 0 to 30,000 counts. Composite overlay images combines both microscopy techniques. Illumination and detection filters are listed in Methods section 2.5.4.

The figure displays qDIC, fluorescence and a composite overlay for various sample regions. In panels A-C, the qDIC clearly shows the morphology and structure of multiple layers of *E.coli*, forming aggregates. Conversely, in panels D and E, the qDIC displays a single layer of *E.coli*. The corresponding fluorescence images confirm the expression of mRhubarb720 and the composite overlays in these panels demonstrate the colocalization of the qDIC and fluorescence signals, highlighting that most bacteria are expressing mRhubarb720.

After the expression of mRhubarb720 by *E.coli* has been confirmed, the next step involved assessing the ability to image the cells using the epr-RS approaches. The SRL spectrum of mRhubarb720 has a distinct double peak at 1627 and 1652 cm^{-1} (Figure 4.2.a) indicating that SRS imaging should focus on either of these wavenumbers. Given the nature of the microscopy method, which involves the excitation of the probes, targeting a single peak is advisable to avoid overstimulation. Excessive scanning could lead to photobleaching, compromising signal quality and reducing the effectiveness of the imaging. The 1652 cm^{-1} peak exhibited approximately 10% higher intensity compared to the 1627cm^{-1} peak when background subtraction was not applied (Figure 4.1). Since background subtraction is not performed when imaging at single absolute wavenumbers, the decision was made to focus on the 1652 cm^{-1} peak due to its greater intensity. Consequently, SRL was conducted over the range of 1645-1675 cm^{-1} in 10 cm^{-1} increments (Figure 4.5). For the CARS imaging, each wavenumber was measured, however, only 1665 cm^{-1} was presented here because the previous experiment in Section 4.2.2 demonstrated that CARS primarily detects non-resonant background signals from fluorescence rather than Raman scattering (See section 4.2.4.1 for more details).

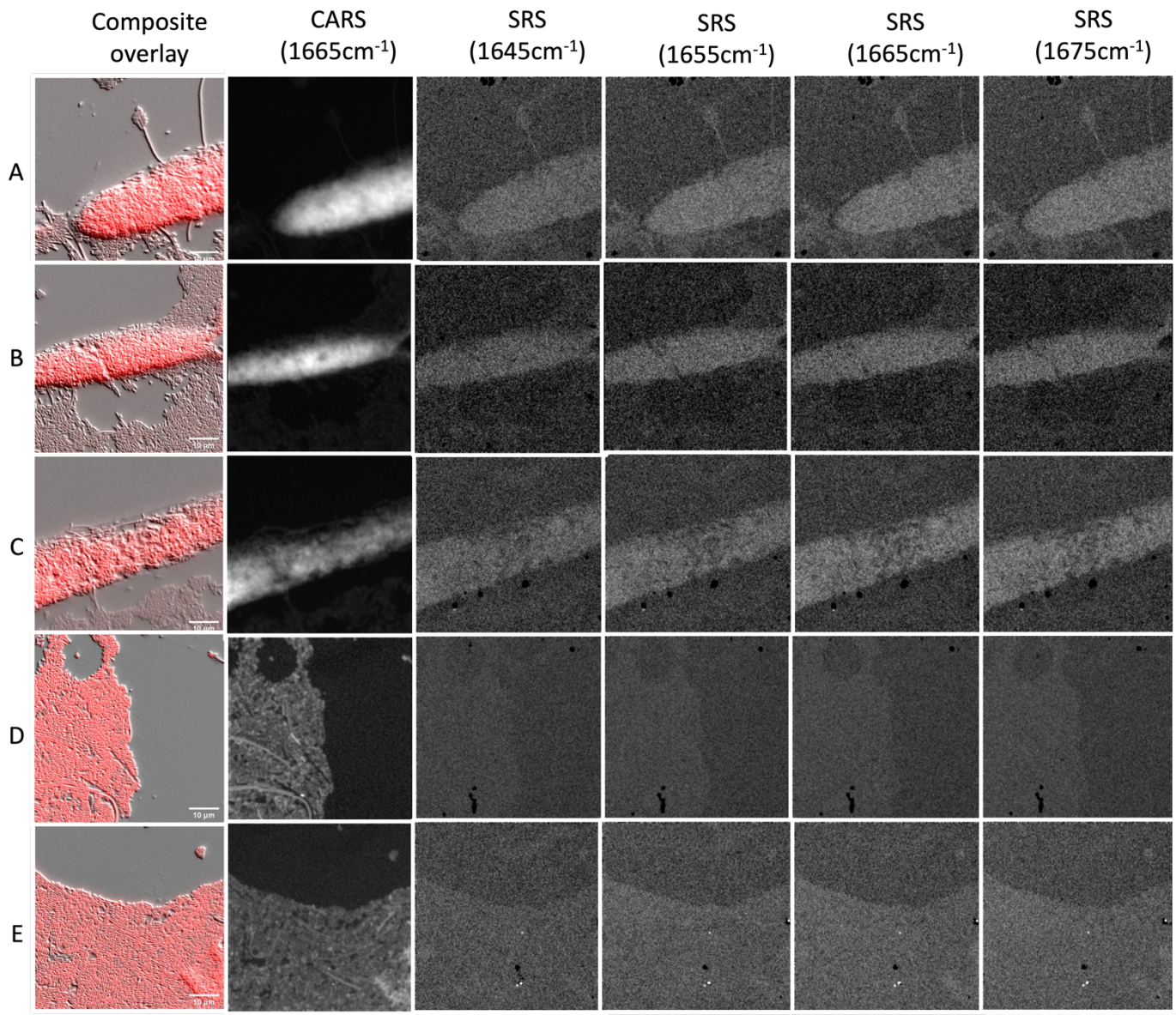


Figure 4.5: CARS and SRS imaging of *E.coli* expressing mRhubarb720. Six repeats were taken per wavenumber, where each repeat was produced via one scan. Images were then averaged over each repeat with 30% Pump power and 100% Stokes power. Grey scale range from m to M for the CARS signal varied over each sample: (A) M = 5.5 Mel/sec, m = 0.2 Mel/sec. (B) M = 4.2 Mel/sec, m = 0.24 Mel/sec. (C) M = 4.0 Mel/sec, m = 0.2 Mel/sec. (D) M = 0.7 Mel/sec, m = 0.3 Mel/sec. (E) M = 1.0 Mel/sec, m = 0.24 Mel/sec. All SRL images have a range of M = 0.6 mV, m = -0.2 mV. Mel refers to mega-electron volts (10^6), m is the minimum value and M is the maximum value.

Observations suggest that CARS imaging provides a signal comparable to fluorescence, particularly when imaging single-layered cells (samples D-E). This outcome aligns with the expectation that the CARS signal is dominated by fluorescence rather than Raman scattering, as previously described. Single cells can be identified in samples D and E,

this is attributed to the minimal scattering or absorption from multi-layered samples, which would otherwise compromise signal quality, as seen in samples A-C. However, further quantitative analysis and higher-resolution imaging are needed to substantiate these observations and verify single-cell resolution. The maximum signal intensities in the multi-layered cells are approximately 5-fold higher than those in single cells. Specifically, the signals for multi-layered samples (A-C) are as follows: 5.5 Mel/sec, 4.2 Mel/sec and 4.0 Mel/sec, compared to the single-layered samples (D-E) with 0.7 Mel/sec and 1.0 Mel/sec, respectively. This increase in signal intensity in multi-layered cells results in greater scattering and absorption of light, which can compromise the resolution in both CARS and fluorescence imaging (Li et al. 2020). However, the exact spatial resolution for these *E.coli* imaging data was not a primary focus in this study. While the observed differences in signal intensity suggest potential impacts of multi-layered samples on imaging quality, more targeted experiments would provide a clearer assessment. For instance, phase-shifted spatial modulation of the signal output could be used in future studies to obtain direct spatial and spectral resolution measurements (Lv et al. 2022).

Unlike CARS, SRS does not produce significant non-resonant background fluorescence signal (see Figure 4.3), which will lead to multiplexing opportunities and a clearer resolution on a multi-layered system (Tsikritsis et al. 2022). This concept is highlighted in sample A, where the cells can be observed to form an offshoot protruding away from the aggregation. The SRS signal at 1655 cm^{-1} was expected to exhibit optimal intensity, with a slight decrease observed in the other wavenumbers based on the *in vitro* experiments previously described. However, this expected trend is less discernible from the multi-layered samples (A-C). This effect is attributed to the high sensitivity of SRS coupled with the increased number of over-expressing cells. It is important to note that this analysis is qualitative, and like the CARS imaging, further experiments would be necessary to gain quantitative data on spatial resolution.

4.2.4 Mammalian cell imaging with mRhubarb720.

To evaluate the feasibility of utilising pre-resonant coherent Raman spectroscopy for cellular imaging, mRhubarb720 fused to various proteins were constructed for production in mammalian cells. Such FP fusions are commonly used in fluorescence-microscopy based cell

imaging approaches. This selection will ultimately look at localisation of certain proteins fused to mRhubarb720 to different locations in the cell. If implemented correctly, this technology could aim to enhance specificity and sensitivity in detecting molecular interactions and structural dynamics within different cellular environment. The HeLa cell line, renowned for its consistency, reproducibility, and high transfectability, was chosen as the model cell system for this study. Three specific proteins were fused to mRhubarb720: Histone2B (H2B), LAMP1, and LifeAct (Figure 4.6). Alphafold3 models were generated for all three constructs to evaluate the structural suitability of tagging the biomolecules with the probes. It is imperative that by tagging the biomolecules, the protein functions are not obstructed.

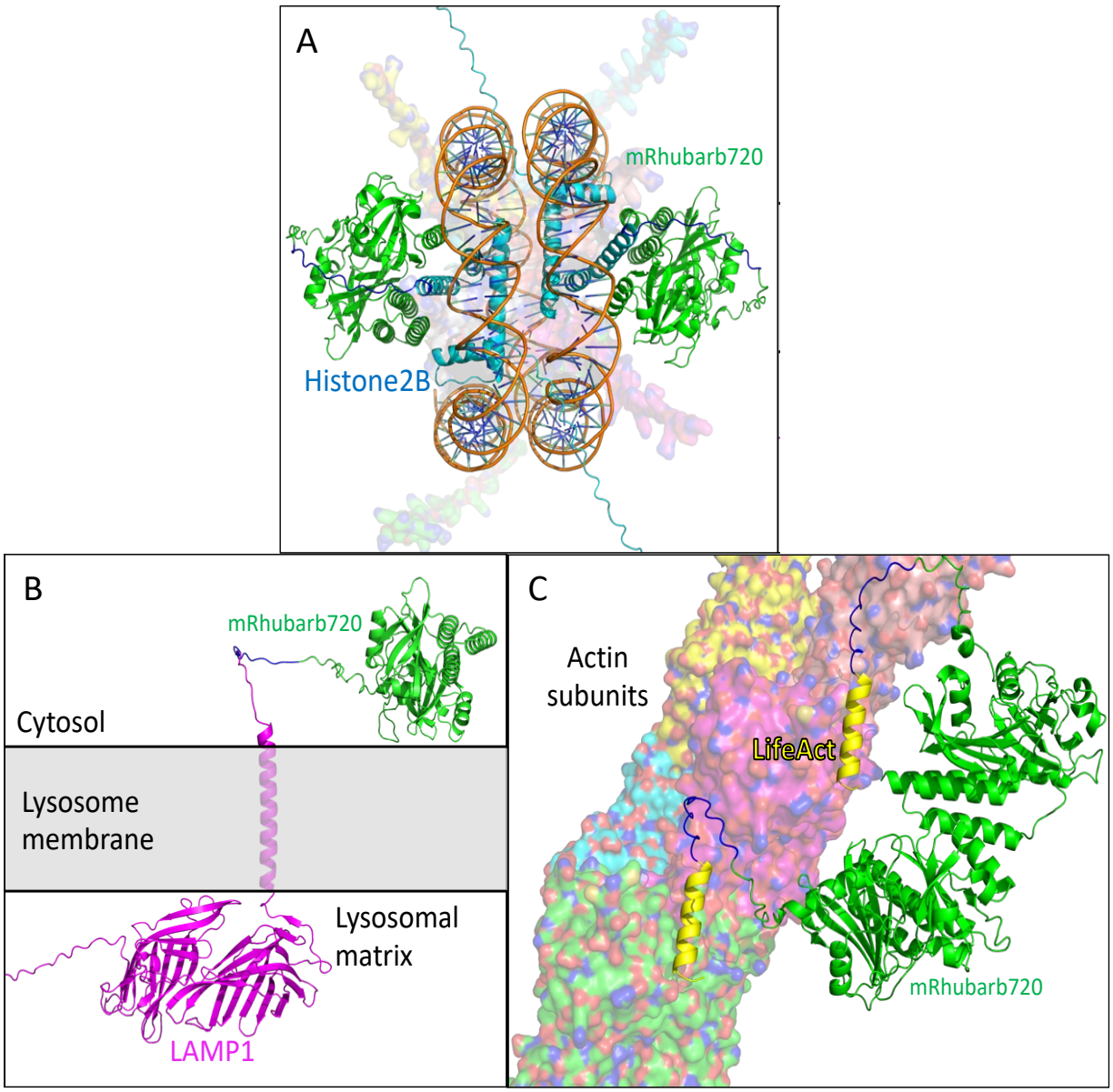


Figure 4.6: Alphafold3 models of target biomolecules coupled to mRhubarb720. (A) Histone2B (cyan) complexed with mRhubarb720 (green) is shown within the nucleosome, with DNA (orange) coiled around it. The sequences of all histone subunits and the associated DNA sequence were submitted to Alphafold3, and the lowest energy model is presented. (B) Lysosomal-associated membrane protein 1 (LAMP1, magenta) bound to mRhubarb720, with the lowest energy model selected. Structural adjustments were made to illustrate the position of the lysosomal membrane. (C) LifeAct (yellow) complexed with mRhubarb720, shown bound to F-actin subunits (surface). Dark blue cartoon represents the linker peptide between target and probe.

Each model provides structurally sound complexes throughout the secondary structures of both the probe and target biomolecule. However, Alphafold still has difficulty

with incorporating post-translational modifications to proteins, especially with covalent linkages, which is why BV was not incorporated into the predictions. Accurate modelling of fusion proteins is crucial, as fusions can affect the function of the target. For instance, fusing an FP to the N-terminal domain of a target containing a signal sequence could inhibit protein function, and potentially compromise the experiment.

The H2B-mRhubarb720 complex is expected to position two probes either side of the nucleosome. When chromatin condensation is activated, the histones will cluster together, could potentially lead to an enhanced signal due to the increased local concentration of mRhubarb720 probes. The LAMP1 protein is localised to the lysosomal membrane with the mRhubarb720 contained within the cytosol rather than the lysosomal matrix, due to the C-terminal attachment. This positioning prevents denaturation of the probe, as the intra-lysosomal matrix contains over 40 hydrolytic enzymes with an acidic pH between 4.5-5.0 (Chen et al. 2020). Finally, LifeAct, a short actin-binding peptide that is regularly used by cell biologists to visualise F-actin structures, is predicted to reveal thin filamentous structures. Non-fluorescent control samples for both bacterial and mammalian cells were also measured, for both CARS and SRS, which can be found in Supplementary Figures S1 and S2.

4.2.4.1 Cell imaging of mRhubarb720-Histone-2B fusions

The H2B-mRhubarb720 complex was transfected in HeLa cells in order to image H2B using pre-resonant coherent Raman scattering, where the expectation is to see signal in the cells nuclear environment. Initial imaging involved measuring five repeats per wavenumber, with each repeat averaged over four scans per sample (Figure 4.7).

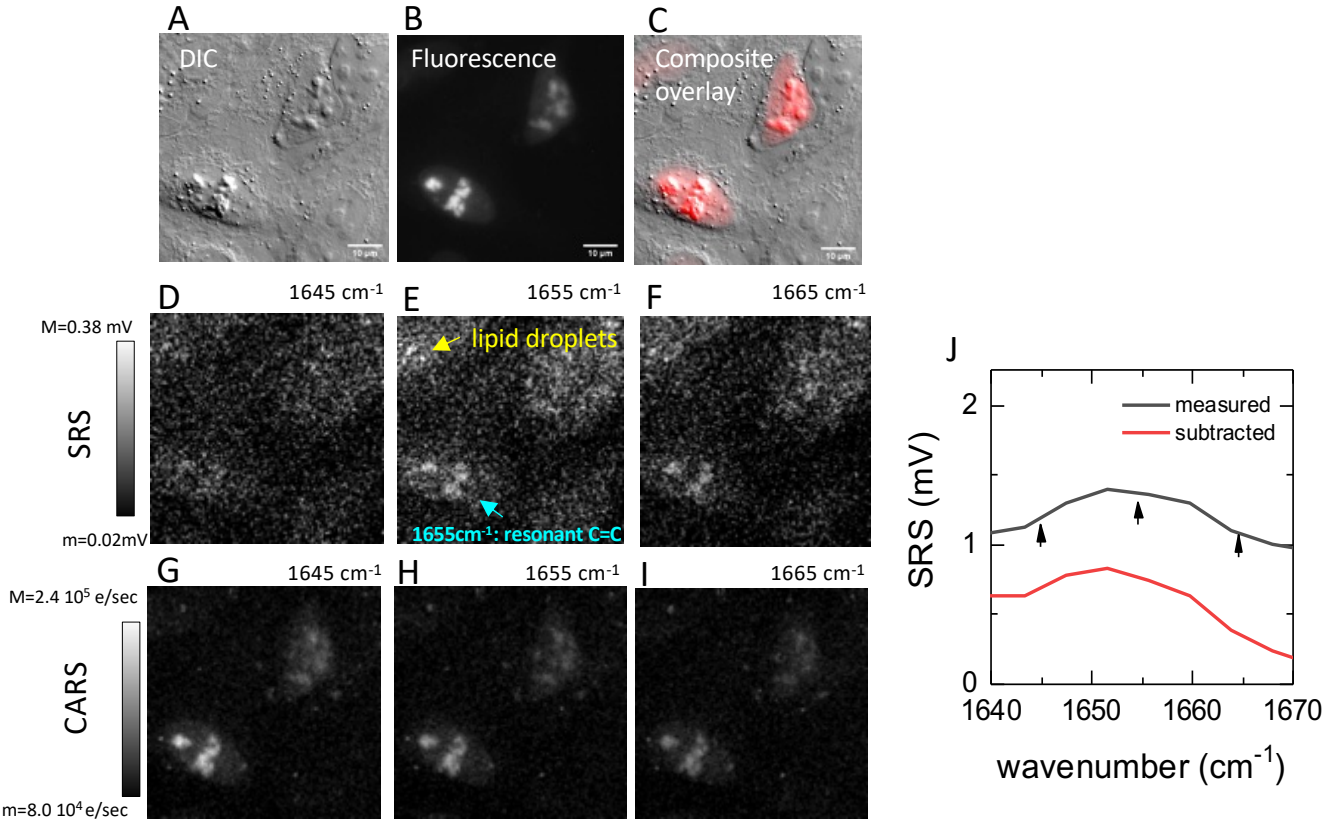


Figure 4.7: Mammalian cell imaging of H2B-mRhubarb using CARS/SRS. Images (A-C) include DIC, fluorescence and a composite overlay, respectively. Both SRS (D-F) and CARS (G-I) images were produced using three wavenumbers, 1645 cm^{-1}, 1655 cm^{-1} and 1665 cm^{-1}, respectively. The graph (J) highlights the *in-vitro* SRS analysis of mRhubarb720 to show location of the peak measurements for cell imaging.

Fluorescence imaging confirmed the localised H2B-mRhubarb720 constructs within the HeLa cells, revealing structures within two HeLa cells (Figure 4.7.A-C). The SRL imaging was performed after epi-fluorescence measurements, where the peak *in vitro* signal (Figure 4.7.J) was measured to be at 1652 cm^{-1}, leading to the expectation of a greater signal at the 1655 cm^{-1} images. SRS imaging qualitatively revealed a distinct resonant effect of the C=C bonds in BV, demonstrated by a clear spectral peak at 1655 cm^{-1}, as opposed to the image at 1645 cm^{-1} (Figure 4.7.D-F). ImageJ analysis was performed on the lower left HeLa cell in each image, where it was calculated that 1655 cm^{-1} was 44% and 38% higher intensity than 1645 cm^{-1} and 1665 cm^{-1}, respectively. In addition, structures resembling lipid droplets, known to be rich in C=C bonds, were observed in the SRS without any fluorescent signals, which

suggest their identity. Based on the spectral characteristics and morphology, these structures are consistent with lipid droplets, though additional validation, such as chemical staining, would be required to definitively identify them. However, for CARS (G-I), a decrease in signal intensity was observed when shifting from 1645 cm^{-1} to 1665 cm^{-1}, suggesting a photobleaching effect commonly seen in fluorescence microscopy. Quantification of the lower left HeLa cell displayed a consistent decline in intensity, with the signal at 1645 cm^{-1} being a 17% higher than at 1655 cm^{-1}, and 35% higher than 1665 cm^{-1}. This is a uniform decrease of intensity is characteristic of a photobleaching effect.

To investigate the apparent photobleaching observed in Figure 4.7.G-I, fewer averages were taken, with six repeats per wavenumber, but each repeat was obtained over a single scan, compared to the previous method of five repeats per wavenumber averaged over four consecutive scans (Figure 4.8). This adjustment aimed to reduce photobleaching or potential photoisomerisation of the BV. As a result, we expect to see an increase in intensity at the Raman peaks (1655-1665 cm^{-1}) for both CARS and SRS. Additionally, measuring a CARS signal without a Stokes beam, as shown in Figure 4.3, indicates that CARS is dominated by fluorescence and CARS background, rather than resonant vibrational features. If this is the case, it would present significant limitations for using CARS as a reliable microscopy method in this context.

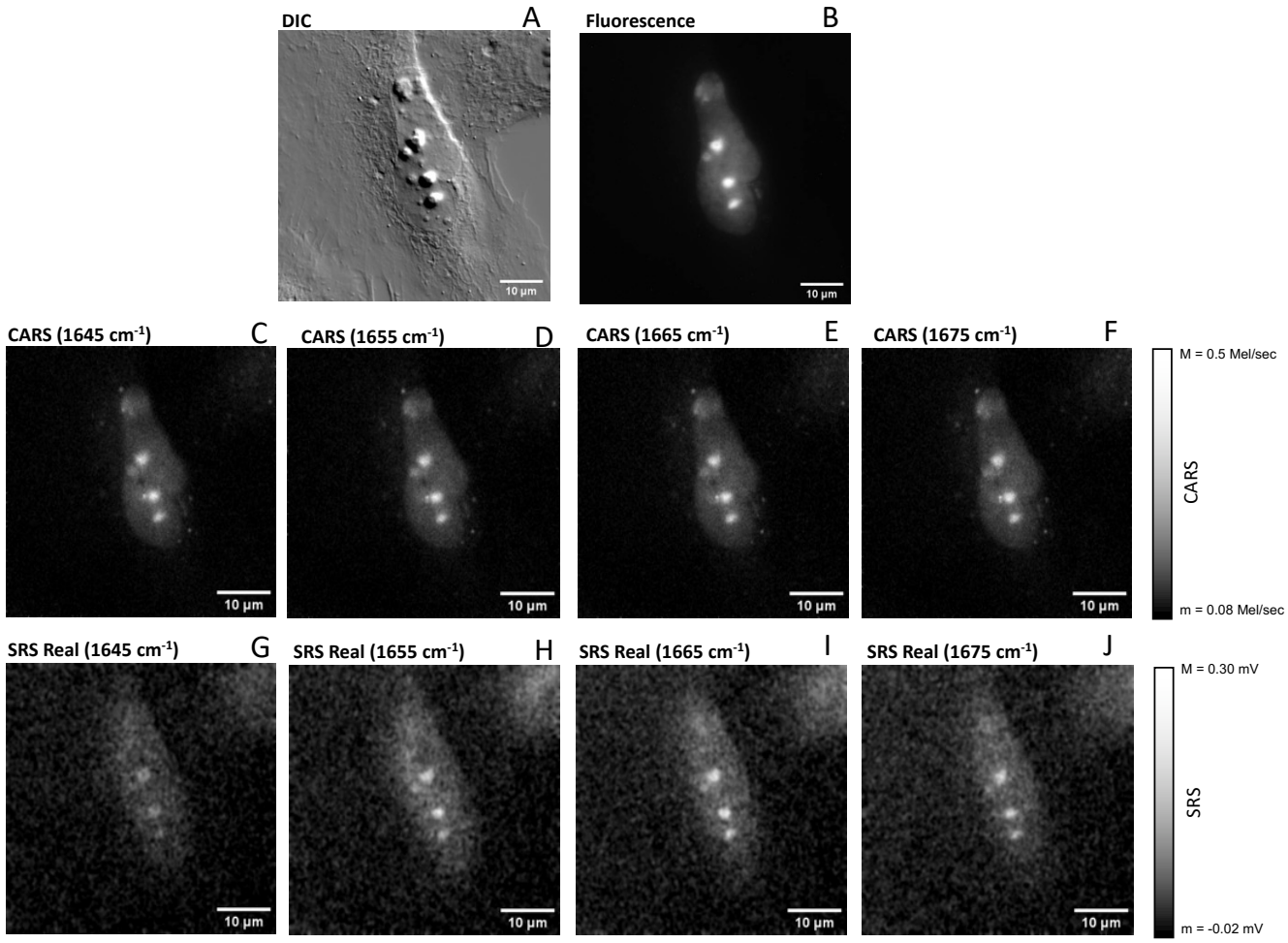


Figure 4.8: Mammalian cell imaging of H2B-mRhubarb720 using a reduced number of repeats for CARS/SRS. DIC and fluorescence (A-B, respectively) show a single cell fluorescing via the H2B-mRhubarb720 construct. CARS (C-F) and SRS (G-J) measured over four wavenumbers: 1645 cm^{-1}, 1655 cm^{-1}, 1665 cm^{-1} and 1675 cm^{-1} (left to right).

The CARS imaging shows no variation in signal across all the measured wavenumbers, indicating that the CARS photomultiplier tube (PMT) is likely detecting fluorescence rather than CARS, as demonstrated by Figure 4.3.B. Quantitative analysis of the intensity measurements across the four images further supports this, revealing minimal percentage differences, mostly under 5%, which suggests no significant change in intensity. The previous imaging presented in Figure 4.7 exhibited a photobleaching effect; however, this effect was mitigated by reducing the number of repeats. These findings confirm that CARS is not suitable for detecting of the vibrational modes of Raman amplified fluorescent proteins using the current set-up.

In contrast, SRS demonstrates an 46% increase in intensity between at 1645-1655 cm^{-1}, followed by an 8% decrease in intensity between 1655-1665 cm^{-1}, and a further 19% decrease in intensity between 1665-1675 cm^{-1}. These findings highlight the potential for multiplexing with genetically encoded Raman-active probes. To demonstrate the change intensity, a false colour added to the same SRS images to enhance the contrast between the wavenumbers (Figure 4.9).

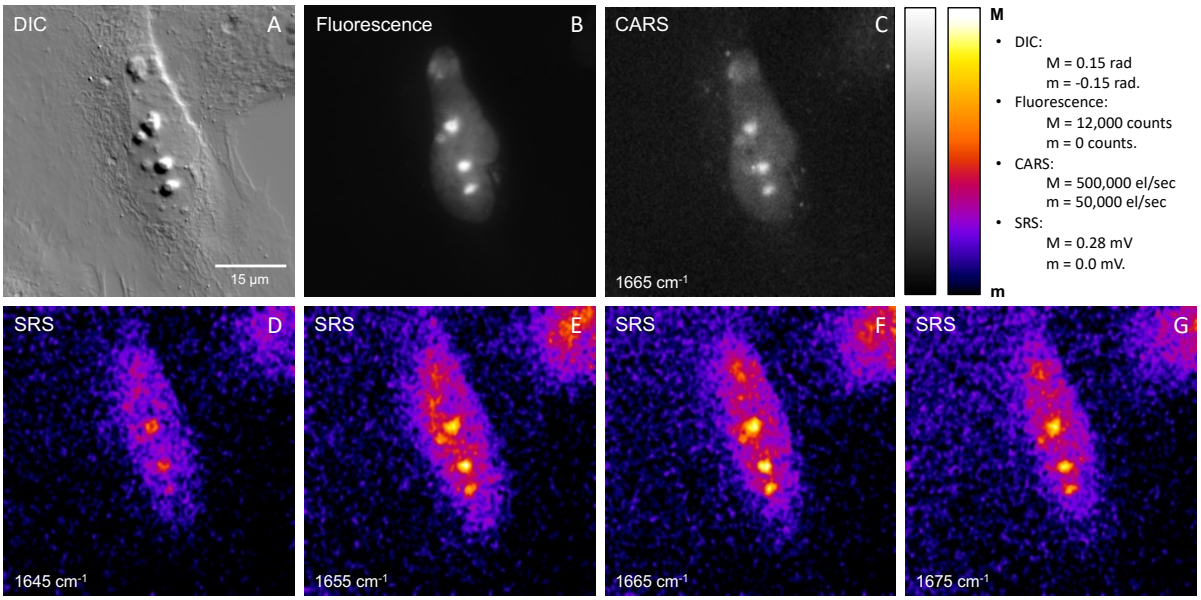


Figure 4.9: Mammalian cell imaging of H2B-mRhubarb720 using false colour to highlight SRS intensity shifts. (A) DIC, (B) fluorescence and (C) CARS at 1665 cm^{-1} present the same information highlighted in the previous figure. (D-G) SRS images were edited using ImageJ to provide false colouring for improved contrast of the Raman signal.

The false colour imaging provides initial qualitative analysis and was utilised to offer an enhanced visualisation of the probe, making it easier to distinguish the intensity between wavenumbers. There is a notable difference between 1655-1665 cm^{-1} and the other wavenumbers where enhancement of the C=C is increased (Figure 4.9.E-F). Subsequent imaging focused on accurately measuring regions of interest to analyse intensity across each wavenumber. This involved selecting two regions within the same image sample, including a non-fluorescent region as a negative control. The mean SRS intensities were calculated for both regions of interest and the control, and the differences were then determined (Figure 4.10).

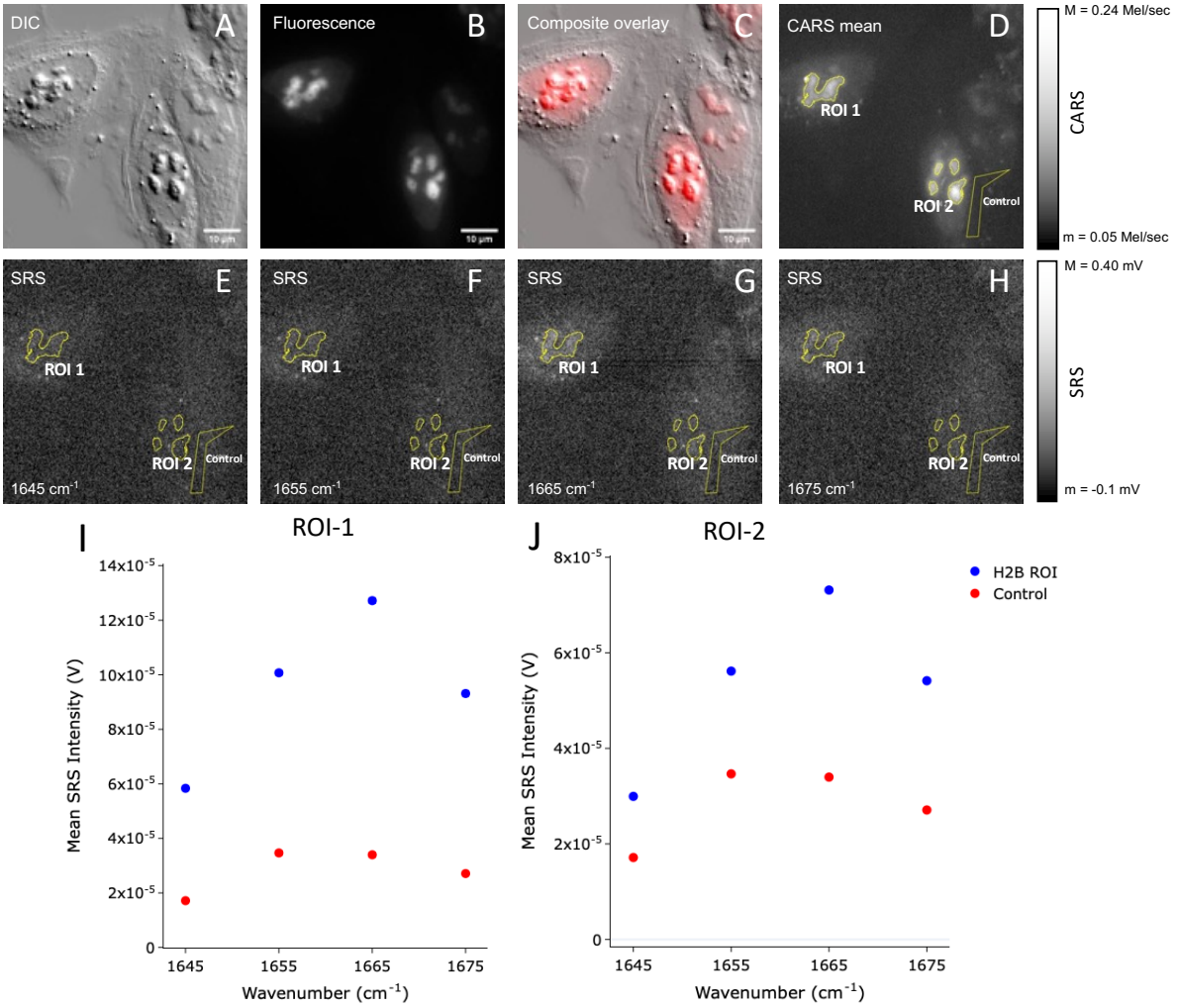


Figure 4.10: Differential analysis regions of interest in the H2B-mRhubarb720 mammalian cell imaging. (A) DIC, (B) fluorescence and (C) composite overlay display the location of the mRhubarb720 within the cells. (D) CARS was averaged across all four wavenumbers that highlight two regions of interest (ROI) and the control region (highlighted in yellow). (E-H) SRS measurement highlights ROI's and the control area, used to quantitatively measure SRS intensity. (I) ROI-1 and (J) ROI-2 differential analysis is presented to show mean SRS intensities of both the ROI (blue) and control (red).

Both regions of interest (ROIs) exhibit a maximum mean SRS intensity at 1665 cm^{-1}, contrasting with the predicted 1655 cm^{-1} maxima from *in vitro* experiments. This ~10 cm^{-1} redshift could be attributed to environmental interactions within the cellular environment, which can influence the electronic and vibrational properties of the molecules (Czamara et al. 2021). Furthermore, the presence of various cellular components and the surrounding matrix can cause shifts in vibrational frequencies due to diverse molecular interactions (Chen

et al. 2023). An important observation is the decrease in intensity on either side of the 1655 cm^{-1} peak, suggesting a narrowing of the peak's intensity, which is critical for sharp spectral analysis and multiplexing. ROI-1 exhibits a higher SRS intensity compared to ROI-2, as depicted by the difference plots on both graphs, with the same control applied to both plots.

The preceding imaging experiments confirmed the localisation of H2B-mRhubarb720 constructs within HeLa cells, revealing distinct structures. SRL imaging showed a significant enhancement of the mRhubarb-BV complex with a notable peak at 1655 cm^{-1}, in contrast to the CARS imaging, which proved insufficient due to a decrease in signal intensity indicative of photobleaching. Despite attempts to mitigate this effect, CARS failed to detect vibrational modes reliably with no variability across the measured wavenumbers. In contrast, SRS demonstrated a maximum 46% increase in intensity at 1655 cm^{-1} compared to 1645 cm^{-1}, with subsequent decreases at higher wavenumbers. This variation highlights SRS's potential for multiplexing with Raman-active probes.

4.2.4.2 Cell imaging of mRhubarb720-LAMP1 fusions

The lysosomal associated membrane (glyco)protein 1 (LAMP1) plays a crucial role in maintaining the structural stability of lysosomes, facilitating autophagy, mediating the transport and fusion of vesicles, and protecting lysosomal enzymes from degradation (Eskelinen and Saftig 2009; Saftig and Klumperman 2009). While LAMP1 is predominantly localised to the lysosomal membrane, its diverse functionalities require its presence in other cellular compartments, such as endosomes, autophagosomes, and the plasma membrane (Reddy et al. 2001; Eskelinen 2006). Consequently, imaging studies are expected to reveal a widespread distribution of LAMP1 signals throughout the cell, with noticeable absence of signal in the nuclear region. This observation is confirmed in Figure 4.11.

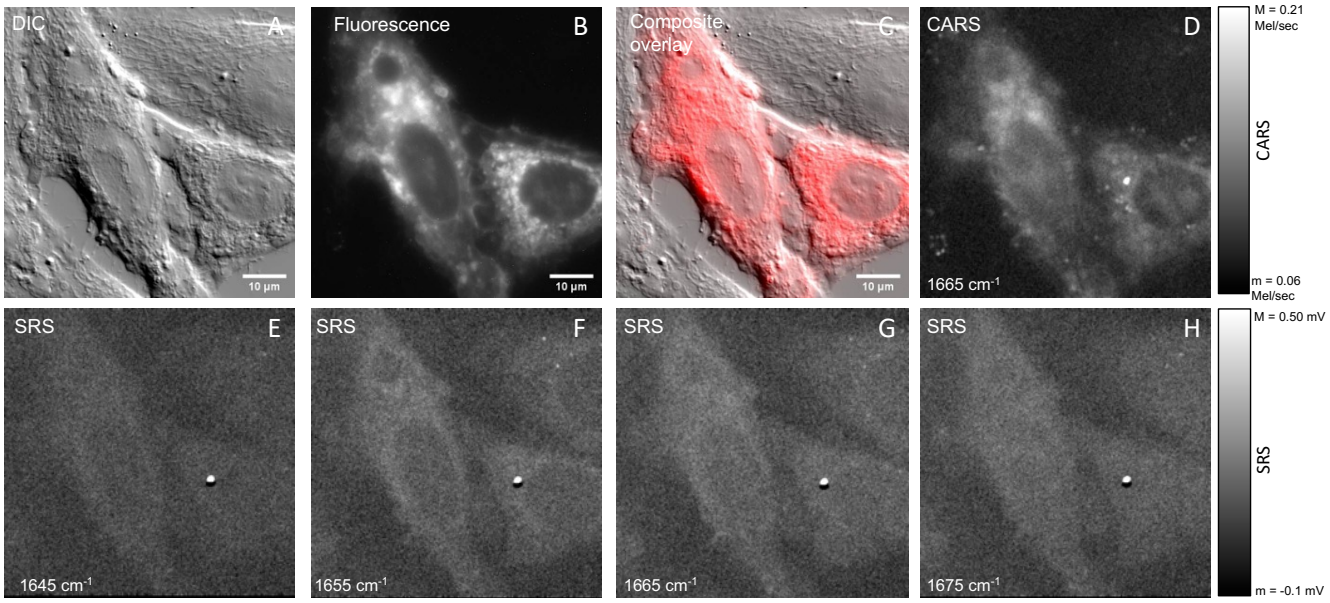


Figure 4.11: Mammalian cell imaging of LAMP1-mRhubarb720. (A) DIC, (B) fluorescence and (C) composite overlay images confirm LAMP1-mRhubarb720 expression within HeLa cells. (D) CARS measured solely at 1665 cm^{-1}, whereas (E-H) SRS was measured across four wavenumbers; 1645-1675 cm^{-1}.

The CARS imaging was comparable to fluorescence, which aligns with expectation given that the PMT detects fluorescence rather than anti-Stokes photons, as previously described. Using ImageJ, SRS voltage for a region of interest (ROI-1) was calculated at 1655 and 1665 cm^{-1}, with an average intensity of 0.17 mV for both wavenumbers (Figure 4.12). A notable drop-off in intensity for 1645 and 1675 cm^{-1} was observed with 0.10 mV and 0.15 mV, respectively. The signal strength at these SRS peaks remains low, only the lipid bilayers are visible in the images, with no evident contribution from mRhubarb720 (see fluorescence image B in top right, which shows no signal). Therefore, the signal observed in the top right in the SRS images (Figure 4.11.F-H) corresponds solely to the lipid bilayers. Significant improvements are necessary, either through modifying the probe or the microscope set up, to achieve a high resolution genetically encoded Raman-active probe. Increasing the density of the probe will enhance the SRS output, as observed in the H2B-mRhubarb720 constructs. Since the LAMP1 protein is relatively dispersed throughout the cell, a reduced SRS output is observed.

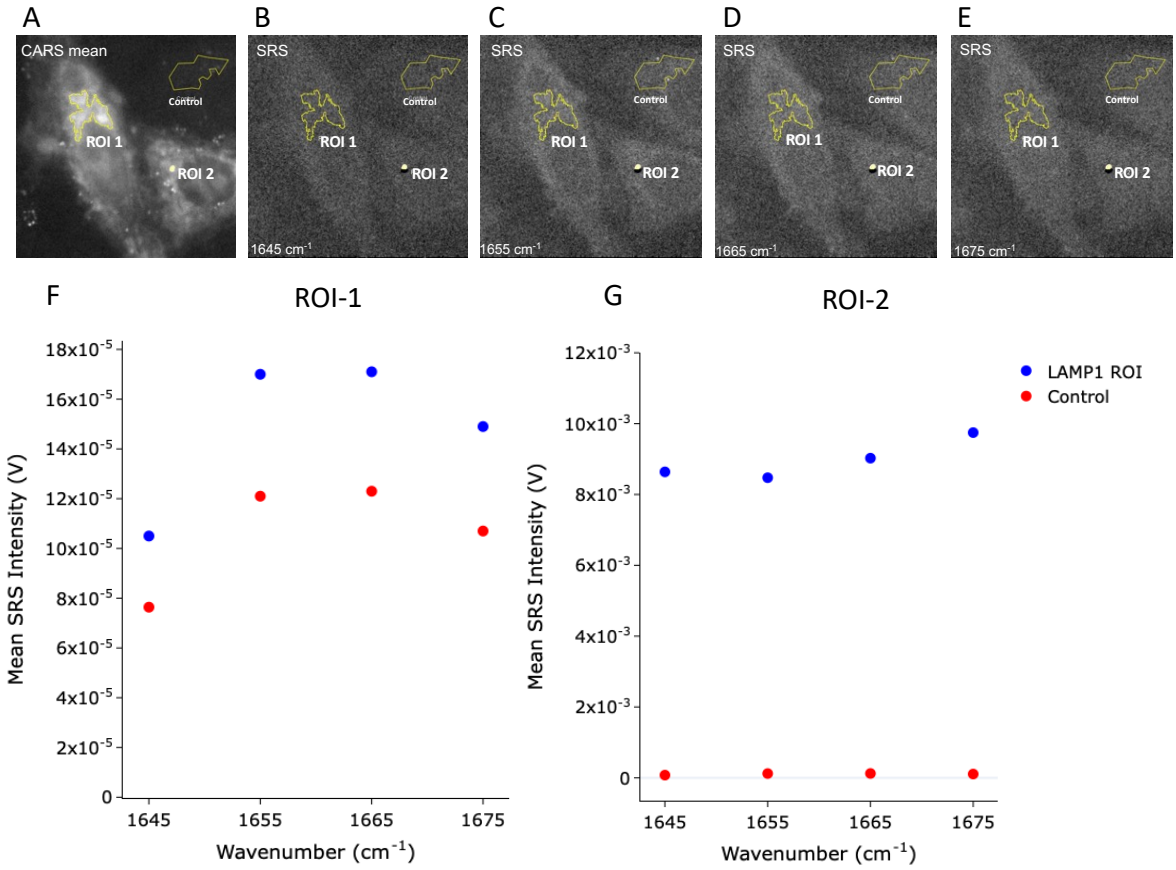


Figure 4.12: Differential signal analysis. (A) CARS and (B-E) SRS analysis highlighting two regions of interest (ROI). These were selected due to their high Raman intensity, with the control focussed on an area of no mRhubarb720 expression within a cell. (F-G) Differential analysis graphs are presented to show mean SRS intensities of both the ROI (blue) and control (red).

A clear artifact is observed in both CARS and all the SRS images, marked by a prominent high-intensity white signal with no variability (ROI-2). This was initially predicted to be a highly condensed lipid-rich droplet, due to the hydrocarbon composition of lipid bodies, which would have an increased intensity is expected at 1655-1665 cm^{-1}. However, this was not the case as intensity was high throughout each wavenumber measured.

4.2.4.3 Cell imaging of mRhubarb720-LifeAct fusions

The last target, LifeAct, was utilised to image the F-actin filaments within the HeLa cells. This protein is a small 17-amino acid peptide that selectively binds to hydrophobic pockets over two adjacent actin subunits. By fusing LifeAct to mRhubarb720, the aim was to

visualise filamentous structures within the cytoskeleton which is important for studying cellular processes such as cell motility and cell division (Figure 4.13).

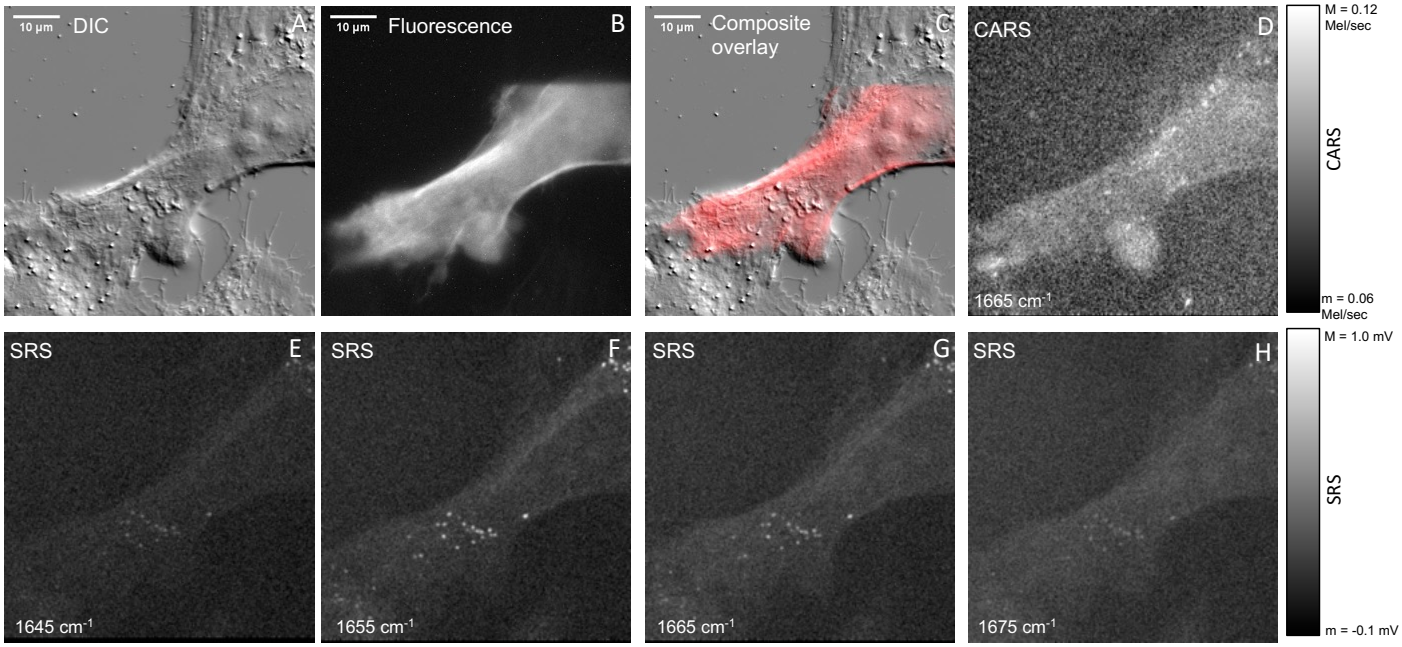


Figure 4.13: Mammalian cell imaging of LifeAct-mRhubarb720. (A) DIC, (B) fluorescence and (C) composite overlay images confirm LifeAct-mRhubarb720 expression within HeLa cells. (D) CARS measured solely at 1665 cm^{-1}, whereas (E-H) SRS was measured across four wavenumbers; 1645-1675 cm^{-1}.

The expectation was to detect filamentous structures that outline the F-actin; however, this was not clearly observed. The fluorescence does confirm mRhubarb720 to be present within cells but even the fluorescence has a lack of fibrous arrangements, as signal is omnipresent without any clear structure. Since CARS detects photoluminescence, the signal was not clear either and could not perceive the F-actin filaments. The SRS also shows no structures analogous to F-actin, however, observations show a number of lipid droplets enhanced at 1655-1665 cm^{-1}.

Due to LifeAct being an indirect probe for F-actin, it will compete with other actin-binding proteins such as myosin and cofilin which will disrupt actin dynamics and the overall structure and in turn disrupt cell morphology (Belyy et al. 2020) Alternatively, mRhubarb720 may cause a steric clash with the binding of the actin subunits, as Figure 4.6.C illustrates the close interaction between the mRhubarb720 probe with the actin targets, with the subunits tightly packed together. The cytoskeleton also regulates proteins such as perilipins,

important in facilitating lipid droplet budding. Disruption of actin filaments can impair this process, leading to altered lipid droplet formation and increased lipid storage (Olzmann and Carvalho 2019; Eynaudi et al. 2021). The SRS imaging in Figure 4.13.E-H supports this hypothesis, showing that the LifeAct-mRhubarb720 fusion exhibited the highest number of lipid droplets among the three fusion proteins explored in this study. The lack of visible actin-like structures in the images further reinforces the idea that actin disruption promotes lipid storage and suggests that the mRhubarb720 probe may not have been optimally designed to support proper actin formation.

4.2.5 Photobleaching limitations

Both mammalian and bacterial cell imaging generated photobleached samples when taking multiple scans per wavenumber, as evidenced in Figure 4.7.G-I. Originally, the process involved taking five scans and averaging the signal per wavenumber, therefore, imaging four wavenumbers would comprise of 20 scans. The photobleaching phenomenon observed in cell imaging was not evident in the *in vitro* samples. This is attributable to the proteins being in solution, which allows fast exchange of fluorophores thus avoiding local bleaching.

To understand and quantify photobleaching, single-cell bacteria were chosen as the focus due to their high expression levels and the uniformity of their signal, making it easier to quantify compared to mammalian cells, where the signal is localised to specific cellular components. This presented its own challenges, as the single-cell bacteria were freely floating in solution and highly mobile, causing difficulty in tracking cells over extended periods of imaging. To highlight the extent of the bleaching, it was decided to measure the fluorescence before and after 20 CARS scans, which involves averaging five scans of each wavenumber (1645, 1655, 1665 and 1675 cm^{-1}) (Figure 4.14).

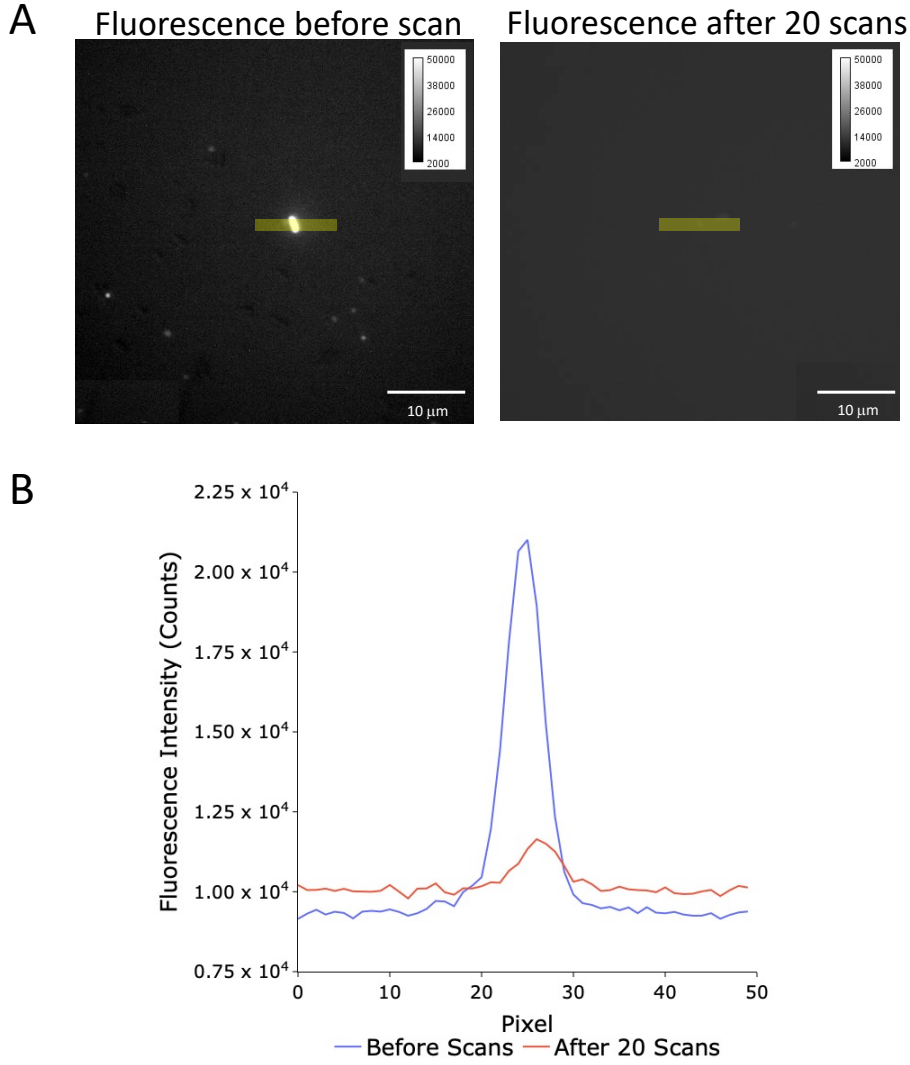


Figure 4.14: Fluorescence on single-cell *E.coli* to measure photobleaching. (A) Fluorescence imaging (excitation: 701 nm, emission: 720 nm) before the CRS scan and after 20 scans. The yellow highlight specifies the area used for the fluorescence intensity traces (summed vertically), plotted via ImageJ. (B) Plot of fluorescence intensity measured across selected pixels before and after CARS scans.

After 20 CARS scans, a significant reduction in fluorescence intensity was observed, with approximately 86% of the bacterial fluorescence being lost. Since the CARS signal was analogous to fluorescence, this allowed for the measurement of photobleaching per scan. To achieve this, repeated measurements were taken at 1665 cm^{-1}, as this corresponded to the peak signal observed from the SRS in mammalian cell imaging (Figure 4.15).

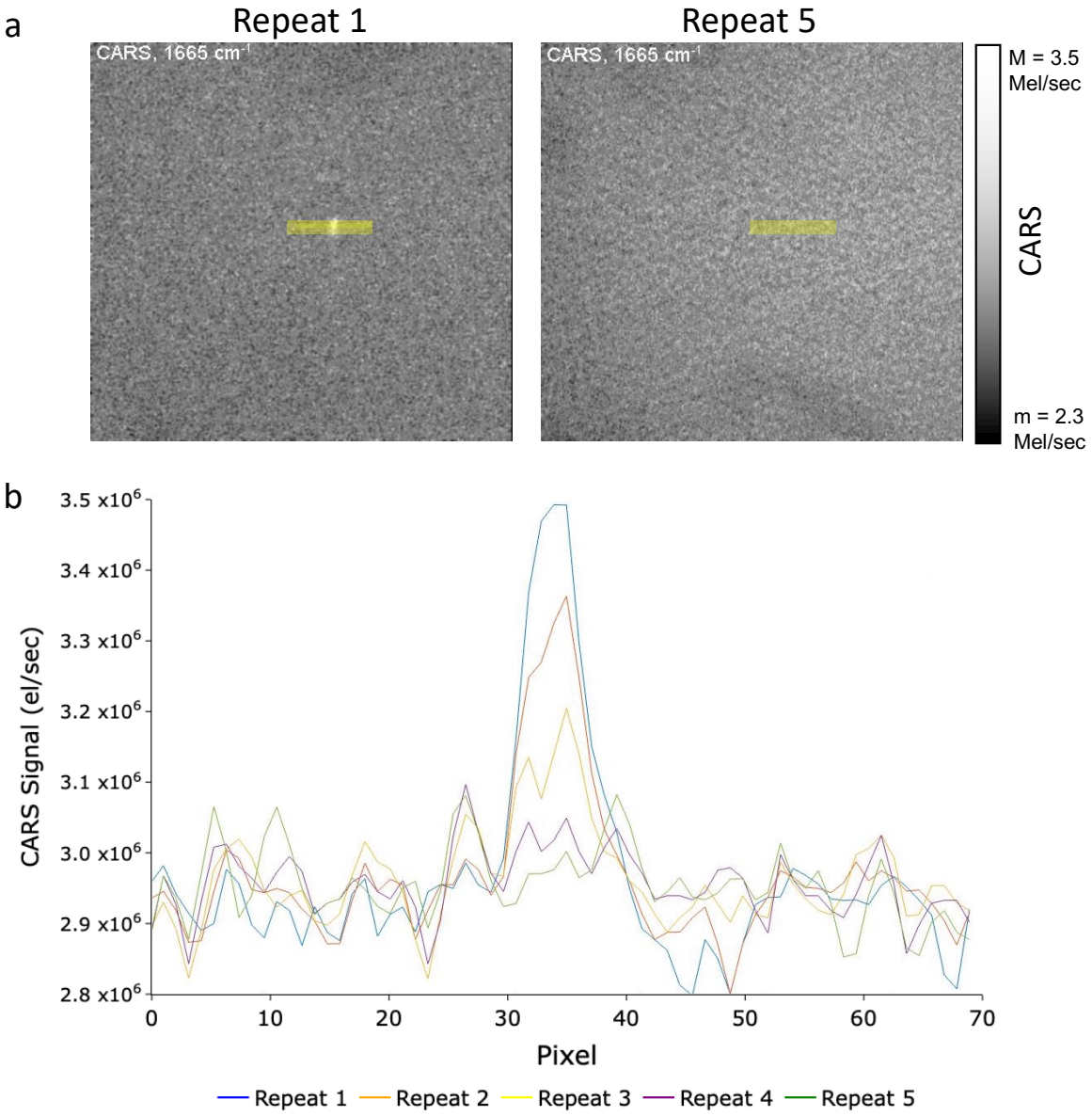


Figure 4.15: CARS on single-cell *E.coli* to measure photobleaching. (a) CARS images from Repeat 1 (left) compared to Repeat 5 (right). The yellow highlight specifies area calculated for fluorescence intensity, plotted via ImageJ. (b) The CARS signal intensity for all five repeats measured at 1665 cm^{-1}.

The CARS is observed to be predominantly bleached by the fourth repeat, with a notable reduction in intensity per subsequent repeat. The homogeneous background is non-resonant CARS from the coverslip and water medium. The near-IR FPs bound to BV are known to undergo photoisomerisation between the C and D ring of the BV (C15-C16) when irradiated with far red, near-IR photons (Yang et al. 2009). The photoisomerised state (Pfr) has no known emission wavelength when excited as the energy is dissipated through rotational movement back to the ground state (Pr). Given that the current microscope set-up

employs lasers within this spectral region, it is predicted that this photoconversion contributes to the observed photobleaching. This conversion is fully reversible, either through dark reversion or rapidly upon re-irradiation of near-IR photons (Stepanenko et al. 2017).

4.3 Conclusions

In conclusion, first steps to genetically encoding probes for use in cellular based Raman imaging has been demonstrated. The relatively narrow spectral linewidths of these probes, coupled with the electronic pre-resonant effect, with signal enhancements comparable to what shown in the literature for synthetic ATTO dyes, open up a promising avenue towards super-multiplexing with genetic encoding. A genetically encoded Raman-active probe was applied and imaged in mammalian cells, resulting in a significantly enhanced SRS signal in H2B-mRhubarb720 fusions. This represents a novel advancement in the microscopy field, potentially aiding in overcoming the limitations of conventional fluorescence microscopy. The detection of a signal in the fingerprint region of SRS is significantly influenced by the characteristics of near-IR FP probes. Analysing these probes, alongside free BV, reveals the importance of understanding vibrational frequencies for effective cell imaging. The distinct double peak observed in mRhubarb720 demonstrates its effectiveness as an SRS probe, while free-BV shows no significant SRS profile, highlighting the approach for genetically encoding Raman-active probes for cell imaging. The varying intensities of near-IR FPs like miRFP670nano3 and emiRFP670 further emphasise the proportionality between SRS intensity and pre-resonant excitation conditions. The red-shifted nature of mRhubarb720 absorbance makes it the most enhanced near-IRFP probe, highlighting the importance of tuning the pump beam to match the molecule's excitation energy. Further progress could be achieved for increasing SRS signal intensity for more blue-shifted near-IR FPs by catering the frequencies of the Pump/Stokes lasers to be closer in proximity to the pre-resonant excitation conditions of the probes. Overall, these findings suggest that near-IR FPs, particularly those with red-shifted absorbance like mRhubarb720, hold significant promise for enhancing the accuracy and effectiveness of SRS-based cell imaging.

An important finding in these experiments is that a signal persists even when the Stokes beam is turned off. This suggests that the detected photons in this condition do not arise from CARS, which requires both pump and Stokes beams, but rather from an alternative mechanism such as two-photon excited fluorescence. In contrast to CARS, which is a coherent four-wave mixing process, fluorescence arises from incoherent spontaneous emission following nonlinear absorption. Therefore, the persistence of signal in the absence of a Stokes beam implies that the system is undergoing two-photon absorption of the pump laser, resulting in fluorescence emission rather than a vibrationally resonant Raman signal. Additionally, there was no variability in CARS signal intensity when measuring across four wavenumbers when imaging the H2B-mRhubarb720 mammalian cells.

Both mammalian and bacterial cell imaging confirm that increasing the number of probes within the focal volume enhances SRS intensity. The H2B and *E.coli* cells presented increased intensities, in contrast to the LAMP1 and LifeAct samples, where the signal was indistinguishable to background noise. With improvements, SRS offers the potential for multiplexing due to its sharp spectral peaks (~20 cm^{-1}) and enables longer observation times due to the lower energy excitations applied to the cells. This represents a significant enhancement over current fluorescence-based microscopy techniques.

The next step is to further develop both the probes and the microscopy setups. Microscopy developments include SRS set-ups featuring pump and Stokes wavelengths in the visible spectrum, so that FPs with chromophores absorbing in the blue to red range can be used, as well as high-speed SRS multiplex acquisition, for detection of multiple probes. Probe developments include the incorporation of electronically coupled Raman bonds vibrating in the biologically silent region (e.g. via reprogrammed genetic code approaches, see Chapter 5 and 6) reducing non-specific background and further expanding the utility of such probes.

5 Experimental and computational analysis of para-cyano-phenylalanine incorporation in mRhubarb720: Investigating through-space enhancement, protein dynamics, and the electronic properties of Biliverdin XI-alpha.

5.1 Introduction

In the previous chapter, electronic pre-resonant stimulated Raman spectroscopy (epr-SRS) was successfully demonstrated to detect molecular vibrations of biliverdin XI-alpha (BV) when bound to mRhubarb720, including on fusion to target proteins in cells. The primary focus was in the fingerprint region (1550–1850 cm^{-1}), a well-established spectral range critical for identifying unique biomolecular structures through stimulated Raman scattering (SRS). This region is crucial for identifying common vibrational modes in biomolecules and provides valuable insights into protein structure (Adar, 2022; Cai et al., 2022). Here, the fingerprint region has been central for generating sharp, distinct spectral peaks, particularly on the C=C stretching within the protein-embedded far-red/infra-red chromophores that could for be the basis of next-generation imaging techniques, aiming to overcome the limitations of fluorescence microscopy (see Chapter 1.2.3 and Chapter 4.1).

To further advance this technology, the focus must shift to the biologically silent window (1800–2800 cm^{-1}), where minimal interference from endogenous biomolecules allows for the detection of specific labels; one route to achieve this is through the incorporation of non-canonical amino acids (ncAAs). These ncAAs can serve as Raman-active vibrational tags, enabling precise monitoring of molecular interactions, similar to BV, if vibrationally coupled to the electronically excited chromophore via through-bond or potentially through-space enhancement. It is predicted that vibrational interactions between non-bonded functional groups can lead to an increase of spectroscopic signals, such as Raman scattering. While through-space enhancement has not been extensively studied in epr-SRS, it has been demonstrated in surface-enhanced Raman spectroscopy (SERS) (Caligiuri et al. 2024). For example, graphene-based systems have shown such enhancement

via π-π stacking and dipole-dipole interactions, where the spatial configuration of molecules relative to the 2D carbon material has significantly impacted the Raman signal (Barros and Dresselhaus 2014; Pérez-Jiménez et al. 2020). These interactions boost the local electromagnetic field, amplifying the Raman signal by orders of magnitude. This concept could be further extended to non-covalent interactions between ncAAs and nearby chromophores to improve detection in the biologically silent window, by leveraging the electronic resonant Raman effect using the Pump/Stokes laser system. If possible, this will achieve more sensitive detection of ncAAs in complex biological environments.

The ncAA of choice for this chapter is para-cyano-phenylalanine (pCNPhe), which contains a nitrile (-C≡N) group at the 4-position of the phenyl ring in its aromatic R-group (Tucker et al. 2006). This nitrile moiety, when bound to phenylalanine, exhibits a strong and isolated Raman band at ~2240 cm^{-1}, falling within the biologically silent window (Weeks et al. 2008). The combination of the aromatic group and the cyano groups vibrational characteristics, makes pCNPhe an ideal probe for integrating into protein systems without significantly perturbing the protein structure. This is particularly important given that BV incorporation is not well understood, as the addition of bulkier moieties could potentially inhibit the interaction between BV and the protein. Therefore, pCNPhe's minimal steric influence should ensure that it serves as an effective vibrational tag without disrupting essential protein-BV interactions, particularly when replacing similarly sized mutational targets (e.g. tyrosine and phenylalanine).

Integrating ncAAs into proteins involve reprogramming the genetic code to recognise the amber stop codon (TAG) in order to incorporate the desired ncAA (explained in more detail in Chapter 1.3.5 and Methods Section 2.2). This process involves an engineered orthogonal tRNA and aminoacyl-tRNA synthase pair that specifically recognises the ncAA and TAG codon (Miyake-Stoner et al. 2009). The reprogrammed stop codon allows for site-specific incorporation of the ncAA, enabling the introduction of novel chemical functions into proteins without interfering with their natural expression pathways.

In this chapter, molecular dynamics (MD) simulations will be employed to understand the structural stability and dynamics of mRhubarb720 when incorporating pCNPhe close to BV. These simulations will provide atomic-level insights into how pCNPhe influences the local protein structure and overall conformational flexibility. Using GROMACs, a freely-available simulator optimised for large biomolecular systems, simulations can be carried out with

high-performance computational capabilities (Abraham et al. 2015). The simulations will apply the CHARMM force field parameters to model non-canonical biomolecules, such as BV and pCNPhe, with high precision (Bjelkmar et al. 2010). These simulations will help identify any structural perturbations caused by the pCNPhe and will be compared to the experimental data generated to determine any trends between the proteins dynamics within the simulation and any spectroscopic shifts observed. Identifying such trends could inform future redesign efforts, enabling a more efficient workflow that saves time and resources. A more in-depth description of the MD approach can be viewed in Chapter 1.4.2.

Additionally, quantum mechanical (QM) simulations will be utilised to compare the electronic properties and vibrational frequencies of BV when bound mRhubarb720 and free-BV in solution. The QM approach provides highly accurate predictions of properties such as dipole moments, excitation energies, absorption profiles and most importantly, the predicted Raman spectrums. The Becke, 3-parameter, Lee-Yang-Parr (B3LYP) functional combined with the 6-31G+(d,p) basis set is applied to these simulations (Lee et al. 1988). The B3LYP is a hybrid density functional theory (DFT) method, well-known for its balanced accuracy and computational efficiency, is particularly useful for studying electronic properties in biomolecular systems. The 6-31G+(d,p) basis set includes polarisation (+) and diffuse functions (d,p) which are important for accurately modelling electronic distribution in systems that exhibit a charge separation (Clark et al. 1983). This combination should reliably predict the electronic properties of large biomolecules like BV, in sufficient simulation times. The full approach and the parameters utilised was discussed in Chapter 1.4.3 and Chapter 2.7.

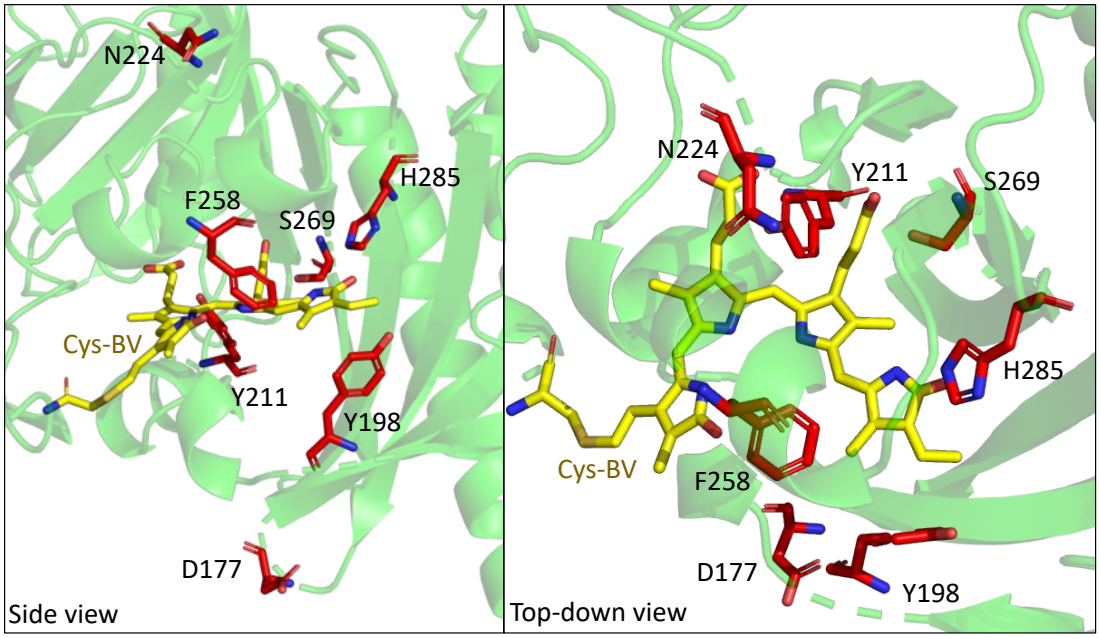
The primary aim of this research chapter is to assess whether pCNPhe incorporation can induce through-space enhancement within the biologically silent region if coupled to BV in mRhubarb720. An investigation on the structural and dynamic consequences of pCNPhe incorporation using MD simulations will follow, analysing how its impacts the BV's structural behaviour and interactions within the protein environment. Additionally, a QM approach is employed to explore the electronic properties of BV, aiming to identify the optical properties observed between mRhubarb720-bound BV and free-BV. Finally, these simulations aim to distinguish alternative vibrational modes beyond the experimentally studied 1550-1850 cm^{-1} wavenumber range that could be targeted for future epr-RS imaging using mRhubarb720.

5.2 Results and discussion

5.2.1 Analysis of incorporating pCNPhe into the BV-binding environment

When designing Raman-active probes, the ability to detect vibrationally active bonds that are not located within the electronically active chromophore allows for greater flexibility in probing specific molecular environments. By introducing ncAA's through site-directed mutagenesis, different types of ncAAs can be incorporated, each offering distinct vibrational signatures. This enables the exploration of varied vibrational modes across different regions of the cell or molecular system, without disrupting the chromophore's structure or function while also enabling multiplexing during epr-SRS. For successful detection, these bonds should be in close proximity to the delocalised system of the chromophore and enhancement should fall within the biologically silent window, as this improves the likelihood of detecting the bonds by minimising interference from the surrounding biological material. To achieve this, specific mutational targets were selected to be close in space to BV, and pCNPhe was incorporated at these positions, enabling potential electronic interaction and facilitate effective through-space vibrational coupling (Figure 5.1). A number of mutational sites were targeted to gain a full assessment of the BV cofactor environment. By introducing mutations at various strategic points around the BV cofactor, the focus is on determining whether through-space enhancement occurs at different locations. In addition, this approach facilitates a comprehensive investigation into how different regions of the surrounding protein matrix influence its structural and functional properties, offering further insights into interactions that may affect the chromophores behaviour.

A



B

C

Residue	Rationale
D177	Control mutation placed over 13 Å away from the BV, to reduce the likelihood of interfering with BV incorporation.
Y198	Aromatic structure similar to pCNPhe and close to D-ring.
Y211	Nitrile group will be situated between both carboxyl group of BV.
N224	Another control mutation, above alpha helix supporting BV (16 Å away).
F258	Aromatic structure similar to pCNPhe and situated in the alpha helix supporting BV, with nitril placed between C and D-rings.
S269	Nitrile is placed directly above C-ring.
H285	Direct H-bond interaction with D-ring, nitrile would insert above ring in this given conformation.

Figure 5.1: Schematic and rationale of mutational targets in mRhubarb720. (A) Side view and (B) top-down view of the BV cofactor environment. Cys-BV (yellow) and mutational targets (red) are represented as sticks with the rest of the protein scaffold as a transparent cartoon. (C) Rationale of mutational targets of incorporating pCNPhe.

Each residue highlighted in red was mutated in order to incorporate the pCNPhe ncAA. The selection of these targets was based on positioning the nitrile group close to the BV cofactor (Y198, Y211, F258, S269, and H285; see Figure 5.1.C), while also including two targets further from the cofactors cavity (D177 and N224). Given that BV incorporation is a post-translational modification, it was crucial to ensure that these modifications do not

interfere with this process. Tyrosine and phenylalanine residues were specifically targeted because their aromatic structures are analogous to pCNPhe, which is most likely to maintain the integrity of the protein scaffold. To evaluate this, each variant was subjected to a DC protein concentration assay, independently of BV incorporation, and standardised to the 280 nm peak for the absorbance. The 280 nm values were used to standardise the samples for comparison, despite the potential minor impact of ncAAs on this wavelength. The spectral properties of each variant were then measured and compared against the WT (Figure 5.2). The observations indicate that all variants exhibit significantly lower molar extinction coefficients (EC) than the WT, but since these values may represent a mix of holo- and apo-forms, they are not fully accurate. Furthermore, the fluorescence properties of each variant did not deviate significantly from the WT's emission maxima.

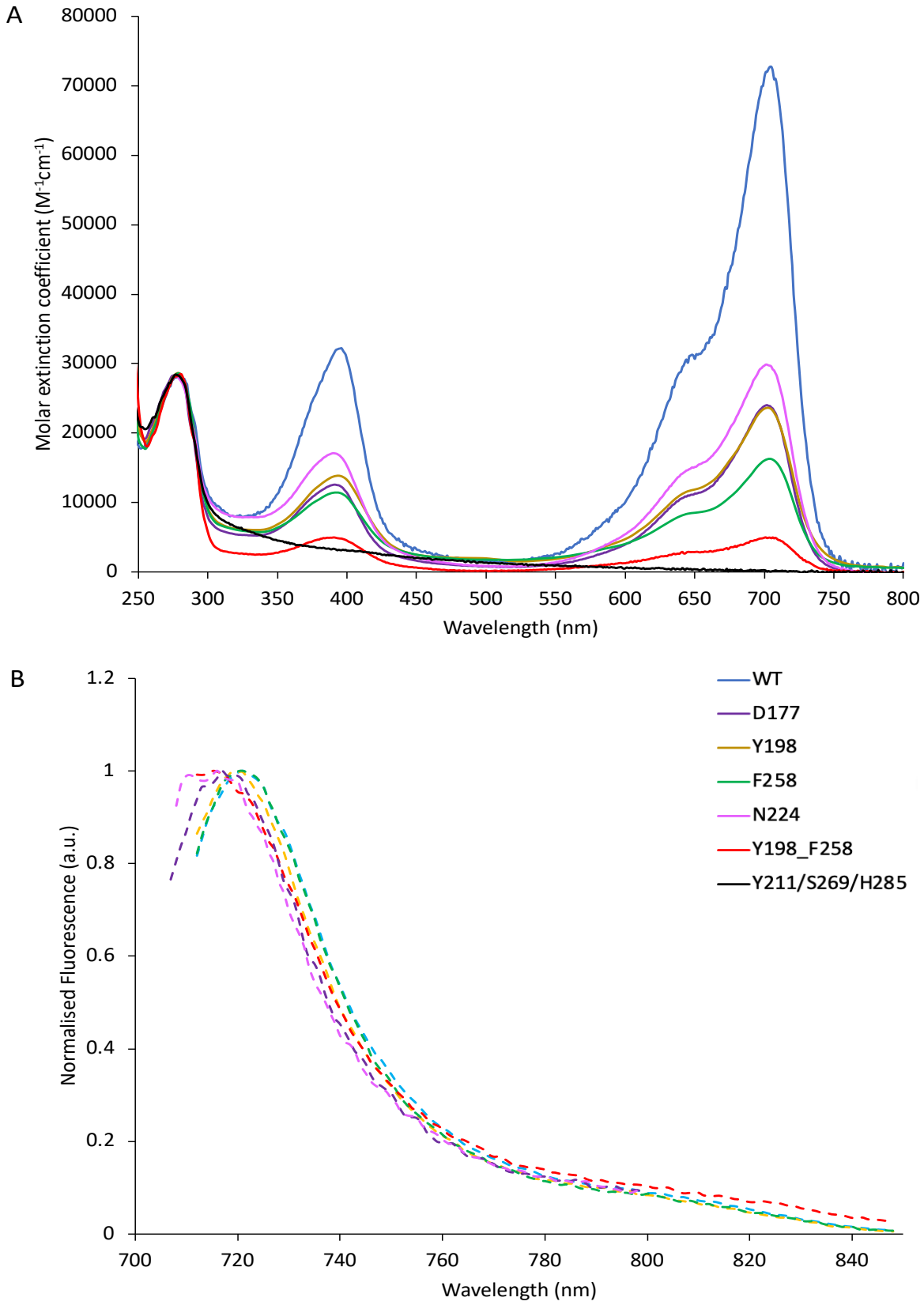


Figure 5.2: Spectral properties of mRhubarb720 variants. (A) Absorbance spectra, in respect to the predicted extinction coefficients and (B) fluorescence emission, normalised to the wavelength maxima. Proteins were diluted to 5 µM for each measurement.

Three mutations (Y211/S269/H285) did not result in any absorptive or fluorescent properties, likely due to their close proximity to the BV-cofactor, which may be preventing BV incorporation and possibly maturation. The peaks observed around ~395 nm is indictive of the BV-binding mechanism, as described in Chapter 3.2.2, suggesting that BV incorporation is reduced in all mutants. The double mutant Y198_F258 exhibits the lowest EC across the scan, which is expected given the presence of two mutations in the BV environment. Fluorescence was normalised to their respective peaks, as intensity was proportional to their absorptive properties. Although the peaks varied slightly, no shifts larger than 5 nm were observed (Table 5.1).

Table 5.1: Summary of spectral data for mRhubarb720 variants. Excitation (Ex), emission (Em), *molar extinction coefficient (EC), quantum yield (QY) and *brightness.

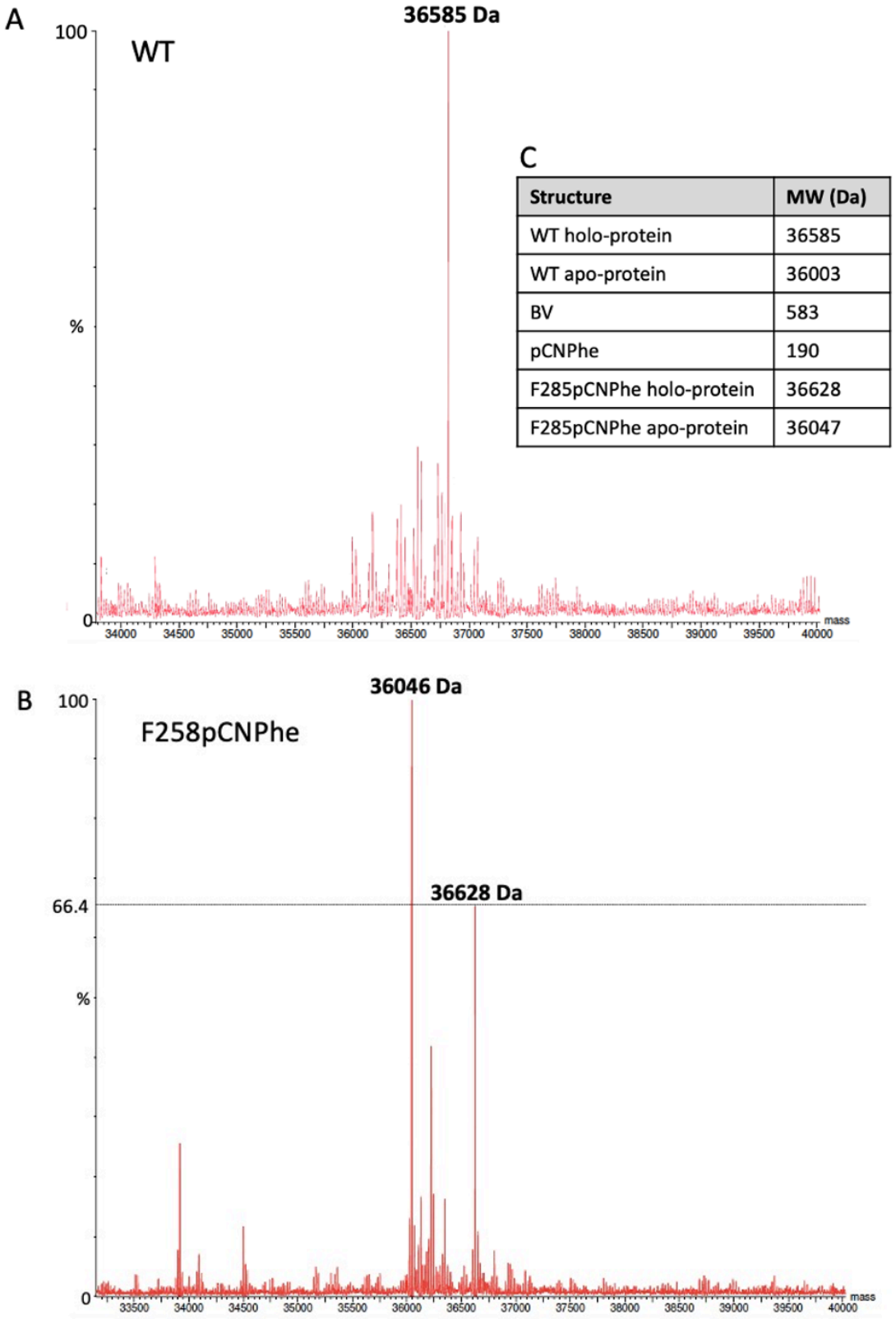
Mutation	Ex λ (nm)	Em λ (nm)	EC ($M^{-1}cm^{-1}$)	QY (%)	Brightness ($M^{-1}cm^{-1}$)
WT	703	720	72,738	6.46	4.70
D177	703	717	23,997	6.43	1.54
Y198	702	719	23,487	6.74	1.58
N224	703	716	29,697	6.35	1.89
F258	702	721	16,214	6.79	1.10
Y198_F258	703	715	4,948	6.18	0.31
Y211/S269/H285	Non-Fluorescent				

*These values are presumed to be homogenous holo-protein samples.

Calculating the brightness of a FP involves multiplying the EC by the QY. The resulting value, expressed in inverse molarity per centimetre ($M^{-1}cm^{-1}$), reflects the proteins ability to absorb and emit light. This implies that all mutations disrupt the function of the FP, by reducing the EC, which in turn decreases the overall brightness of the FP. As outlined in the Chapter 2.3.4, determining the QY for the variants required the use of a reference. Ideally, this reference would have been a near-IR dye with fluorescent peaks similar to mRhubarb720-WT. However, the WT itself was used as the reference, relying on the QY value reported in literature (Rogers et al. 2019). Since this approach depends on the accuracy of the reported value, the QY values determined were relative to the WT rather

than absolute measurements. As discussed in Chapter 3, reported values are not always reliable. The excitation peaks for each variant displayed non-significant differences compared to the WT, with a maximum blueshift of 1 nm observed in the Y198 and F258 mutants. This shift could be attributed to experimental error, as the double mutant Y198_F258 exhibits an excitation peak analogous to that of the WT.

To investigate the cause of the differences in EC, it was decided to submit the WT plus one of the mutants for mass spectrometry analysis to identify the masses within the samples. The F258 mutant was chosen for analysis as it displayed the lowest EC compared to WT for a single mutation (Figure 5.3).



C

Structure	MW (Da)
WT holo-protein	36585
WT apo-protein	36003
BV	583
pCNPhe	190
F285pCNPhe holo-protein	36628
F285pCNPhe apo-protein	36047

Figure 5.3: Mass spectrometry of WT and F258 sample. (A) WT sample and (B) F258pCNPhe sample with the mass peak for the F258 sample is normalised to the largest peak at 100%, the two largest peaks are annotated. (C) Calculated mass of proteins (via ExPASy) and reported masses of pCNPhe and BV.

The largest F258 peak was measured at 36,046 Da, which is characteristic of mRhubarb720 containing a nitrile group without the presence of BV. The second largest peak measured at 36,628 Da represents mRhubarb720 with the nitrile group as well as the inclusion of BV (out by 1 Da). For comparison, mRhubarb-WT has a predicted mass of 36,585 Da. Mass prediction was carried out using the ExPASy ProtParam tool, adding BV (583 Da) and subtracting one hydrogen (1 Da) for WT. For the F258pCNPhe mutant, the mass was predicted by substituting pCNPhe with phenylalanine, adding the mass of the nitrile group (26 Da), the BV (583 Da), and then subtracting the mass of two hydrogens (2 Da; from the BV covalent linkage and the replacement of hydrogen by the nitrile). This indicates that the sample is predominantly composed of apo-protein.

By normalising the percentage of these peaks, it is possible to estimate the relative proportions of apo- and holo-protein within the sample at the time of measurement. Specifically, the apo-protein is calculated to be at most 60% of the sample contents, while the holo-protein makes up to 40% – based solely on these peaks. However, given the presence of additional other peaks in this sample (due to possible protein degradation), these calculated values represent the upper limits, and actual ratios may be lower. Another method to determine these ratios is by calculating the 280/395 nm ratio of peaks from Figure 5.2.A, where the F258 holo-protein is estimated to be 35% of the overall sample, with the remaining 65% likely representing F258 apo-protein.

5.2.2 Epr-SRS analysis of selected pCNPhe mutations

In this section, the SRS spectra of mRhubarb720 was analysed with the incorporation of the pCNPhe at key positions within the proteins structure. The aim was to investigate through-space enhancement of the nitrile group and whether it was possible to observe features within the biologically silent window. The SRS experiments were conducted on the pCNPhe in solution, WT and three key mutants (D177, Y198 and F258). The D177 was selected as a control, positioned further away from the BV-binding pocket, where no significant electronic or vibrational interactions is expected. On the other hand, Y198 and F258 was chosen due to their location near the BV cofactor, allowing for potential through-space vibrational coupling. The following SRS spectra were collected thanks to Dr David Regan.

Firstly, the fingerprint region (1550-1950 cm^{-1}) was investigated in order to examine the effect of pCNPhe incorporation on the BV double peak (1627-1652 cm^{-1}), previously discussed in Chapter 4.2.2. This analysis provides insight into how vibrational interactions are altered by the proximity of pCNPhe to the BV, however, no new peaks were expected in this Raman spectral range (Figure 5.4).

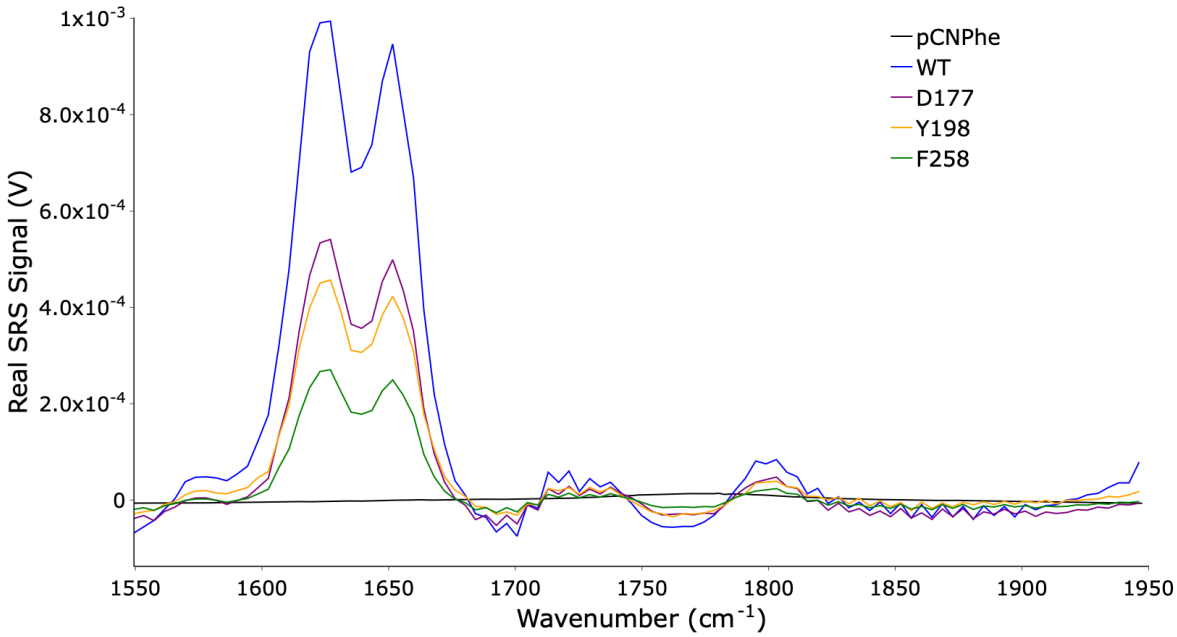


Figure 5.4: Background subtracted SRS fingerprint scan of pCNPhe and selected variants. Each protein variant, WT (blue), D177 (purple), Y198 (yellow) and F258 (green), was measured at the sample concentration of 1 mM (apo-/holo- mixture for mutants). The pCNPhe (black) was also measured at 1 mM. All protein variants were solubilised in 50 mM Tris, whereas pCNPhe was in 50 mM Tris and 50 mM NaOH.

The notable difference between WT and the mutants is the intensity of their distinctive BV double peaks at 1627-1652 cm^{-1}. As previously mentioned, the apo-/holo-protein ratio is most likely central to this finding, as incorporating pCNPhe at different positions will change proportion of BV-binding. Since each mutant has peaks at 1627-1652 cm^{-1}, the inclusion of pCNPhe is not changing the vibrational frequencies of BV. The reduction in signal intensity can be quantified where D177 is 46% weaker than the WT, Y198 and F258 exhibited more substantial reductions with 54% and 73%, respectively. These reductions correlate with the absorption results (Figure 5.2.A), albeit not at the same

proportion. This aligns with the original hypothesis that pCNPhe incorporation would primarily reduce the vibrational signal output, as a function of total mRhubarb720 protein concentration, rather than introducing new peaks or altering the current BV double peak observed in WT. The pCNPhe in solution provides context for understanding the baseline Raman signal within the fingerprint region, as the pCNPhe curve shows insignificant signal, particularly around the 1627-1652 cm^{-1} BV peaks.

The biologically silent window in the SRS spectrum is a particular region of interest, as it typically lacks endogenous Raman-active modes from biomolecules, making it ideal for detecting ncAA's like pCNPhe, if through-space coupling can be achieved. However, there is a lack of significant enhancement from incorporated pCNPhe in this region (Figure 5.5). The same concentrations for the ncAA and variants were measured, with pCNPhe in solution displaying a spectral peak at 2244 cm^{-1} with an intensity of 0.029 mV. The variants did not exhibit any sharp or significant enhancement at this wavenumber, suggesting that pCNPhe is not actively contributing to the vibrational coupling.

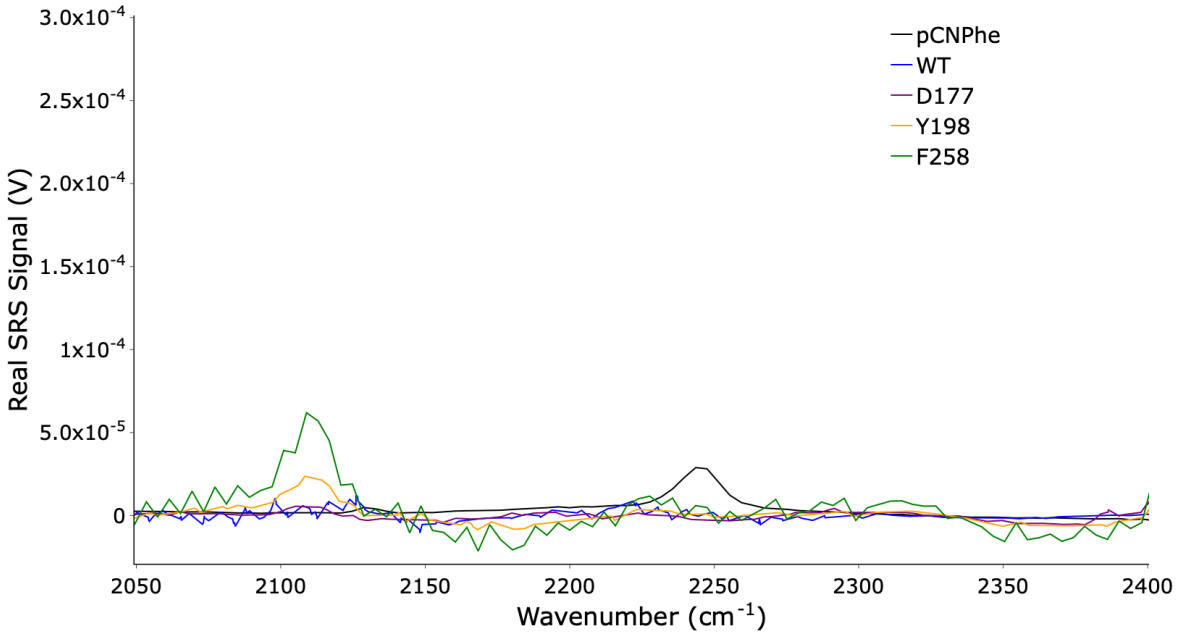


Figure 5.5: Background subtracted SRS biological silent window scan of pCNPhe and selected variants. Each protein mutant, WT (blue), D177 (purple), Y198 (yellow) and F258 (green), was measured at 1 mM, and the pCNPhe (black) was also measured at 1 mM. All protein variants were solubilised in 50 mM Tris, whereas pCNPhe was in 50 mM Tris and 50 mM NaOH.

Despite the proximity of Y198 and F258 to BV, the minimal signals suggest that these mutations are not capable of through-space enhancement. This was expected for D177 due to the position of the mutation. Therefore, in future experiments it may be required to incorporate the Raman-active bond within the electronically excited chromophore in order to promote molecular vibrations in this region while performing epr-RS. One factor influencing the observed minimal signals of the variants is the percentage of nitrile bonds present. In the pCNPhe sample, every residue contains a nitrile group (190 Da per residue), whereas in the mRhubarb720 protein with pCNPhe, there is only one nitrile group per entire protein (37 kDa). Although the samples have the same concentration, the percentage of nitrile bonds in the protein sample is significantly lower, which reduces the signal intensity in SRS measurements. Additionally, signal interference or background may also be playing a role. The strong fluorescence from mRhubarb or nearby chromophores could potentially overwhelm or mask the nitrile signal via two-photon absorption, making it harder to detect in the protein-bound system (see Chapter 4.2.1 for more information). The peak observed around 2110 cm^{-1}, more prominently for the Y198 and F258 samples, can be attributed to a nitrogen gas artefact (Murdachaew et al. 2014). The intensity and position of this nitrogen peak can vary based on the optical pathway in the system. Since the variants will differ in how the light and sample interacts with the nitrogen present in the optical path, it can cause subtle changes in the wavenumber and intensity of the signal (Sands et al. 2007).

5.2.3 Molecular dynamics analysis of mRhubarb720 and their pCNPhe variants

Each mutation of mRhubarb720 resulted in detrimental effects on its spectral properties, which posed a significant challenge since maintaining a high molar extinction coefficient is critical for enhancing the Raman peaks in the fingerprint and biologically silent region via potential through-space vibrational coupling. The aim of these *in silico* experiments was two-fold: first, to understand the mechanisms underlying protein stability of the mRhubarb720 apo-protein and the incorporation of ncAAs; and second, to establish a computational workflow for analysing mutations prior to *in vitro* experiments in the future, helping to reduce both time and cost of implementing the experimental work.

Various analysis was performed on the protein as well as the BV, including the Root Mean Squared Deviation (RMSD), Root Mean Squared Fluctuation (RMSF), number of hydrogen bonds between BV and the surrounding environment, Surface Accessible Surface Area (SASA) and Radius of Gyration (Rg). The RMSD is used to quantify structural differences by comparing individual conformations to the mean structure across the simulation, providing insights into how much each frame deviates from this average. Lower RMSD values indicate the structure remains close to the mean conformation, whereas higher values suggest conformational shifts from the reference over time. RMSF, while related, specifically measures the flexibility of individual residues (for the protein) or atoms (for the BV), by tracking their fluctuations throughout the simulation. The total number of hydrogen bonds between BV and its environment, and SASA will mainly assess the solvent interactions of BV. Lastly, the Rg assesses the distribution of atoms relative to the molecules centre of mass, where a lower Rg indicates a more compact structure, and a higher Rg suggests an expanded conformation, potentially a sign of protein unfolding. Using these analyses will help assess how the incorporation of pCNPhe affects the structural stability and dynamics of both the protein and BV, providing insights into any potential destabilisation or conformational changes caused by the mutations.

5.2.3.1 mRhubarb720-WT

The WT was analysed first to establish a baseline for comparison with the variants, as it exhibited the most favourable spectral properties due to the sample being homogeneous with holo-protein, whereas the mutants contained a mixture of holo- and apo-protein. The RMSD analysis for the WT revealed that its structure remained relatively stable throughout the simulation, with only minor deviations from the mean conformation (Figure 5.6). All simulations indicated that the WT maintained its structural integrity effectively, as there were no unexpected spikes or abrupt changes in RMSD. This consistent stability provides a valuable reference for evaluating the effects of mutations on the proteins structural dynamics and stability, as detailed in the following sections.

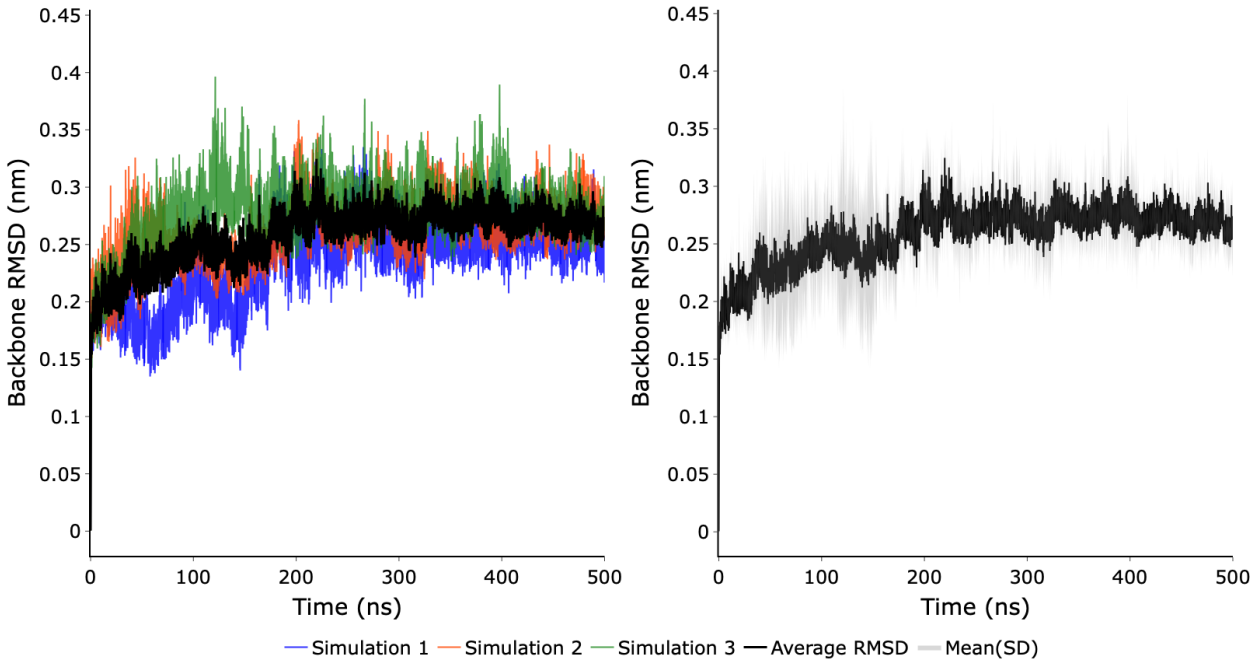


Figure 5.6: mRhubarb720 WT RMSD analysis over three simulations. (A) Each simulations RMSD: Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green), with the average RMSD (black). (B) The average RMSD with the standard deviation (light grey).

The average backbone RMSD across all the simulations ranged from 0.2-0.3 nm, with deviation of ± 0.05 nm, suggesting that the protein maintains a stable conformation without large-scale conformational changes or unfolding. Simulations that show larger RMSD values, such as over 0.4 nm (when assessing backbone deviations), often correspond to significant confirmational shifts, partial unfolding, or other structural instabilities (Shukla and Tripathi 2020).

The BV was also measured, with the RMSD computed on all atoms of the ligand, offering insights into its conformational changes during the simulations (Figure 5.7). The overall fluctuations range between 0.07-0.2 nm, indicating moderate movement within the cofactors binding site, where the ligand remains stable across all three simulations.

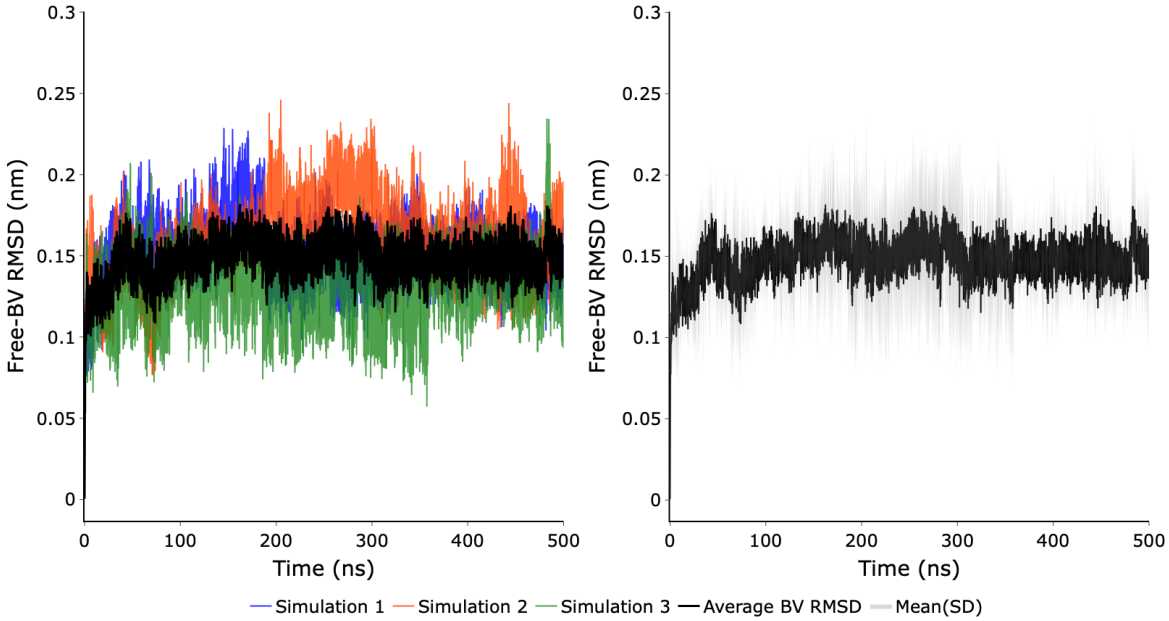


Figure 5.7: mRhubarb720 WT Biliverdin XI-alpha RMSD analysis over three simulations. (A) Each simulations RMSD: Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green), with the average RMSD (black). (B) The average RMSD with the standard deviation (light grey).

The RMSD stabilises between 0.11-0.17 nm, with consistent behaviour across the simulations which is reflected in the minimal standard deviation (SD), observed in Figure 5.7.B. This indicates that the cofactor position, relative to the protein scaffold, converges in all simulations. The low RMSD values are typical of a well-structured protein-ligand complex, implying that BV does not undergo any significant conformational changes or displacement during the course of the simulations. This will serve as a benchmark for comparison with other variants, allowing for insight into how the pCNPhe mutations might affect BV stability and binding efficiency.

As mentioned, the C-α RMSF profile analyses the flexibility of each residue within the protein, which depicts the fluctuations of regions within the mRhubarb720 WT. Each simulation was measured, as seen in Figure 5.8.A, where most residues exhibit RMSF values below 2.5 Å, indicating relatively low fluctuation for the majority of the protein. However, several regions particularly around residues 75-90, 195-205 as well as the C-/N-terminal domains exceed this threshold, showing higher RMSF values with fluctuations reaching up to 3.7 Å on average over the three simulations (Figure 5.8.B).

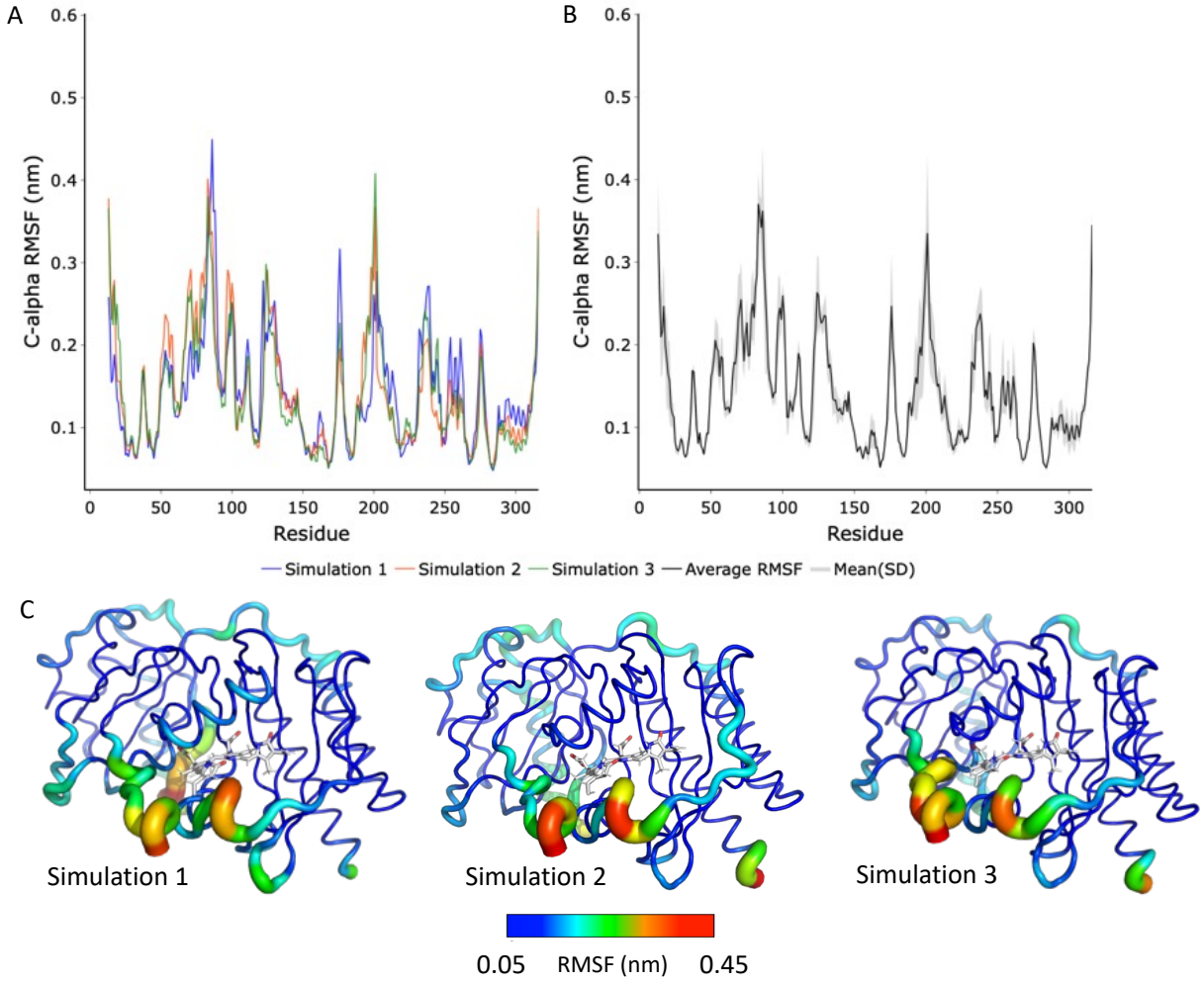


Figure 5.8: mRhubarb720 WT C-α RMSF analysis over three simulations. (A) RMSF of individual simulations: Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green). (B) The average RMSF of all simulations (black) with the standard deviation (light grey). (C) B-factor putty representation of each simulation coloured by residue fluctuation from minimum (blue) to maximum (red) illustrated by the colour scale. BV is depicted by grey ball and sticks in each structure.

The average RMSF trends are consistent across those observed in the individual simulations, indicating the reproducibility of the data as highlighted by the standard deviation. One region that displays higher RMSF comprises residues 198-205, close to the BV-binding pocket, implies a highly flexible and potentially functionally dynamic region of the protein. The B-factor putty representations maps the RMSF values onto the protein structure, as observed in Figure 5.8.C, which highlights these residues reside below the BV within mRhubarb720. Further analysis of the simulation data reveals that residues 201-203 (S201, D202, I203) initially remain situated beneath the BV chromophore for the first 180 ns.

However, between 180 and 500 ns, these residues move away, suggesting a dynamic rearrangement of the local protein structure around BV (Figure 5.9). Despite the proximity of these residues to the chromophore, no direct hydrogen bonding or strong electrostatic interactions are observed, indicating that their initial position may be driven more by steric factors or solvent interactions.

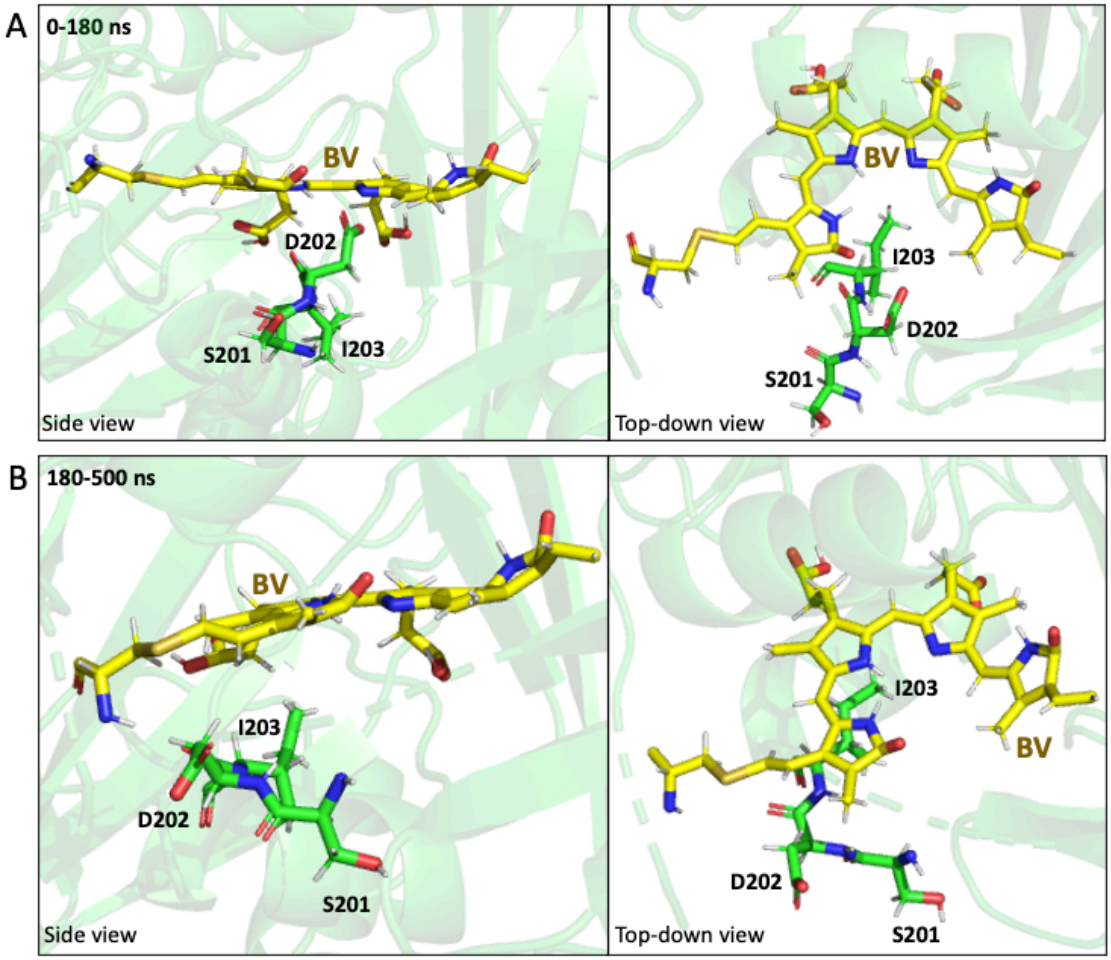


Figure 5.9: mRhubarb720-WT protein scaffold reconfiguration. Simulation 1 schematic of (A) 0-180 ns and (B) 180-500 ns, displaying the molecular conformation of S201, D202 and I203 (green) in respect to BV. The left panels are the side view and right represents the top-down view of the BV-binding environment.

During the initial phase of the simulation (0-180 ns), the close proximity of these flexible residues to BV may influence the local environment around the chromophore. A slight increase in the RMSD for Simulation 1 is observed around 180 ns, suggesting some degree of flexibility is introduced in the surrounding protein structure (Figure 5.6.A). For

Simulations 2 and 3, this movement occurs within the first 30 ns and all three simulations do not revert back to the starting configuration, which suggests that the system has transitioned into a new local energy minimum. This behaviour indicates that the initial structure was likely in a higher-energy state, and the observed conformational change reflects the system settling into a more energetically favourable configuration. Even with this movement, the overall protein scaffold continues to provide sufficient support to BV, preventing any major conformational changes. However, after 180 ns the residues move away from BV, but despite this shift, no significant changes are observed in the RMSD of BV across the 500 ns of the simulation (Figure 5.7). This indicates that BV remains structurally stable throughout and suggests that BV stability is maintained primarily by the broader protein framework rather than specific interactions with these residues. Other than this region, the BV-binding site appears relatively rigid across all simulations.

It is also important to analyse the RMSF of BV to assess the flexibility of individual atoms within the cofactor (Figure 5.10). It has previously been reported that the D-ring of BV is particularly susceptible to photoisomerisation upon irradiation with far-red/near-IR photons, converting from a Pr state (red-light absorbing) to a Pfr state (far-red-light absorbing). This photoisomerisation event plays an evolutionarily role in the physiological responses in plants/photosynthetic bacteria (Yang et al. 2009d). Therefore, monitoring atomic fluctuations within the D-ring is crucial for determining if isomerisation occurs in the absence of irradiation, as this would be detrimental to the proteins function.

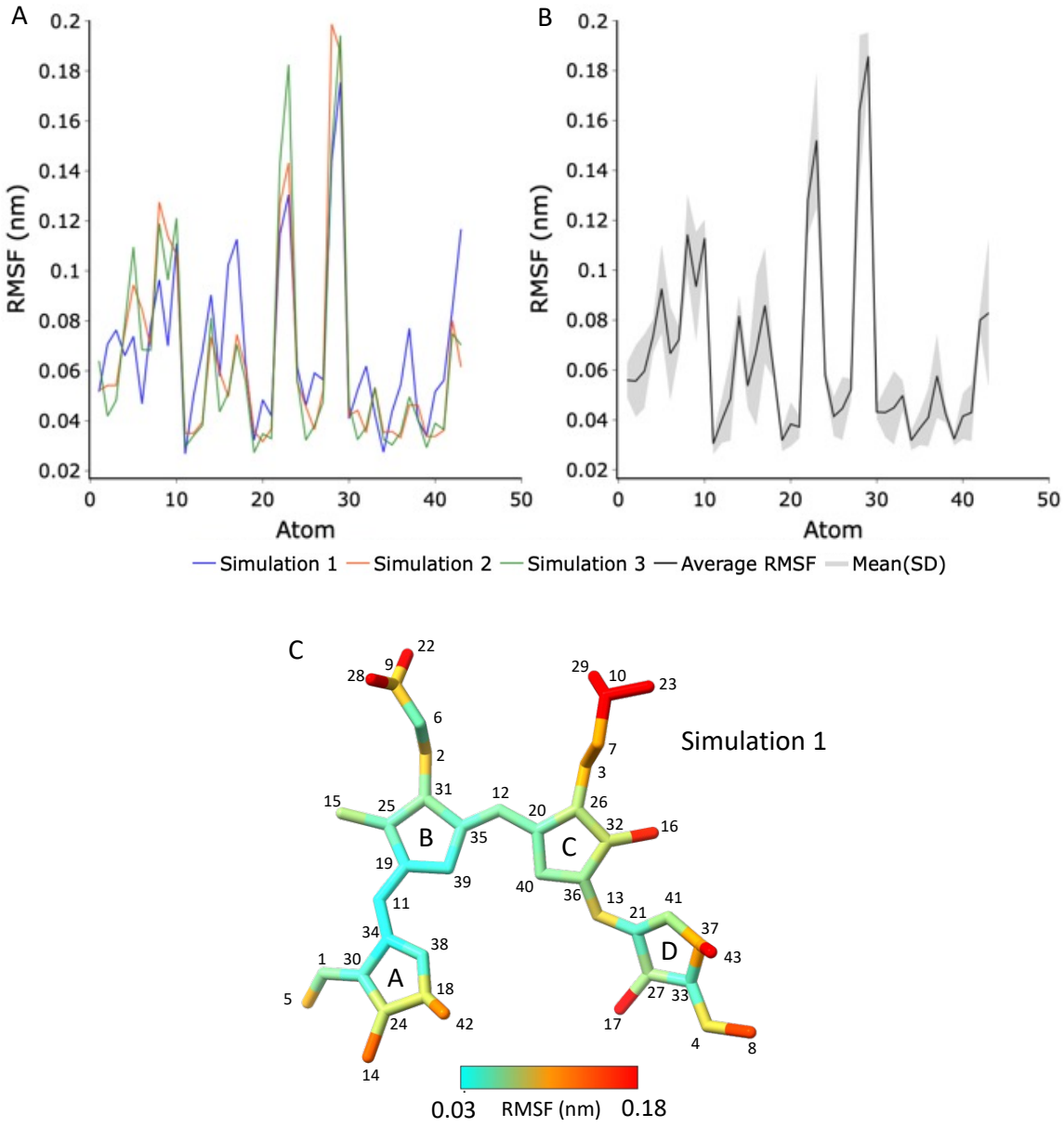


Figure 5.10: Biliverdin XI-alpha all atom (without H) RMSF analysis over three simulations. (A) RMSF of individual simulations: Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green). (B) The average RMSF of all simulations (black) with the standard deviation (light grey). (C) RMSF mapping of Simulation 1 coloured by atomic fluctuation from minimum (cyan) to maximum (red) illustrated by the colour scale. Each atom is labelled with their corresponding atom number for the graph, as well as the nomenclature for each ring.

The average RMSF values in Figure 5.10.B indicate that the greatest fluctuations are observed in both carboxyl branches extending from rings B and C (atoms 9-10, 22-23, 28-29)

with RMSF values between 0.12-0.18 nm. Atom 8 is symmetrically equivalent to atom 5, which forms the thioester bond with Cys15 in mRhubarb720. As atom 8 does not participate in covalent bonding with the protein scaffold, it can exhibit greater flexibility. The structural representation of BV, shown in Figure 5.10.C, provides a spatial mapping of the flexibility, where it highlights the highest atomic fluctuations occur at the carboxyl branches.

Throughout the simulations, several residues intermittently form hydrogen bonds with BV, primarily involving the carboxyl branches. R249 and S252 interact with the branch extending from the B-ring, while R217 and S269 binds to the branch from the C-ring. One limitation of the PyMOL hydrogen bond finder function is that a hydrogen atom cannot form simultaneous hydrogen bonds with different acceptors. Despite this, the hydrogen bonds observed tend to break and reform consistently over the simulation, leading to a dynamic interaction where bonding is equally distributed among the acceptor atoms.

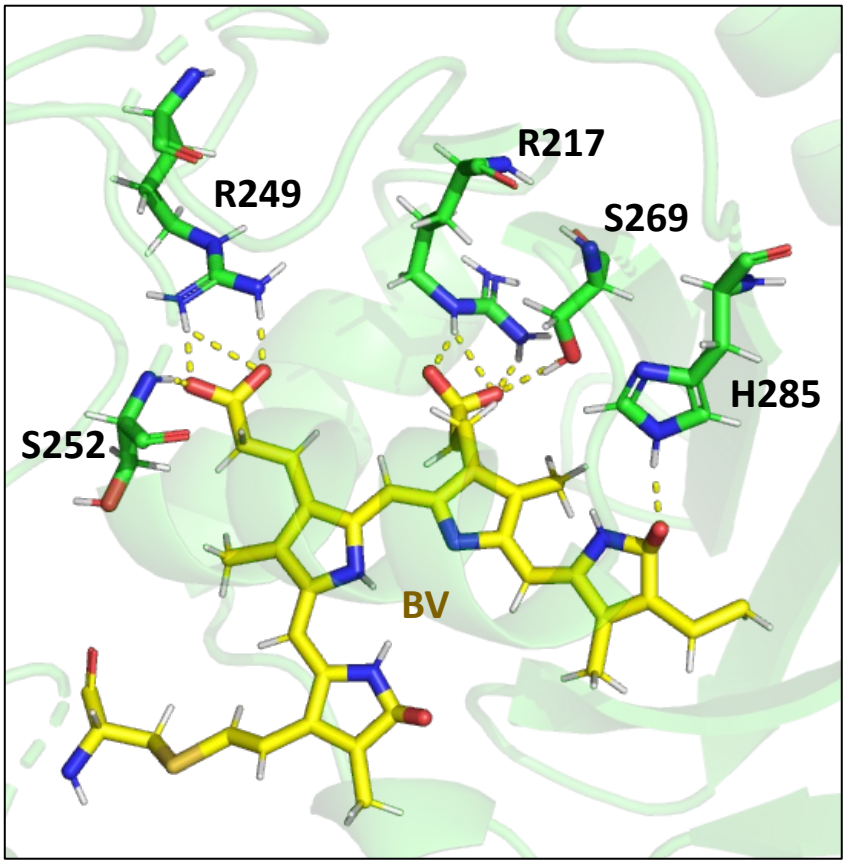


Figure 5.11: mRhubarb720-WT protein-BV hydrogen bond interactions. Taken from Simulation 1, showing the hydrogen bonds (yellow dashed lines) binding to the residues labelled.

The hydrogen bond formed by H285 specifically stabilises the oxygen attached to the D-ring, which is hypothesised to help maintain BV in its Pr-state conformation. A notable feature of the ring structure is the flexibility of the D-ring compared to the other rings from the RMSF analysis. This increased flexibility may suggest a propensity for movement, which could influence the cofactors photoconversion state. However, it is important to note that despite this flexibility, no isomerisation of the C15=C16 double bond is observed throughout the 500 ns simulation, indicating that while the D-ring displays some mobility, it does not lead to significant structural changes in the chromophore during this time frame. It is important to note that the observations above focus solely on the protein-BV interactions, although solvents do interact with both the protein and BV. However, the visualisation of hydrogen bonds involving solvent molecules was not possible due to the large file size, which caused the visualisation software to become unresponsive.

Building on the protein-BV interactions, the following analysis will examine the total number of hydrogen bond interactions observed between BV and both the protein and surrounding solvent molecules across the simulations (Figure 5.12). This analysis will provide a more comprehensive view of how these solvent interactions contribute to the stability and dynamics of the chromophore within its environment.

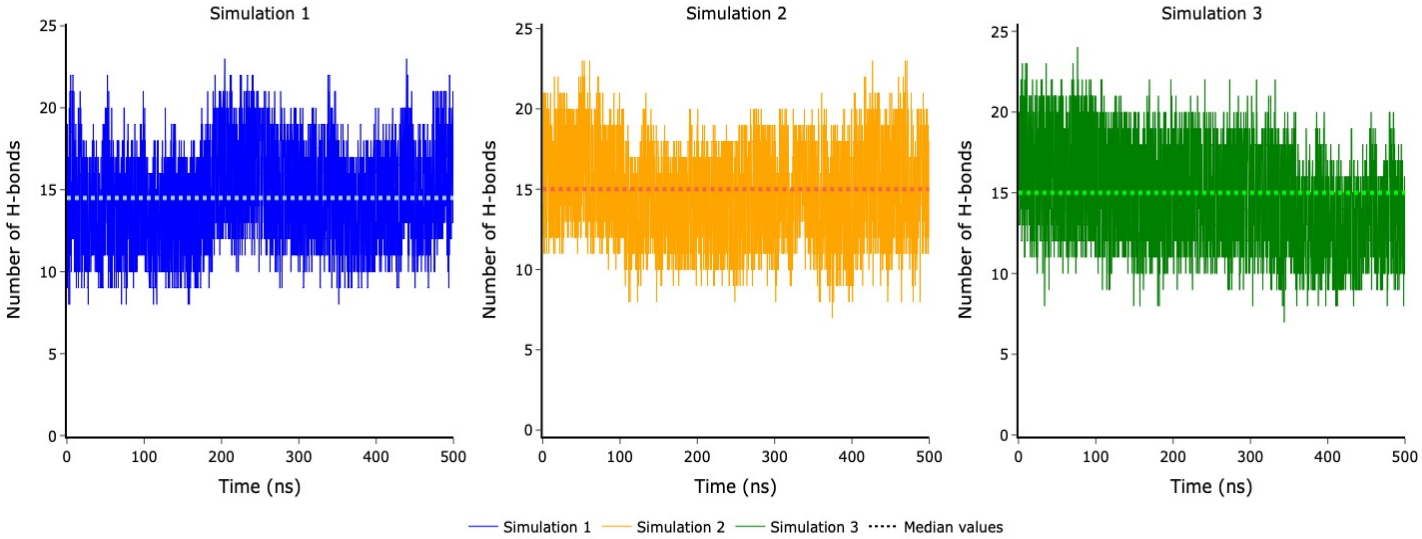


Figure 5.12: Total number of hydrogen bonds between BV and the surrounding environment. These calculations considered the H-bonds between BV and the protein scaffold and solvent present in the system. Each simulation was plotted and was presented with each corresponding median values.

For Simulations 2 and 3, the median number of hydrogen bonds is exactly 15, while for Simulation 1, the median is 14.5. Since a fractional hydrogen bond is not physically meaningful, this value can reasonably be rounded to 15 for consistency. This consistency across all three simulations indicates that BVs interaction with the protein scaffold and solvent environment is stable. Figure 5.12 provides insight into how these hydrogen bonds are distributed between the protein scaffold and the solvent. With 5 oxygen atoms available to bond with the protein scaffold, up to 10 hydrogen bonds could theoretically form, however, the H285 – D-ring interaction consistently forms only one hydrogen bond with its oxygen. This suggests that, on average, approximately 9 of these bonds are attributable to the scaffold, with the remaining hydrogen bonds contributed by interactions with the solvent. Unlike PyMOL, which incorrectly identifies interactions such as oxygen-oxygen bonds or simultaneous hydrogen bonds from a single hydrogen atom, GROMACS applies stringent geometric criteria that ensures each hydrogen bond is accurately assigned. In GROMACS, each hydrogen can form only one bond at a time, and only appropriate donor-acceptor pairs are recognised as true hydrogen bonds. This analysis will provide a baseline for the future comparisons, in particularly when investigating how mutations changes may affect the stability of BV through hydrogen bond interactions.

Following the analysis of hydrogen bond interactions, the next step in understanding the stability and dynamics of the BV within the protein environment involves examining the SASA. By analysing the SASA of both the protein and BV, insights can be gained into how exposed the chromophore is to the surrounding solvent, which may have important implications for its interaction with the protein scaffold (Figure 5.13). For the protein, the SASA remains relatively stable across the three simulations, with values consistent around 160 nm^2. The standard deviation is small, indicating that the protein maintains a steady exposure to the solvent throughout the simulation. This stability suggests that the proteins overall conformation is preserved and that its interactions with the surrounding solvent remain constant over time. The minor fluctuations observed are expected in long-term molecular dynamics simulations and do not appear to affect the overall protein structure or its interactions with BV.

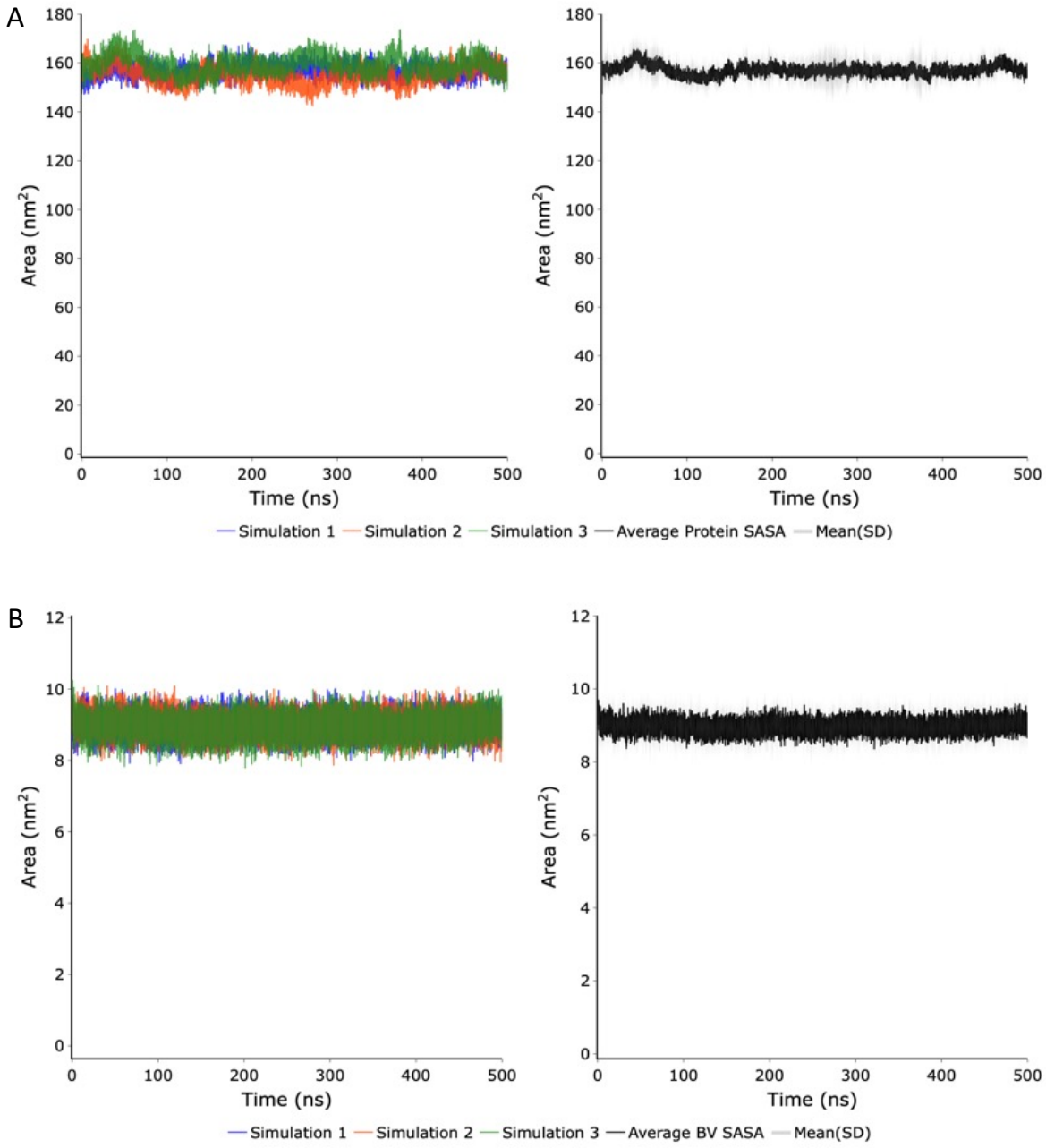


Figure 5.13: Surface Accessible Surface Area (SASA) for mRhubarb720 protein scaffold and BV. (A) The SASA of the entire protein (excluding BV) and (B) the SASA of BV specifically. Each simulation: Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green), with the average SASA (black). In both panels, the left-hand plots show the individual SASA values for each simulation, while the right-hand plots show the mean average SASA across all three simulations, with the standard deviation included (grey)

The SASA for BV shows consistent values across the three simulations, with the SASA centred around 9 nm^2. The lack of significant fluctuation in the chromophores SASA suggests

that the BV remains well-embedded within the protein scaffold, with limited exposure to the solvent. This stable exposure profile further reinforces the idea that BVs interactions with the surrounding protein and solvent environment are maintained throughout the course of the simulation, with no significant rearrangements or conformational shifts that would lead to increased solvent exposure. In Chapter 3.2.4, the crystallisation of mRhubarb720 showed that the BV is mostly enclosed within the GAF domain, with a small gap near the thioester bond where BV binds (Figure 3.5.A). This suggests that the chromophore is well-protected by the surrounding protein structure, minimising its exposure to the solvent. Consequently, any mutation that increases the solvent accessibility of BV could potentially destabilise the binding environment. Given the well-maintained nature of the binding pocket, increased solvent exposure would likely disrupt the balance of interactions between BV and the protein scaffold, possibly leading to a reduction in chromophore stability or altered optical properties.

The final MD analysis conducted on both the protein scaffold and BV was the Rg, highlighting the distribution of atoms in respect to the centre of mass of the molecule. The mRhubarb720 protein and on the BV individually, was conducted and displayed in Figure 5.14. Observations show minimal deviation from the average across the 500 ns, this signifies a relatively stable structure throughout the simulation. The average protein Rg remains constant at around 2.05 nm, and the low SD suggests that there is little variability in the proteins overall size and compactness across all three simulations, reinforcing the idea that the protein does not undergo significant conformational changes or unfolding. This result also implies that the proteins tertiary structure does not exhibit large-scale expansions or contractions.

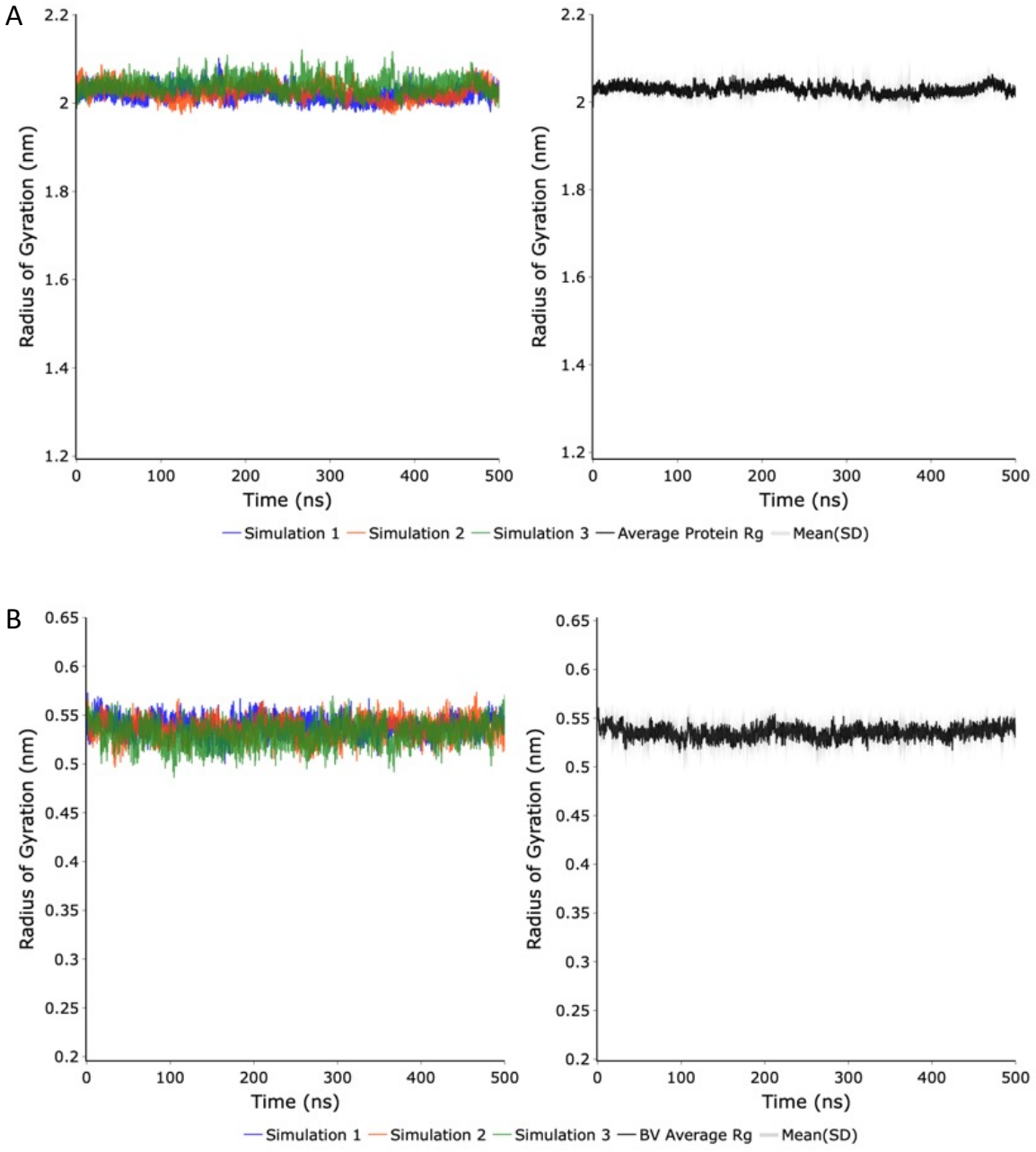


Figure 5.14: Rg analysis of mRhubarb720 WT and BV over three simulations. (A) The Rg of the entire protein (excluding BV) and (B) the Rg of BV specifically. Each simulation: Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green), with the average Rg (black). In both panels, the left-hand plots show the individual Rg values for each simulation, while the right-hand plots show the mean average Rg across all three simulations, with the standard deviation included (grey)

The cofactors Rg fluctuates around an average value of 0.54 nm, with minor variations detected over 500 ns simulations. The consistent Rg for BV reflects the fact that it remains tightly embedded within the protein, with minimal movement or conformational

flexibility. The lack of significant deviation in the Rg suggests that BV is stabilised by the surrounding protein scaffold, with no indications of significant fluctuations that might affect its optical properties. Each MD analysis discussed in this section will provide a baseline for the apo-protein, free-BV and the mutational simulations explored in the following sections.

5.2.3.2 Apo-protein

Investigating the conformational dynamics of the apo-protein is important for understanding the functional behaviour of mRhubarb720, especially since all functionally expressed mutants contain a proportion of the apo form within the sample. The apo-protein serves as a key reference point for assessing how various mutations influence the proteins structural integrity and interactions with the solvent. The presence of the apo-protein in mass spectrometry results indicates its stability, where it can exist in a viable state during experimental conditions. Understanding the conformational landscape of the apo-protein allows for a clearer interpretation of how mutations may affect the proteins dynamics and interactions, including any changes in binding efficiency or solvent accessibility.

The RMSD was conducted first to observe any conformational changes throughout each simulation (Figure 5.15). Notably, the apo-protein exhibits significant fluctuations in RMSD across all three simulations, indicating the presence of different conformational states. These variations suggest that the apo-protein is capable of adopting multiple alternative conformations, which may be influenced by its surrounding environment and the inherent flexibility of the protein structure.

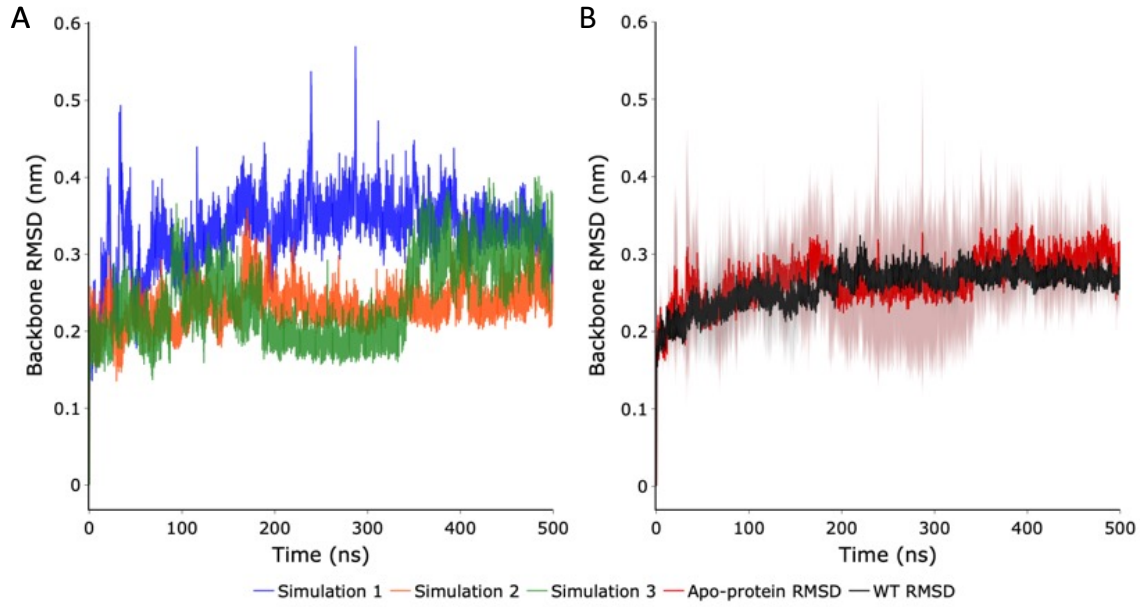


Figure 5.15: mRhubarb720 Apo-protein RMSD analysis over three simulations. (A) Each simulations RMSD: Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green). (B) The average apo-protein RMSD (red) was compared to WT (black), with the standard deviation highlighted in their respective colours (apo – red; WT – black).

The averaging of these fluctuating states results in an overall RMSD that aligns closely with the average WT RMSD values, however, this alignment is misleading. While the average RMSD of the apo-protein mimics that of the WT, it does not accurately reflect the dynamic behaviour of the apo-protein. The WT protein maintains a consistent conformation throughout the simulation, with minimal fluctuations, indicating a stable and rigid structure. In contrast, the apo-proteins ability to shift between different conformations highlights its dynamic nature, suggesting that it may have a more adaptable structure in the absence of the bound chromophore. To analyse this, snapshots at 250 ns and 500 ns for each simulation were compared to the WT simulation in order to observe any conformational changes (Figure 5.16). The WT structure shows minimal change between the two time points, with the outline of the bound BV visible, demonstrating its consistent conformation throughout the simulations.

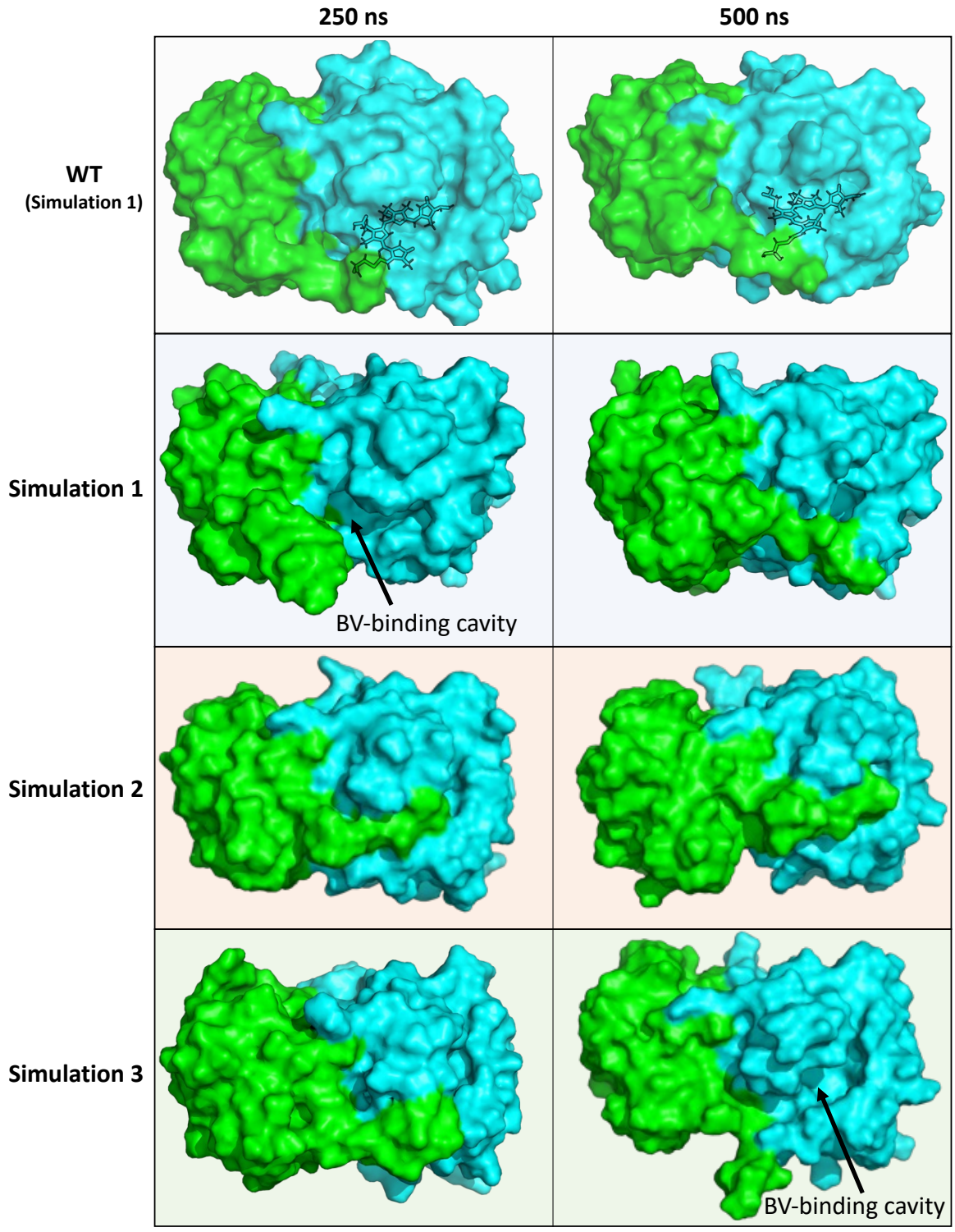


Figure 5.16: Comparing molecular configurations of apo-protein simulations to WT at 250 ns and 500 ns. Each image shows the surface of the protein with the PAS domain (green) and GAF domain (cyan) to distinctly highlight where the binding site resides. The WT depictions are set at 20% transparency in order to observe the Cys-BV outlined in black. The binding cavity in the GAF domain is annotated if exposed.

In Simulation 1 at 250 ns, the apo-protein reveals an exposed binding cavity, while at 500 ns, the N-terminal domain, where the PAS domain (green) arm responsible for binding BV, partially obstructs the binding site. In Simulation 2, the N-terminal domain remains securely positioned within the binding site at both time points, suggesting a stable interaction that maintains the integrity of the cavity. Meanwhile, Simulation 3 transitions from having the N-terminal domain fixed within the binding site at 250 ns to exposing the binding site by 500 ns. This shift explains the sudden jump in RMSD observed in this simulation at 350 ns, indicating a significant conformational change.

These findings highlight the importance of ensuring effective BV incorporation when mutating residues within the BV-binding site. While the apo-protein is dynamic, it remains relatively stable in certain conformations, which is why it is detected in the experimental samples. If initial binding is partially inhibited due to mutations, the formation of apo-protein alongside holo-protein may occur. Therefore, the placement of mutational targets is crucial, otherwise, samples could become heterogeneous which can cause difficulties with the interpretation of the results.

The implications of these observations extend to the RMSF, where the N-terminal domain (i) is the most flexible region of each simulation (Figure 5.17.A). The second peak (ii) is located on a loop connecting the PAS and GAF subunits, where this loops flexibility increases in the absence of BV. For the third peak (iii), it is close to the flexible region in WT, previously mentioned. Whereas WT's flexible residues are S201-I203, the apo-proteins higher flexibility is slightly shifted to H197-P199 which can be attributed to the need for these residues to adjust in order to accommodate the N-terminal BV-binding site in the absence of BV.

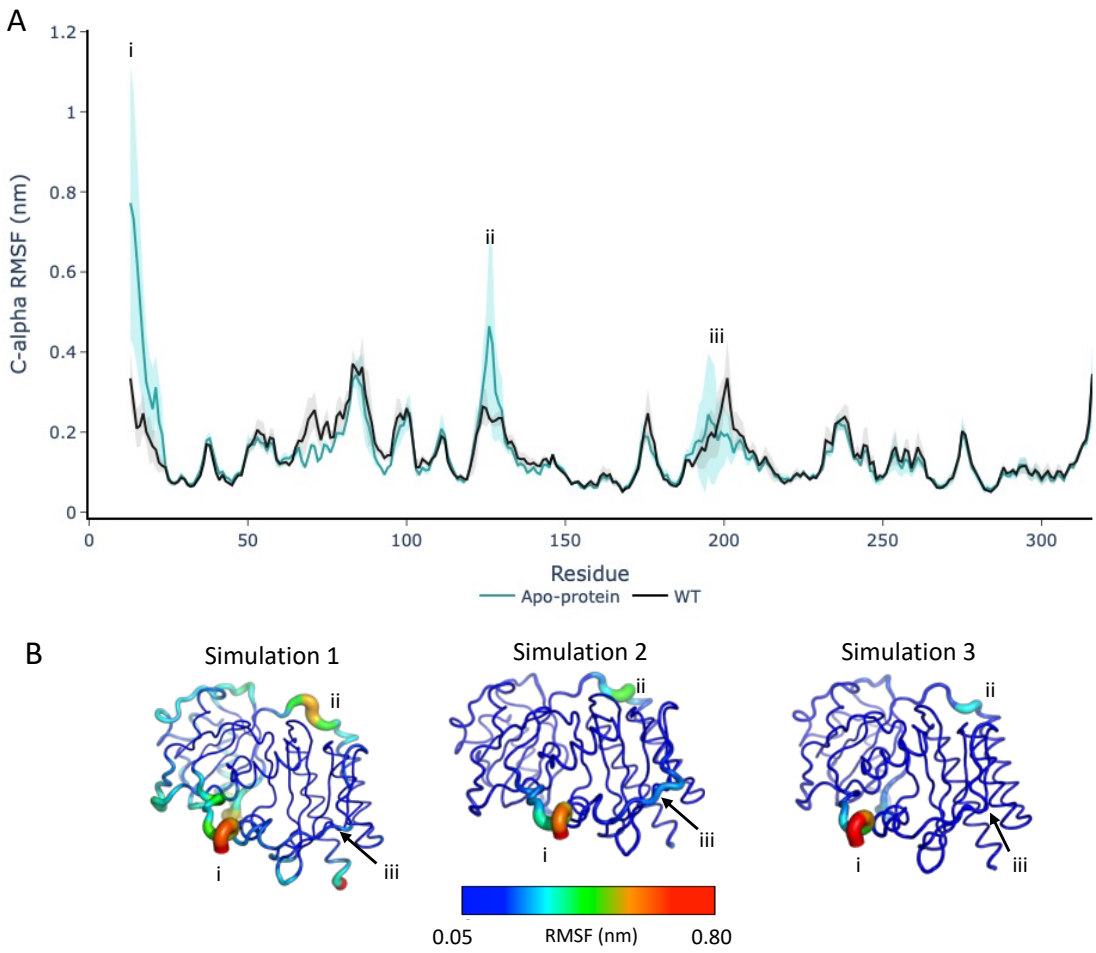


Figure 5.17: mRhubarb720 apo-protein C-α RMSF analysis over three simulations. (A) The average apo-protein RMSF over three simulations (cyan), plotted against the average RMSF of WT (black) with the standard deviation depicted in their respective colours. (B) B-factor putty representation of each apo-protein simulation coloured by residue fluctuation from minimum (blue) to maximum (red) illustrated by the colour scale.

The B-factor for all three simulations show that the N-terminal domain remains a focal point of flexibility (Figure 5.17.B). However, the differences observed among the simulations illustrate the dynamic nature of the apo-form. For instance, Simulation 2 exhibits a slight reduction in movement indicating a more stable conformation, which is also supported by the RMSD analysis. Whereas Simulation 1 and 3 demonstrates increased movement in the same region. While the WT protein showed more flexibility in its B-factor figures, this observation is somewhat misrepresented by the shorter RMSF scale. The increased B-factor values for the WT may suggest greater mobility, but the structural

integrity and overall conformation of the WT remain stable throughout the simulation. In contrast, the apo-protein demonstrates a larger RMSF range, which will skew the flexibility of majority of the protein where it is similar to WT.

The SASA values for the apo-protein exhibit a slight but consistent increase compared to the WT, which can be attributed to the exposed BV-binding site observed throughout the simulations (Figure 5.18.A). This suggests a greater interaction with the solvent, which aligns with the conformational changes noted in previous analyses.

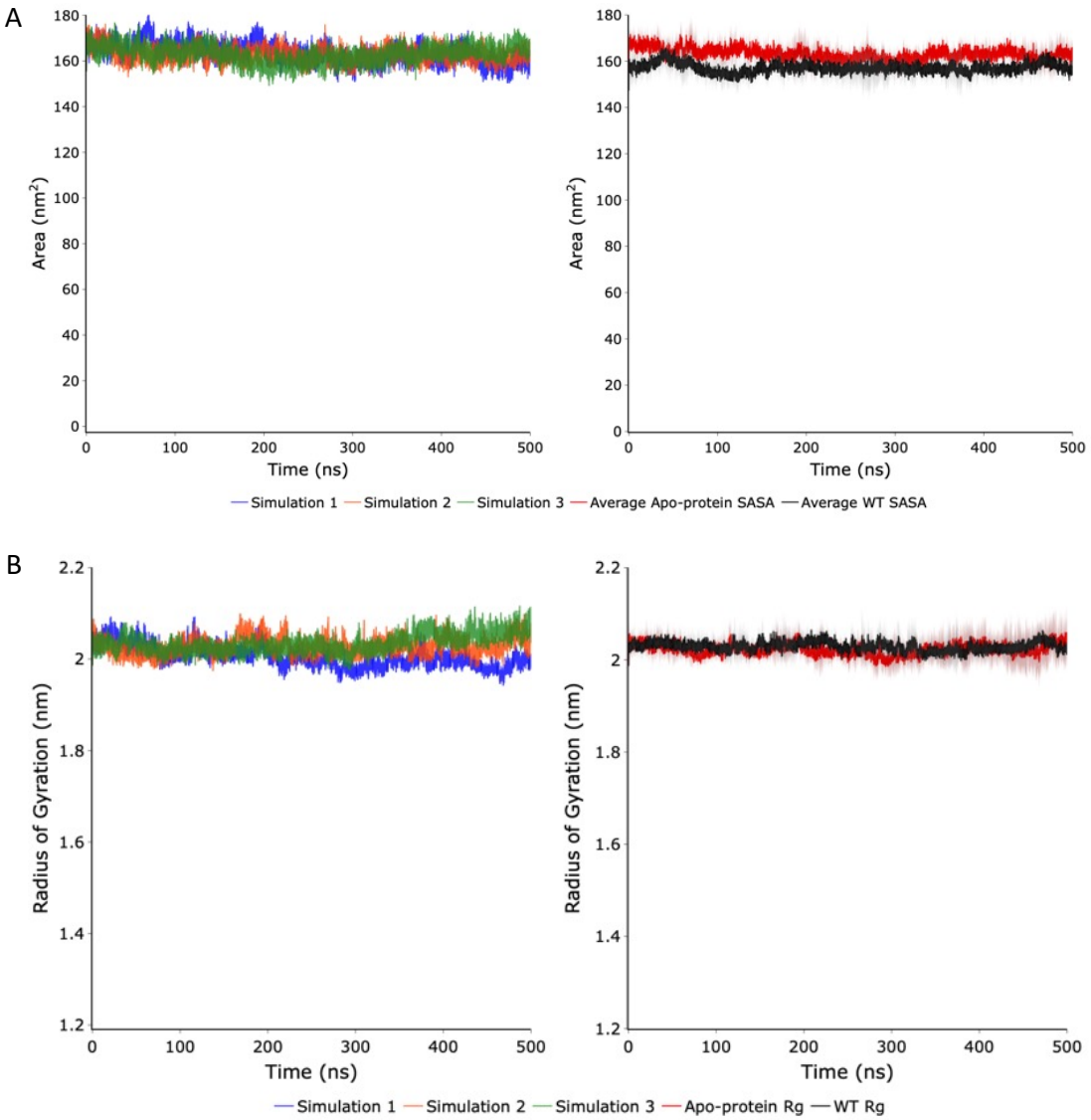


Figure 5.18: SASA and Rg analysis of apo-form compared to WT. (A) SASA analysis and (B) Rg analysis with each simulation: Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green) in the left-hand plots show the individual Rg values for each simulation, while the right-hand plots show the mean average apo-form compared to the WT across all three simulations, with the standard deviation included (Apo – red; WT – grey).

The Rg values for the apo-protein do not significantly change compared to those of the WT (Figure 5.18.B). This stability can be attributed to the behaviour of the N-terminal domain, which may effectively cancel out the variations in Rg when it transitions between positions away from the BV-binding site and those within it. A closer examination reveals distinct trends in the individual simulations. Specifically, Simulation 1 shows a gradual reduction in Rg over the 500 ns period, suggesting a compacting of the protein structure as it stabilises. In contrast, Simulation 3 exhibits an increase in Rg during the same timeframe, indicating a more expanded conformation. These changes align well with the snapshots observed in the previous figure, which depict the differing states of the apo-protein at 250 ns and 500 ns.

5.2.3.3 Functionally-expressed pCNPhe mutants

The mutants developed and experimentally analysed (D177, Y198, N224, F258 and Y198/F258) earlier in this chapter were also computationally investigated to understand the relationship between pCNPhe incorporation and potential impact on the holo-protein and the cofactor. Firstly, RMSD analysis was conducted and plotted to show the average deviation from 50-500 ns, excluding the equilibration of the simulation (0-50 ns) to avoid potential data skewness (Figure 5.19). The average RMSD for the WT across all three repeats was 0.264 ± 0.034 nm, while the apo-protein had an average RMSD of 0.252 ± 0.060 nm. The five functionally expressed mutants exhibited the following average RMSD values: D177 (0.240 ± 0.069 nm), Y198 (0.262 ± 0.035 nm), N224 (0.232 ± 0.060 nm), F258 (0.252 ± 0.046 nm), and the double mutant Y198/F258 (0.219 ± 0.044 nm). Statistical analysis was performed using the Kruskal-Wallis test, a non-parametric method for comparing medians across groups without assuming normal distribution. This was achieved by applying the test to each individual simulation against all three WT simulations and if there was a non-significant p value ($p < 0.0001$) associated with the test, then that particular simulation will be treated as such. While useful, its application to large MD datasets is limited due to the high variability in protein flexibility and increased number of data points, which may complicate the interpretation of the results.

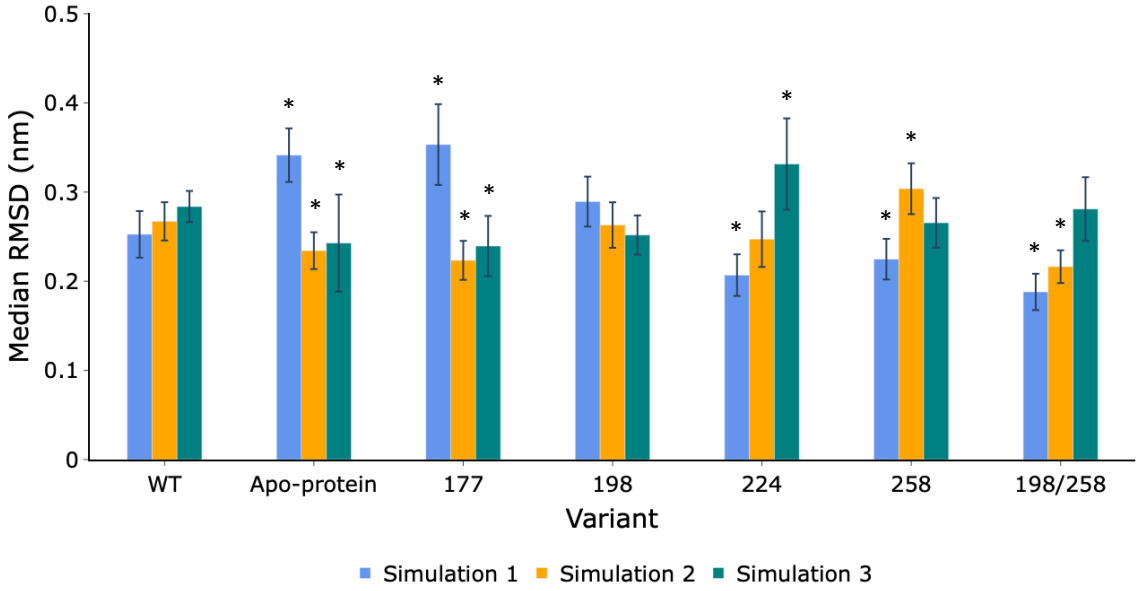


Figure 5.19: Average RMSD of WT, apo-protein and pCNPhe mutants calculated over 50-500 ns time frame. All three simulations are represented by their median values and the error bars represent the standard deviation of the averaged RMSD values for each simulation. Kruskal-Wallis test was performed on each simulations to compare against each WT simulation, with * denoting significance ($p < 0.0001$). All individual mutant RMSD graphs can be seen in Figures S3-7.

Notably, the apo-protein and variant N177 (which served as a control) consistently exhibited significant deviations from WT across all simulations. The apo-proteins increased RMSD, as expected, reflects the conformational flexibility caused by the absence of the BV cofactor, confirming the crucial role of BV in maintaining a set protein configuration. In contrast, the significant deviation observed in variant N177 is interesting, given its designation as a control. This finding suggests that even minimal structural modifications can lead to noticeable perturbations in protein dynamics. Variant Y198 stands out as the only variant that sustained non-significant deviations relative to WT in all three simulations. This indicates that the mutation at this position does not significantly affect the protein's overall conformation, potentially maintaining native-like stability. On the other hand, variant Y198/F258, despite being a double mutant, showed significantly lower RMSD values, implying that these mutations may lead to a more rigid conformation.

The BV RMSD was calculated to measure the cofactors deviation from the mean reference structure (Figure 5.20). RMSD values for each variant were calculated as follows: WT displayed a median RMSD of 0.148 ± 0.023 nm, with variants D177 and Y198 showing

similar stability at 0.146 ± 0.022 nm and 0.140 ± 0.018 nm, respectively, suggesting these mutations do not significantly impact BVs structural stability. Variant Y198, in particular, had the lowest RMSD among all single mutation variants, possibly indicating a slight stabilising effect on BV. Interestingly, the double mutant Y198/F258 showed the lowest median RMSD (0.139 ± 0.019 nm) overall, suggesting a possible stabilising interaction when both mutations are present, potentially restricting BV movement or increasing its rigidity within the protein environment.

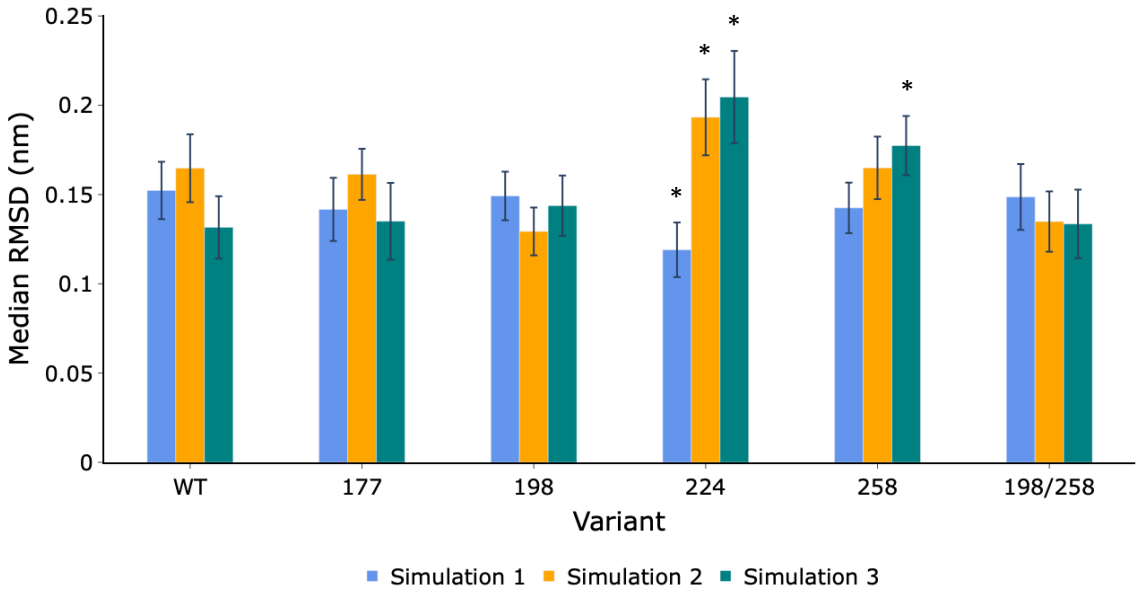


Figure 5.20: Average BV RMSD of WT and pCNPhe mutants calculated over 50-500 ns time frame. All three simulations are represented by their median values and the error bars represent the standard deviation of the averaged RMSD values for each simulation. Kruskal-Wallis test was performed on each simulations to compare against each WT simulation, with * denoting significance ($p < 0.0001$). All individual RMSD graphs can be seen in Figures S8-12.

Variant F258 displayed an elevated RMSD (0.156 ± 0.024 nm), but statistical significance was only observed in the third simulation, indicating that the increased deviation may not be consistent across all conditions. In contrast, variant N224 exhibited a notably higher RMSD at 0.166 ± 0.042 nm, showing complete statistical significance across all simulations ($p < 0.0001$). The unique behaviour of N224 can be attributed to the mutations position at the interface between the PAS and GAF domains (Figure 5.21). Originally intended as a control, this mutation was expected to produce minimal structural

impact, however, its location appears to influence the internal conformation of BV, particularly in simulations 2 and 3, where the BV-binding site displays a noticeably altered structure. In hindsight, this placement likely made N224 a less ideal control due to its proximity to a critical domain interface.

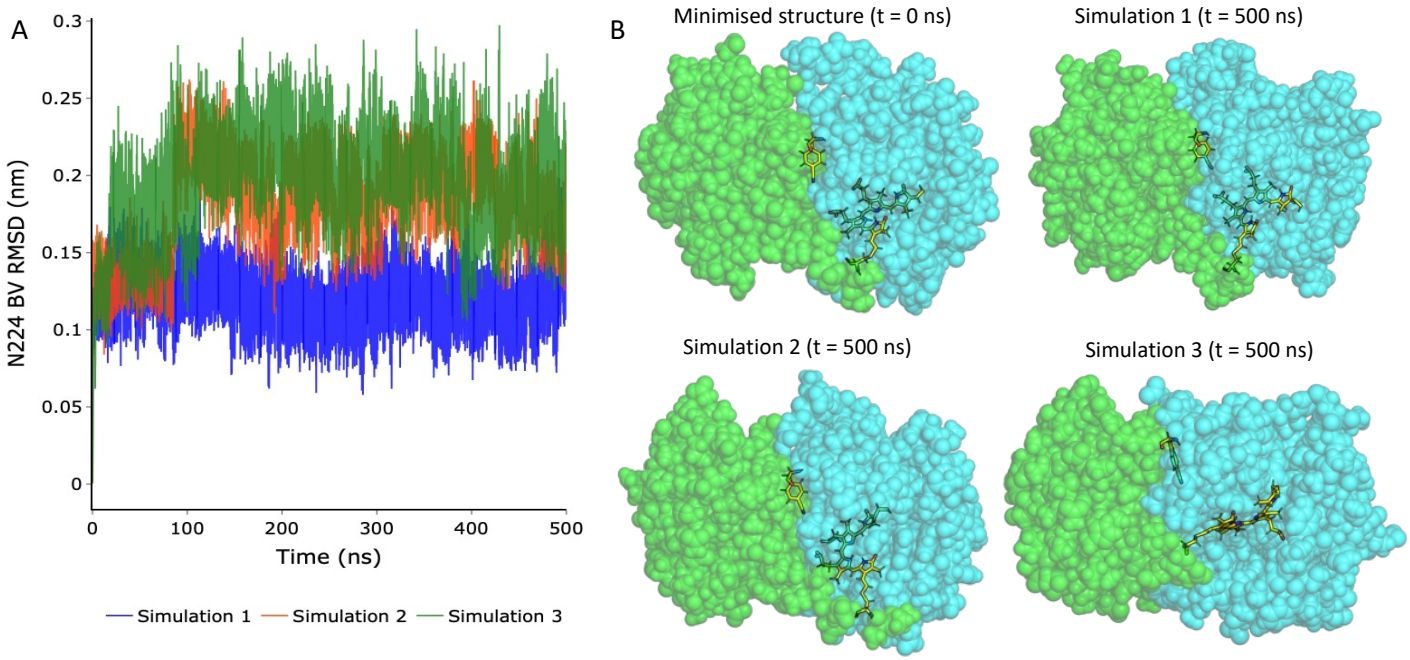


Figure 5.20: N224 BV RMSD focus with molecular configuration of each simulation after 500 ns. (A) BV RMSD plot of N224 simulations, with Simulation 1 (blue), Simulation 2 (orange) and Simulation 3 (green). (B) Comparison of the minimised structure for N224 to each simulation at 500 ns, both the Cys-BV and pCNPhe were highlighted in yellow sticks, with the PAS (green) and GAF (blue) denoted via spheres.

The PAS and GAF domains often function in structural support and signal transduction, and a mutation at their intersection will likely disrupt their interactions, inducing a shift in the proteins overall conformation. This shift appears to propagate to the BV-binding pocket via allosteric coupling, leading to an environment less optimised for BV binding. In simulations 2 and 3, this structural rearrangement likely weakens the stabilising interactions within the binding site, such as non-covalent interactions, reducing the stability of BV and contributing to the increased RMSD observed for N224.

Analysing the RMSF of the mutants, key regions show differences in flexibility compared to the WT (Figure 5.21). The N-terminal domain, where BV binds to Cys15, shows consistently higher RMSF values in all mutants compared to WT, suggesting that mutations

may increase flexibility around this critical binding site. This heightened flexibility could indicate an indirect effect of the mutations on the BV-binding regions stability, as any increased movement here may influence the overall dynamics of BV. Residue 126, although distant from the BV-binding site, also exhibits elevated flexibility in most mutants. This observation implies that the mutations may have subtle, long-range effects on protein dynamics, altering flexibility in regions that are not directly involved in BV binding.

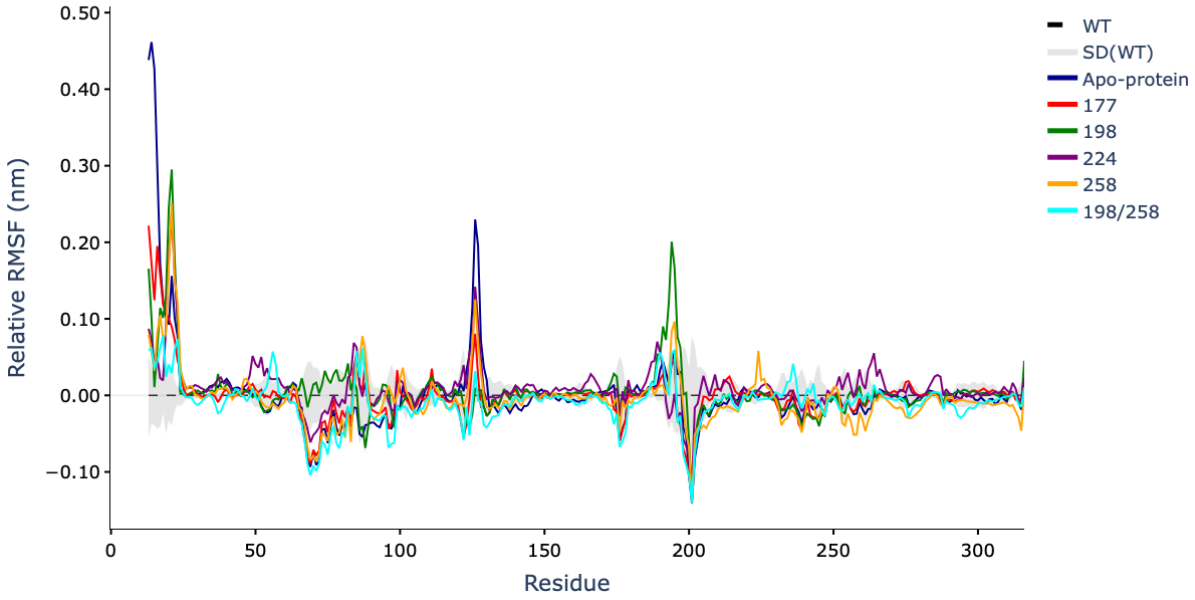


Figure 5.21: Relative C-α RMSF of the pCNPhe variants with respect to WT. All variants were subtracted by the WT values for each residue, where positive relative change depicts more flexible residues and vice versa. WT is represented by the black dashed line (y = 0) with the standard deviation shown by light grey. All individual RMSF graphs can be seen in Figures S13-17.

The largest shift in flexibility resides within the BV-binding pocket, where residues 194-195 show increased RMSF in the mutants, while residues 201-203 display decreased flexibility relative to WT. This shift suggests that the mutations may redistribute flexibility within the binding pocket, enhancing flexibility at 194-195 while stabilising the 201-203 region. This could imply a compensatory rearrangement, where the protein balances increased flexibility in one area by stabilising another, possibly to maintain structural integrity within the pocket. Additionally, it is worth noting the broad similarity in RMSF patterns among all mutants, which indicates that although each mutation introduces specific

changes, there is an underlying consistency in how these modifications impact protein flexibility, particularly within the BV-binding region.

In the BV RMSF analysis, the control mutants D177 and N224 exhibit lower flexibility on average than WT, contrary to expectations for the controls (Figure 5.22). This discrepancy suggests that these choices may have been suboptimal for mutagenesis, as they induce an unexpected reduction in BV flexibility rather than a providing a neutral impact. It is important to note that this MD analysis was conducted following experimental work to further explore the experimental observations. This unexpected reduction in flexibility highlights the importance of selecting appropriate controls for mutagenesis, as these findings may suggest that these mutations inadvertently reduce BV flexibility within the protein, affecting its natural dynamics.

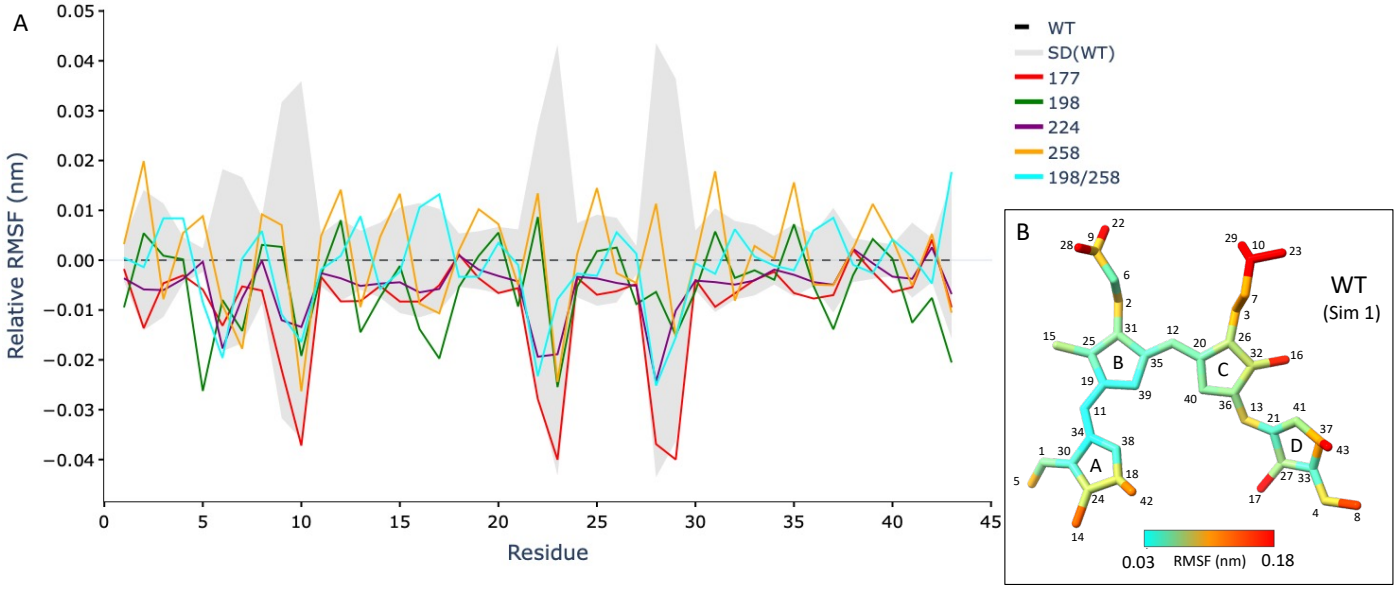


Figure 5.22: Relative all atom (without H) BV RMSF of the pCNPhe variants with respect to WT. (A) All variants were subtracted by the WT values for each residue, where positive relative change depicts more flexible residues and vice versa. WT is represented by the black dashed line (y = 0) with the standard deviation shown by light grey. (B) Structural map of BV with WT Simulation 1's RMSF values visualised. All individual BV RMSF graphs can be seen in Figures S18-22.

In contrast, mutants 198, 258, and the double mutant 198/258 show flexibility profiles similar to WT, indicating these mutations have a minimal effect on BV conformational dynamics. This similarity implies that the structural flexibility inherent to WT BV is preserved in these variants, unlike in the control mutants. The presence of apo-protein observed in these samples experimentally is therefore more likely attributable to binding

mechanisms, such as protein-folding pathways or cofactor incorporation efficiency, rather than any intrinsic instability of BV itself. The largest decrease in flexibility across all variants is observed at the carboxyl groups (atoms 9-10, 22-23 and 28-29) protruding away from the B and C-rings. Although not substantial, this does point to the fact that the cofactor environment is more restrictive when incorporating pCNPhe.

The hydrogen bond analysis between BV and its surrounding environment reveals at least one statistically significant simulation in all variants (Figure 5.23). Despite the observed significance, this does not necessarily imply that there are meaningful or functionally relevant differences in hydrogen bonding patterns across the variants. As mentioned previously, the complexity and variability inherent to MD data mean that significance alone does not necessarily confirm meaningful differences, but rather highlights potential shifts that require further consideration.

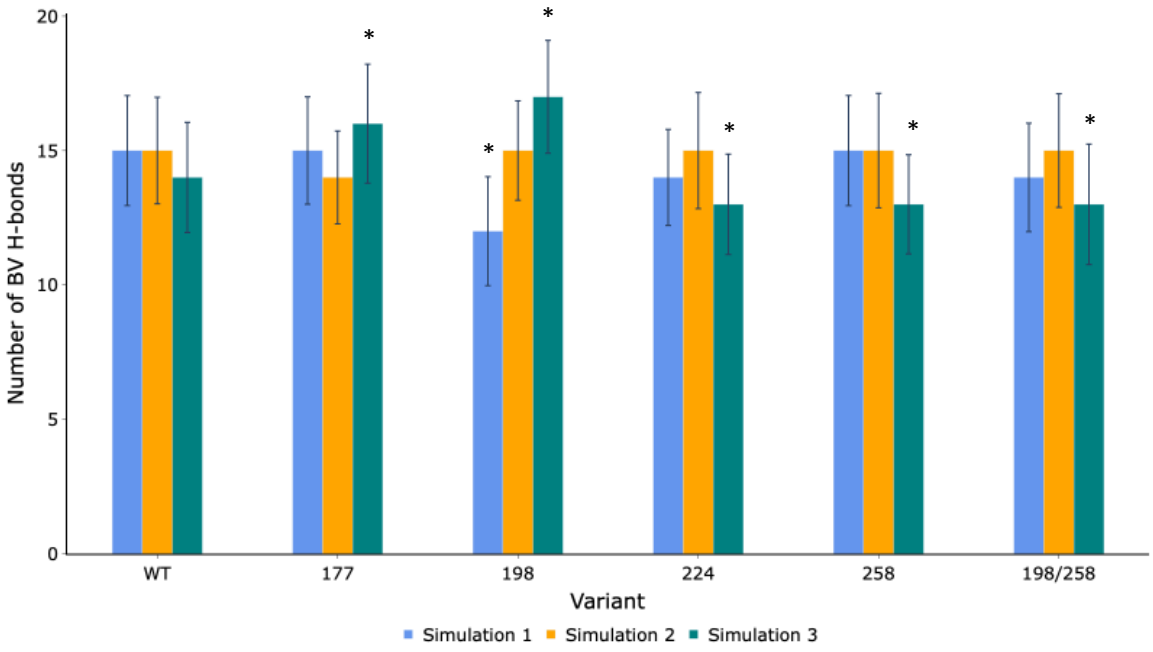


Figure 5.23: Average number of hydrogen bonds between BV and the surrounding environment, calculated over 50-500 ns time frame. All three simulations are represented by their median values and the error bars represent the standard deviation for each simulation. Kruskal-Wallis test was performed on each simulations to compare against each WT simulation, with * denoting significance ($p < 0.0001$). All individual hydrogen bond count analysis graphs can be seen in Figures S23-27.

The average number of BV hydrogen bonds for each variant are as follows: WT at 15.0 (± 2.11), D177 at 15.0 (± 2.22), Y198 at 15.0 (± 2.75), N224 at 14.0 (± 2.26), F258 at 14.0 (± 2.31), and Y198/F258 at 14.0 (± 2.32). While some variants display fluctuations in hydrogen bond counts relative to WT, these variations remain within a comparatively narrow range.

The SASA for the protein component reveals some notable features between the variants (Figure 5.24.A). As anticipated, the Apo-protein exhibits a significantly higher SASA (162.55 ± 3.51 nm^2) compared to WT (157.14 ± 3.75 nm^2), most likely due to the absence of BV, which normally stabilises the protein structure and reduces solvent exposure. This increased SASA in the Apo-protein indicates greater surface area exposure, possibly resulting from the void left within the BV-binding environment. For variant D177, Simulation 1 shows a significantly higher SASA compared to WT, which might suggest an altered conformation or an increased exposure of normally buried residues. This could be due to local structural disruptions caused by the mutation, which may affect how certain domains or loops are positioned relative to the solvent. In contrast, variant F258 in Simulation 3 exhibits a significantly lower SASA. This decrease could be due to tighter packing or a more closed conformation in this particular simulation, potentially influenced by specific interactions or conformational changes induced by the mutation that reduce the exposure of the protein to the solvent. However, since the significant differences for both D177 and F258 is observed in only one out of three simulations for each variant, this suggests that these variations may stem from temporary structural fluctuations rather than consistent changes in solvent exposure caused by the mutations themselves.

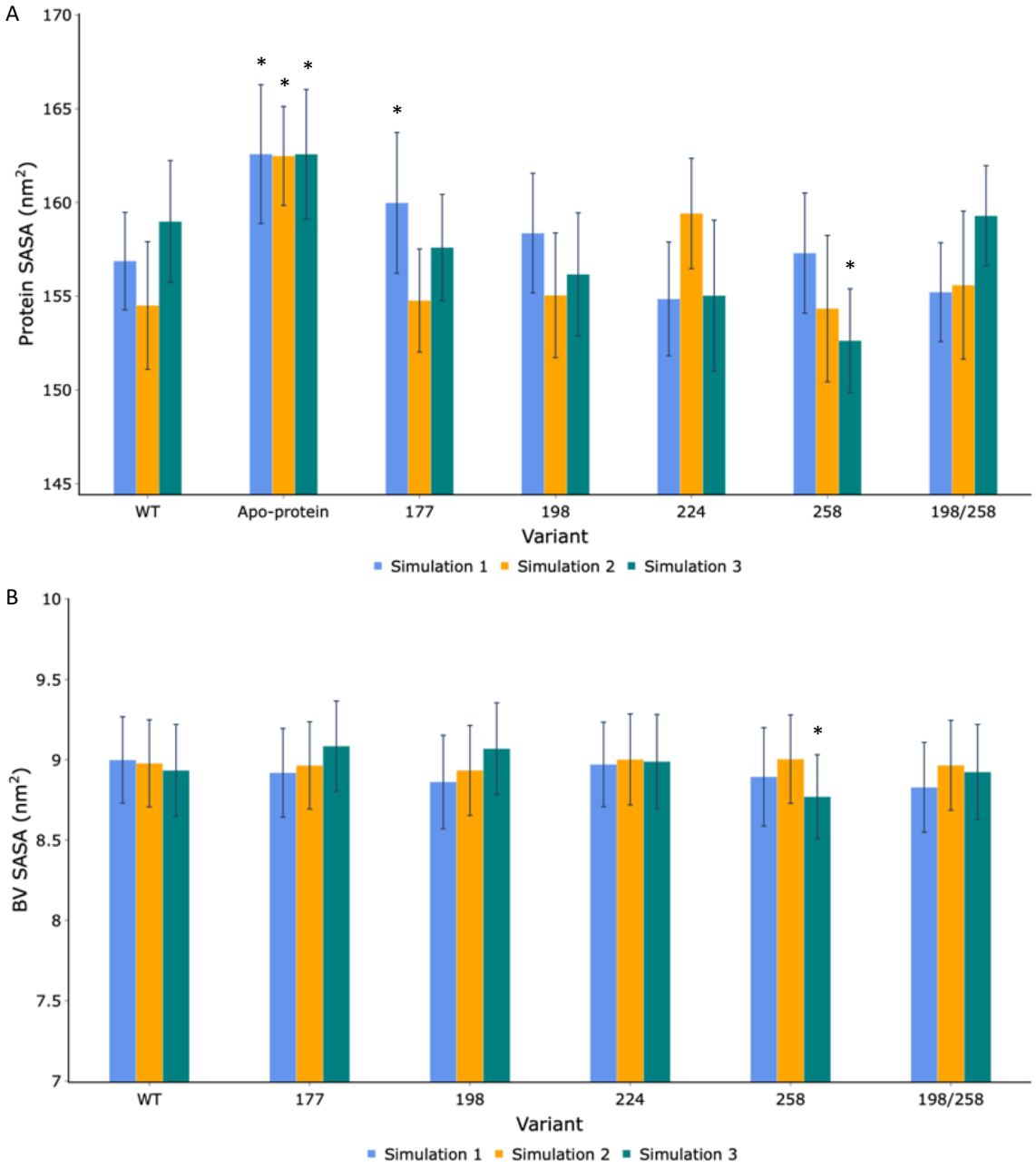


Figure 5.24: Average SASA of protein and BV calculated over 50-500 ns time frame. (A) Protein SASA analysis and (B) BV SASA analysis with all three simulations are represented by their mean averages and the error bars represent the standard deviation of the simulation. Kruskal-Wallis test was performed on each simulations to compare against each WT simulation, with * denoting significance ($p < 0.0001$). All individual SASA graphs can be seen in Figures S28-37.

The BV SASA analysis indicates a consistency in solvent accessibility across the variants, with most simulations displaying values similar to WT. The average BV SASA for WT is 8.98 ± 0.28 nm^2, with comparable values observed in variants D177 (9.00 ± 0.29 nm^2),

Y198 (8.98 ± 0.30 nm^2), N224 (8.99 ± 0.28 nm^2), and Y198/F258 (8.93 ± 0.29 nm^2). However, in Simulation 3 of variant F258, BV SASA is significantly lower than WT, with an average of 8.91 ± 0.30 nm^2. This reduction is consistent with the lower protein SASA observed in Simulation 3, suggesting that variant F258 may adopt a slightly more compact conformation in this specific simulation. Nevertheless, the lack of significant difference in the other simulations for variant F258 indicates that this decrease in SASA is likely a transient effect rather than a consistent structural shift. In the hydrogen bond analysis, no substantial changes in the number of hydrogen bonds between BV and the surrounding environment were observed across the variants. This stability in hydrogen bonding supports the expectation that BV SASA remains largely similar to WT.

The final analysis focussed on the Rg of both the protein and BV, to detect any expansion or contraction in the variants as a potential effect of pCNPhe incorporation. First the protein Rg was analysed where the average Rg for WT is 2.02 ± 0.016 nm, which serves as a baseline for comparison (Figure 5.25). Variants D177, N224, and F258 exhibit significantly higher Rg values, suggesting a more expanded protein conformation in these cases. For variant 177 (2.04 ± 0.018 nm), this expansion may be attributed to the mutations peripheral location within the protein structure, where the larger size of pCNPhe compared to aspartic acid will increase Rg on its own. In variant N224 (2.03 ± 0.018 nm), which is situated at the interface between the PAS and GAF domains, the mutation likely exerts a destabilising effect, as observed previously, potentially causing the domains to shift apart, and resulting in a higher Rg. Variant F258 (2.04 ± 0.018 nm), which resides within the BV-binding pocket, the mutation may induce slight structural adjustments that expands the pocket and increase the overall Rg. In contrast, variant Y198, which is also located within the BV-binding pocket, shows Rg values that remain consistent with WT across all simulations (2.02 ± 0.018 nm). This difference between Y198 and F258 suggests that the specific position and structural role of the residue within the binding pocket influence how each mutation affects the surrounding conformation. The double mutant Y198/F258 displays an average Rg of 2.03 ± 0.024 nm, with a notably larger standard deviation than the other variants. This increased variability, particularly the elevated Rg observed in simulation 3, suggests that the combined mutations may expand the protein, especially given their location within the BV-binding pocket.

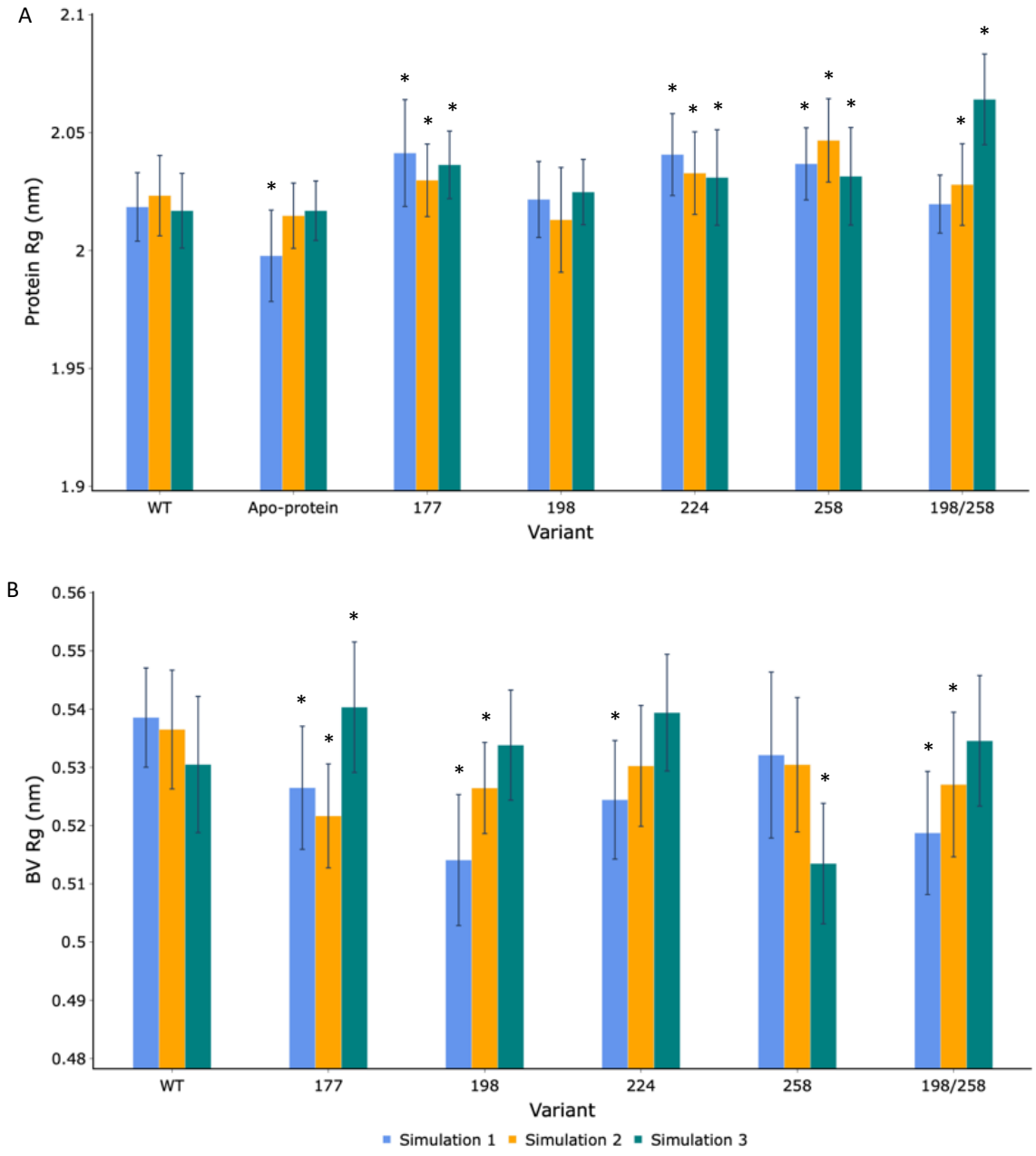


Figure 5.25: Average Rg of protein and BV calculated over 50-500 ns time frame. (A) Protein Rg analysis and (B) BV Rg analysis with all three simulations are represented by their mean averages and the error bars represent the standard deviation of the simulation. Kruskal-Wallis test was performed on each simulations to compare against each WT simulation, with * denoting significance ($p < 0.0001$). All individual Rg graphs can be seen in Figures S38-47.

The BV Rg analysis indicates that most mutations result in a more compact BV conformation compared to WT (Figure 5.25.B). The average Rg for WT is 0.536 ± 0.011 nm, while mutants generally show slightly reduced values: D177 at 0.531 ± 0.013 nm, Y198 at

0.528 ± 0.014 nm, N224 at 0.534 ± 0.012 nm, F258 at 0.528 ± 0.015 nm, and the double mutant Y198/F258 at 0.528 ± 0.014 nm. The analysis shows that a number of simulations across all mutants exhibit a significantly lower BV Rg compared to WT, indicating a general trend toward a more compact BV conformation in these variants. The only outlier is variant D177, where Simulation 3 displayed a higher Rg, suggesting an expanded BV conformation. This fluctuation in D177 may indicate that BVs conformation in this variant can shift between more compact and expanded states, while the other mutants consistently favour a more constrained BV structure. The fact that BV appears more compacted in most mutants, despite the protein Rg showing expansion, suggests that the mutations may restrict BVs internal flexibility while promoting a more open conformation in the overall protein structure.

To combine the complete MD dataset and reveal structural patterns with WT and mutant variants, a principal component analysis (PCA) was performed. This analysis used the median values and standard deviations from each measured metric. Two PCA plots were generated: one including both Protein and BV metrics (excluding the Apo-protein due to the absence of BV), and another including only Protein metrics (with Apo-protein included). This approach enables a clearer comparison of how BV-related metrics contribute to structural variations and helps distinguish the impact of specific mutations on both the Protein and BV-binding pocket. Excluding BV metrics also allows for direct comparisons between the Apo-protein and the variants, providing additional insights into how the absence of BV affects overall protein structure in relation to the mutations. No normalisation or scaling was applied to the metrics prior to PCA, so the resulting principal components reflect the natural scale of the input features. This means that structural metrics with higher numeric ranges may have influenced the variance more strongly.

The first PCA plot analysed both the Protein and BV data which captures 55.8% of the total variance in the first two principal components, meaning that these components describe over half the variability across all measured properties (Figure 5.26.A). WT is positioned distinctly on the left side of the plot, suggesting that the WT structure diverges from the variants in terms of the combined Protein and BV analysis, particularly along PC1, which captures the largest portion of the variance. This separation indicates that the mutations introduce structural variations that distinguish each variant from the native state. The other variants appear spread across both PC1 and PC2 without forming tight clusters,

indicating that each mutation influences the structural metrics in unique ways. The lack of clustering suggests that the structural effects of the mutations are not uniform and may affect different regions or aspects of the protein and BV-binding site. However, the PCA plot does show that the controls, D177 and N224, appear relatively close to one another, especially along the PC1 axis. While not tightly clustered, their relative proximity may indicate similar structural effects due to their shared positioning outside the BV-binding pocket. In contrast, mutations closer to the BV-binding site (Y198, F258 and Y198/F258) are also positioned relatively close to each other along PC1, suggesting they may have a more comparable impact on structural characteristics specific to the BV-binding environment. While Y198 and F258 themselves are not particularly close on the plot, the double mutant Y198/F258 sits between these points, supporting the idea that the combined mutation integrates structural features from both individual mutations. This positioning reinforces the point that mutations within or near the BV-binding site produce related structural effects that differ from those introduced by mutations further away.

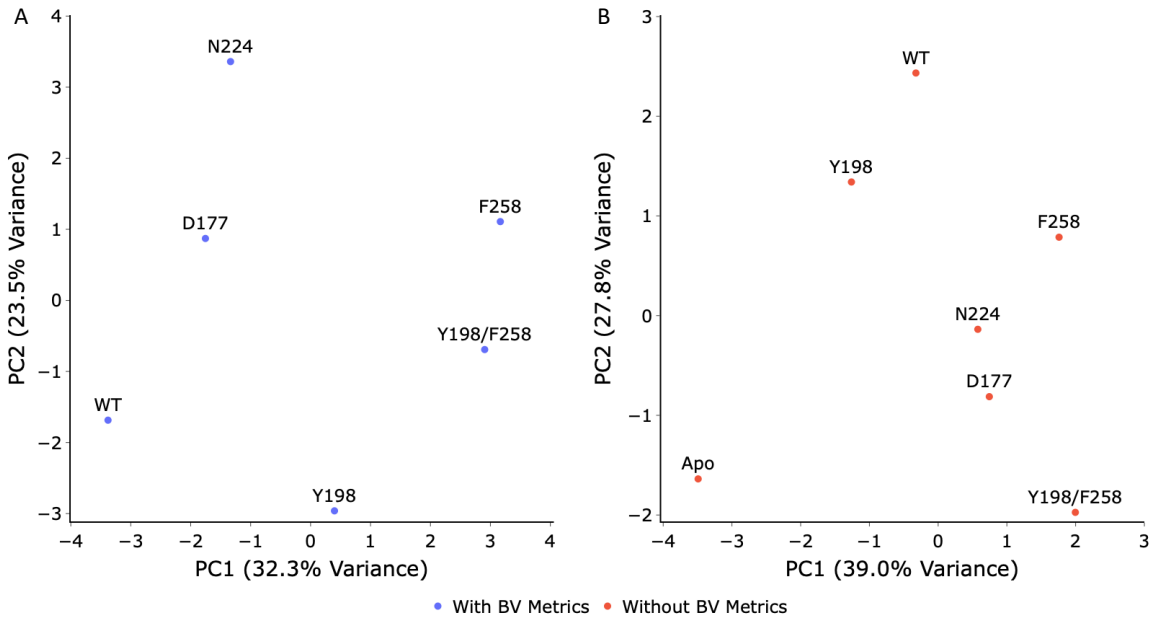


Figure 5.26: Principal Component Analysis (PCA) plots of the MD analysis. (A) PCA plot that includes all protein and BV analysis, which excludes the Apo-protein. (B) PCA plot that includes only the protein metrics so that Apo-protein could be plotted.

The second PCA plot analysed only the protein data which captures 66.8% of the total variance within this graph (Figure 5.26.B). This high variance indicates that the Protein-

only analysis provides a more comprehensive distinction between variants than when BV metrics were included, likely due to the clear separation of the Apo-protein from all other variants. Apo is positioned uniquely, both on the far left along PC1 and at the bottom along PC2, highlighting the significant structural divergence caused by the absence of BV. WT is located at the top of the plot, distinct from the mutants, but closest to Y198. The relative proximity of WT to Y198 suggests that this variant maintains a structure more similar to WT than the other mutations. Moving further from WT are F258, N224, D177, and finally the double mutant Y198/F258, which appears much more structurally divergent. This ordering differs notably from the previous PCA plot with BV data included, where WT and the variants displayed a different pattern of separation. In this Protein-only analysis, the variants are more dispersed along PC2 than PC1, indicating that PC2 captures much of the structural/dynamical variability introduced by the mutations. The positions of N224 and D177, are closer together and may cause similar structural impacts, which may reflect their positioning outside the BV-binding site. In contrast, the double mutant Y198/F258, positioned far from WT along both PC1 and PC2, suggests that the combination of these mutations drives a substantial difference from WT-like structure.

This MD analysis has highlighted how different pCNPhe mutations influenced the structural stability of the protein, particularly in terms of the BV-binding site and overall protein conformation. Although the initial control mutants were not ideal for comparative purposes, this workflow has demonstrated its potential for identifying whether future mutations preserve structural characteristics closer to WT. Going forward, this methodology offers a robust foundation for integrating computational analysis as a predictive tool in mutagenesis studies, enabling more efficient and targeted experimental designs.

5.2.3.4 Free-BV

In Chapter 4.2.1, the Raman profile of free-BV was measured, revealing no significant signals in the fingerprint region, and lacking absorptive properties in the required near-IR range, which is essential for epr-RS due to the Pump-Stokes laser set-up of the microscope. This contrasts with the mRhubarb720 WT, which displayed a distinct double peak at 1627-1652 cm^{-1}, as well as displaying a large extinction coefficient at 701 nm (72.5 $mM^{-1}cm^{-1}$).

Understanding the mechanisms behind this observation is crucial, as the enhancement of the Raman peaks in mRhubarb720 may be driven by the molecules dynamic behaviour.

The MD analysis performed on the WT-BV was also applied to free-BV, where an increase in movement was anticipated due to the absence of a protein scaffold to restrict the BV's conformation. In Figure 5.27.A-B, the RMSD of free-BV supports this idea as all three simulations display an average value of 0.25 nm with occasional spikes reaching up to 0.40 nm. The consistent fluctuations in all three simulations indicate a relatively high degree of movement and flexibility in the absence of a protein scaffold. The RMSD of BV within the WT stabilises at a maximum of 0.15 nm, with a much narrower range of deviation as represented by the SD. These RMSD results highlight the critical role of the protein scaffold in modulating the dynamic behaviour of the BV.

The RMSF values, observed in Figure 5.27.C, further confirm the increased movement in free-BV, as all atoms exhibit greater flexibility compared to the mRhubarb720 WT-BV. Despite this, the trend of atomic flexibility between free-BV and the WT-BV suggests that the intrinsic dynamic behaviour of BV is inherent to its structure, with the protein environment primarily reducing but not fully suppressing atomic fluctuations. For example, the carboxyl group atoms (9-10, 22-23 and 28- 29 remain the most flexible regions in BV. Atoms residing in the D-ring (21, 27, 33, 37 and 41) fluctuate on average about 0.05 nm more than in WT-BV. After minimisation and equilibration, the D-ring of free-BV adopts a Pfr state conformation, as shown in the RMSF map representation, therefore, the Pr/Pfr-state flipping is not expected to occur (Figure 5.27.D).

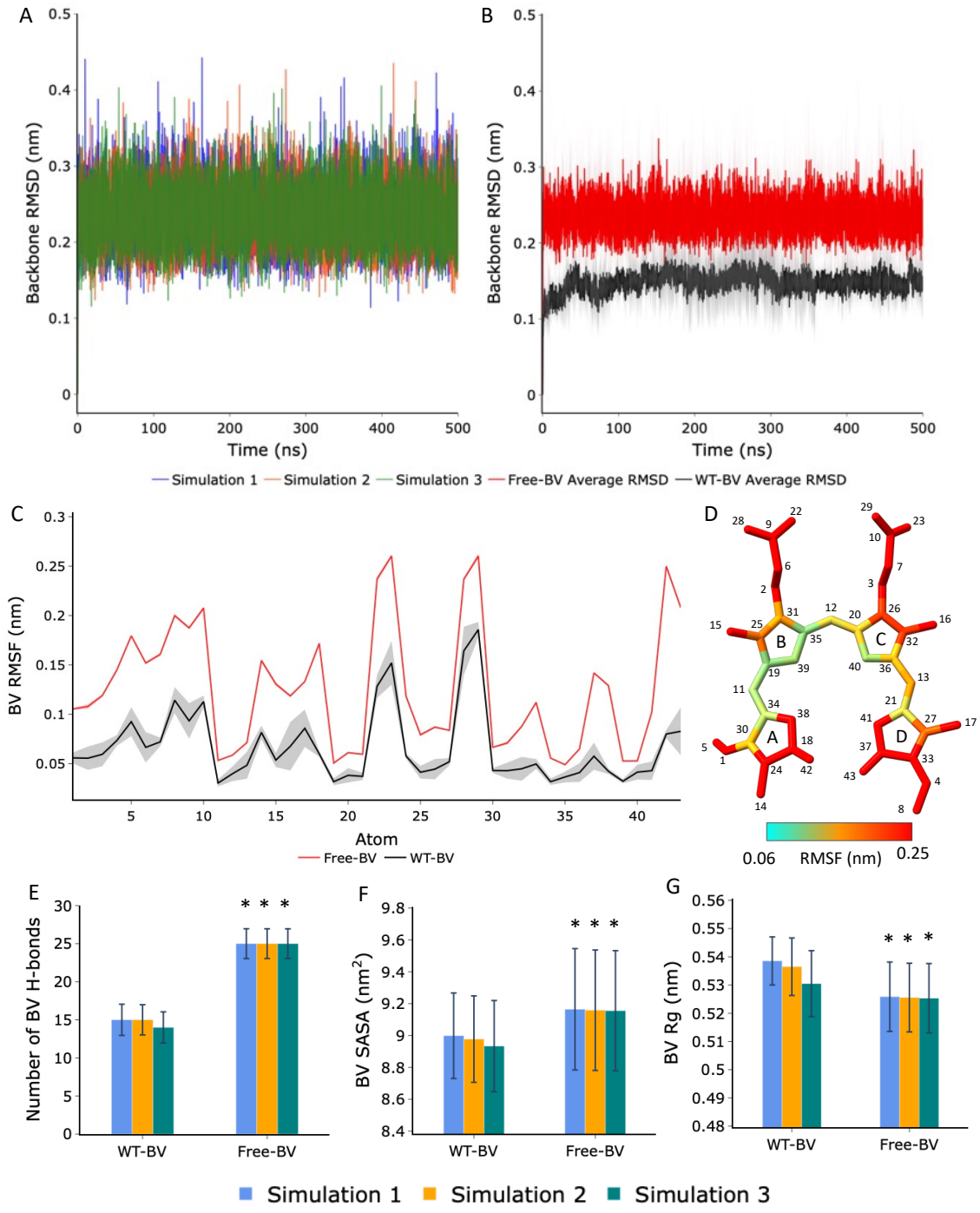


Figure 5.27: MD analysis of free-BV in respect to the WT-BV. (A) RMSD of individual Free-BV simulations. (B) Free-BV RMSD average of all three simulations (red) compared to the average WT-BV RMSD (black). (C) Average RMSF of all simulations of Free-BV (red) and WT-BV (black). (D) Free-BV RMSF mapping representation of Simulation 1. (E) Average number of H-bonds that involve BV in each simulation. (F) Average SASA of each simulation. (G) Average Rg of each simulation. Each averaged graph (C, E-G) is plotted from 50-500 ns time-frame and the Kruskal-Walis test was performed on plots E-G to compare against each WT simulation, with * denoting significance ($p < 0.0001$). All individual simulation graphs (excluding RMSD) can be seen in Figures S48-51.

Free-BV in solution forms a significantly higher number of hydrogen bonds compared to WT-BV. This increase is likely due to the unrestricted access of water molecules in solution, allowing for more hydrogen bond interactions (Figure 5.27.E). In contrast, WT-BV is bound within the BV-binding pocket of mRhubarb, where hydrogen bonding is more limited due to the protein's structural constraints. The SASA for Free-BV is also significantly higher than for WT-BV, though the difference is relatively modest (Figure 5.27.F). This suggests that, while Free-BV in solution is more exposed to solvent, the BV-binding pocket within mRhubarb allows for partial solvent access. Finally, the Free-BV displays a reduced Rg compared to WT-BV, indicating a more compact structure (Figure 5.27.G). This compactness can be attributed to the Pfr-state D-ring, which is isomerised compared to WT, causing the nitrogen (atom 41) to orient inward toward the other pyrrole groups. The protein-bound state likely restricts BVs conformational flexibility, maintaining a slightly expanded structure optimised for interaction with the surrounding residues within the binding pocket. These observations highlight the contrasting structural properties of BV in solution versus bound states. In solution, BV interacts freely with surrounding water molecules, leading to increased hydrogen bonding and a more compact conformation. Within the protein, the BV-binding pocket regulates BVs exposure to solvent and maintains an expanded structure, emphasising the role of the protein environment in modulating BVs conformation and stability.

While the MD simulations provide valuable insights into the dynamics and flexibility of mRhubarb720 and the BV, these results alone do not fully explain the spectral properties observed in the previous chapters. To gain a deeper understanding between the WT-BV and free-BV, the electronic structure will be investigated via a QM approach in Section 5.2.4.

5.2.4 Quantum mechanical approach to understand the electronic properties of BV

The final set of analyses aimed to predict the optical properties of BV in both its mRhubarb720-bound conformation and as Free-BV, using QM methods. These predictions aim to reveal how the local environment in mRhubarb720 influences BVs properties and to predict the experimentally observed spectral characteristics, including absorption and Raman profiles. This computational workflow offers a valuable framework for future

experimental work, enabling predictions that can guide experimental planning and analysis more effectively.

Due to the resolution of the mRhubarb720 crystal structure discussed in Chapter 3.2.3, it was not possible to determine the exact protonation state of BV when bound to mRhubarb720. Therefore, these QM simulations, BV with the deprotonated C-ring was used, as this is the protonation state BV IX-α forms when catalysed from heme (Whitby et al. 2002). Additionally, given that the crystal structure did not resolve the positions of water molecules within the BV-binding pocket, a Polarisable Continuum Model (PCM) using water was included to replicate a biological environment. This allowed for a more accurate simulation of the solvation effects on the chromophores electronic structure. The geometry optimisation involved a three-step process: HF/3-21G followed by B3LYP/6-31G and finally B3LYP/6-31G+(d,p) to ensure that geometry was sufficient for accurate frequency simulations. The following single-point energy, frequency and TD-DFT calculations were also conducted using the B3LYP/6-31G+(d,p) method and basis set.

For the modelling of the BV-bound mRhubarb720 system, the R-groups of three key residues were included: Cys15, the covalent attachment site for BV; R249, which stabilises one of the carboxyl groups extending from the B-ring via H-bonding; and H285, which stabilises the D-ring into a Pr conformation also via a H-bonding interaction (Figure 5.28.A). These residues were fixed during the geometry optimisation to maintain the interactions. Other interacting residues, detailed in Section 5.2.3.1, such as R217, S252 and S269 were left out of the simulation to reduce the simulation times. Their exclusion was also a practical decision due to the complexity of the system and the challenges involved in large systems during the optimisation phase. In contrast, the Free-BV lacks additional protein components, so it remains unconstrained and was optimised using the same process (Figure 5.28.B).

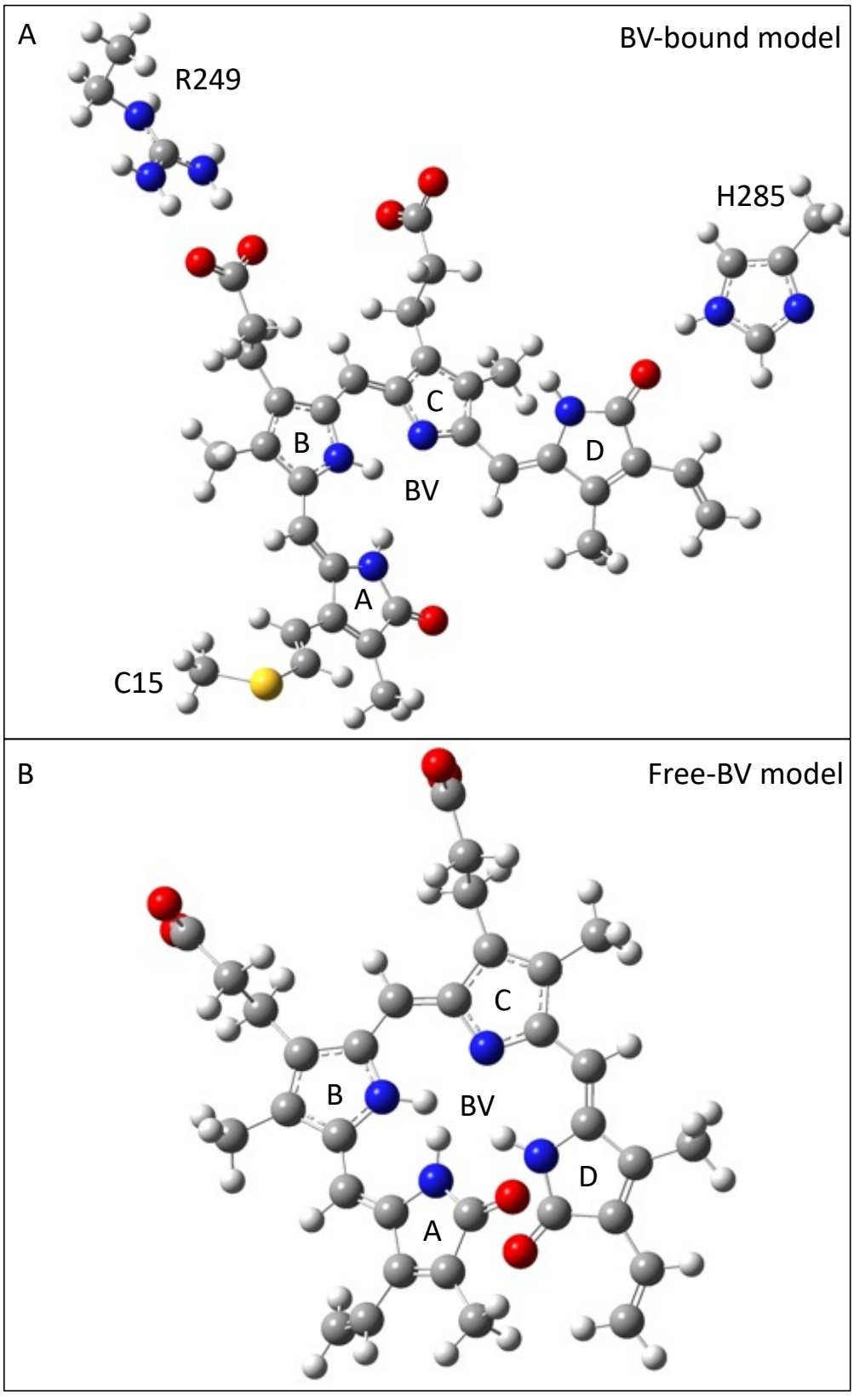


Figure 5.28: Geometry optimised QM models. (A) BV-bound model includes three interacting residues that are fixed in position whereas (B) free-BV model had no constraints. Both models are labelled with their ring nomenclature.

The most notable distinction of the geometry optimised structures is the Pr-state D-ring conformation in the BV-bound form, where the oxygen atom on the D-ring forms a H-bond with H285, effectively stabilising this conformation. This is supported by the Rg analysis from the MD simulations which shows the BV-bound form to adopt an extended conformation of the Free-BV (see Section 5.2.3.3). In contrast, the geometry optimised Free-BV model, relaxes into a more compact conformation with each pyrrole ring facing inward towards the centre of mass. Despite conformational changes, the conjugated double bond system remained intact and consistent in both the BV-bound and Free-BV models. This consistency suggests that BV's resonance structure, essential for its optical properties, is preserved even with structural shifts, such as potential cis-trans isomerisation. While these shifts can affect electronic properties, they do not disrupt the conjugation itself. The geometries for both BV states can be observed in Table 5.2.

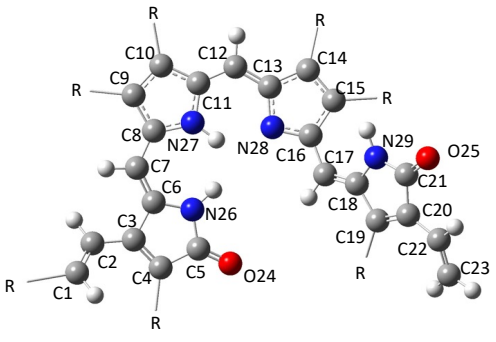
Table 5.2: Geometry optimisation of selected bond lengths, bond angles and dihedrals. Schematic of the atom labels are also shown below, R represents the groups incorporated not involved in the core electronic structure.

Selected bonds	Bond Length (Å)	
	BV-bound	Free-BV
C1-C2	1.31592	1.34577
C2-C3	1.46897	1.46184
C3-C4	1.33360	1.37633
C4-C5	1.49223	1.47469
C5-O24	1.21591	1.25541
C5-N26	1.37638	1.39731
C6-N26	1.39762	1.40403
C3-C6	1.48958	1.47706
C6-C7	1.32859	1.36836
C7-C8	1.44631	1.42874
C8-C9	1.38766	1.42669
C9-C10	1.41586	1.41139
C10-C11	1.38612	1.42673
C11-N27	1.37621	1.38969
C8-N27	1.36182	1.37918
C11-C12	1.42343	1.40931
C12-C13	1.34197	1.38700
C13-C14	1.46391	1.46384
C14-C15	1.34848	1.38153
C15-C16	1.48091	1.46262
C16-N28	1.29851	1.36143
C13-N28	1.42368	1.40667
C16-C17	1.47235	1.43321
C17-C18	1.32570	1.36779
C18-C19	1.49401	1.46589
C19-C20	1.33959	1.37930
C20-C21	1.49175	1.49566
C21-O25	1.21937	1.25588
C21-N29	1.35808	1.38730
C18-N29	1.40018	1.39705
C20-C22	1.46787	1.45134
C22-C23	1.32093	1.34895

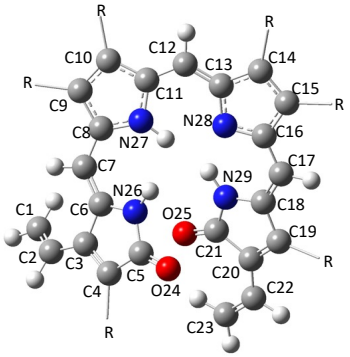
Selected angles	Bond Angle (°)	
	BV-bound	Free-BV
C1-C2-C3	122.997	127.126
C2-C3-C6	121.963	126.272
C3-C6-C7	125.518	127.918
C6-C7-C8	127.798	127.834
C7-C8-C9	126.827	127.771
C8-C9-C10	107.505	107.889
C9-C10-C11	107.084	107.217
C10-C11-C12	129.577	130.858
C11-C12-C13	126.570	125.920
C12-C13-C14	126.845	126.324
C13-C14-C15	106.748	106.146
C14-C15-C16	106.250	106.625
C15-C16-C17	124.582	124.586
C16-C17-C18	121.921	125.919
C17-C18-C19	129.369	128.559
C18-C19-C20	108.341	108.365
C19-C20-C22	133.097	127.382
C20-C22-C23	125.660	126.303
C5-N26-C6	111.475	111.232
C8-N27-C11	109.823	110.285
C13-N28-C16	106.715	105.756
C18-N29-C21	111.637	111.477

Selected dihedrals	Dihedral Angle (°)	
	BV-bound	Free-BV
C3-C6-C7-C8	-179.458	-175.148
C6-C7-C8-C9	157.408	-161.567
C10-C11-C12-C13	-176.935	-175.069
C11-C12-C13-C14	172.311	-171.440
C15-C16-C17-C18	49.112	-165.796
C16-C17-C18-C19	-175.716	-179.698

BV-Bound



Free-BV



The bond lengths, bond angles, and dihedral angles of BV-bound and Free-BV chromophore models reveal distinct structural differences that may influence their Raman intensities and overall vibrational behaviour. The bond length data indicates that BV-bound model exhibits shorter average bond lengths compared to Free-BV, with an average decrease of 0.013 Å across all bonds. Although this difference is subtle, it may impact Raman intensities. Shorter bond lengths in the BV-bound corresponds to stronger bonding as the bond order between the models is exactly the same. Shorter bonds can result in sharper Raman peaks and enhanced polarisability changes during vibrations when electrons between atoms are closer together are shared more freely (Kolesov 2021). In contrast, the slightly longer bond lengths in Free-BV suggest reduced vibrational coupling and weaker polarisability changes, due to the increased distances electrons have to travel. The bond angle analysis indicates that the angles are relatively similar between BV-bound and Free-BV, with minimal deviations observed across the chromophore structure.

The dihedral angles provide insights into the relative alignment and planarity of the aromatic rings within each system. Dihedral angles represent the torsional relationship between two planes formed by four atoms, and their periodic nature (ranging from -180° to +180°), which means that differences exceeding 180° must be adjusted by subtracting them from 360°. This adjustment ensures that the angular differences are calculated correctly, reflecting the smallest rotation required to align the two planes. For example, the dihedrals connecting rings A and B are C3-C6-C7-C8 and C6-C7-C8-C9, and in the BV-bound model, these angles are -179.458° and 157.408°, respectively, resulting in a difference of 23.13°, indicating a near-planar relationship with slight torsional distortion. In Free-BV, the corresponding angles are -175.148° and -161.567°, with a smaller difference of 13.58°, suggesting better planarity between these rings. For rings B and C, the dihedral angles C10-C11-C12-C13 and C11-C12-C13-C14 were examined. In BV-bound, these angles yield a small difference of 10.75°, while in Free-BV, the angles are even smaller with a difference of 3.63°. These results indicate that rings B and C are planar to each other in both systems and this high degree of planarity likely facilitates conjugation and electron delocalisation between these rings, which is crucial for maintaining the chromophores electronic properties. The most significant differences are observed in rings C and D, represented by the dihedral angles, C15-C16-C17-C18 and C16-C17-C18-C19, where the BV-bound model has a large difference of 135.17°, indicating substantial misalignment and torsional strain in this region.

In Free-BV, a far smaller difference of 13.90° is observed showing that these rings are comparatively coplanar. The substantial distortion in BV-bound is caused by interactions with the H258, which imposes structural constraints via hydrogen bonding and stabilises the Pr state chromophore. Due to the fixed nature of the BV-bound model, which constrains the molecules geometry, may have an impact on increasing polarisability across the molecule.

For the following results, it is important to note that single-point energy calculations for determining the dipole moments and molecular orbitals in the BV-bound model were performed without the interacting residues R249 and H285. This approach was taken to assess how different conformations affect BVs inherent electronic properties without potential influences from nearby residues, which could otherwise alter the dipole moment values and orbital maps. The C15 residue was still included, as it is covalently bonded to BV. Nevertheless, for calculations involving predicted excitation energies, TD-DFT, and predicted Raman spectra, still included these interacting residues for the BV-bound model. Their presence is essential in these cases because they contribute to the overall spectral and excited-state properties, ensuring that the full electronic environment around BV is captured in these specific analyses.

5.2.4.1 Electronic properties

Understanding the electronic properties of each conformation will be crucial for predicting their optical behaviours observed in the experimental data. The dipole moments of BV-bound and Free-BV were 63.97 Debye (D) and 58.06 D, respectively. The fact these values are relatively close suggests that the optical properties of the two conformations are not heavily influenced by the changes in dipole moments. The direction vectors of both models were fairly similar and can be observed in supplementary Figure S73. This supports the hypothesis that the structural/environmental factors, rather than the differences in charge distribution, are responsible for the divergent optical characteristics observed.

The energy levels of the highest occupied molecular orbital (HOMO; ground state, S_0) and the lowest unoccupied molecular orbital (LUMO; excited state, S_1) of each model was also calculated (Figure 5.29). The energy gap between these states is critical in determining the excitation energy required undergo an electronic transition, which is related to the molecule's absorption profile. Additionally, the HOMO and LUMO orbitals can also be

visualised to reflect the regions in the molecule where electrons are likely to be found during a transition between ground and excited states.

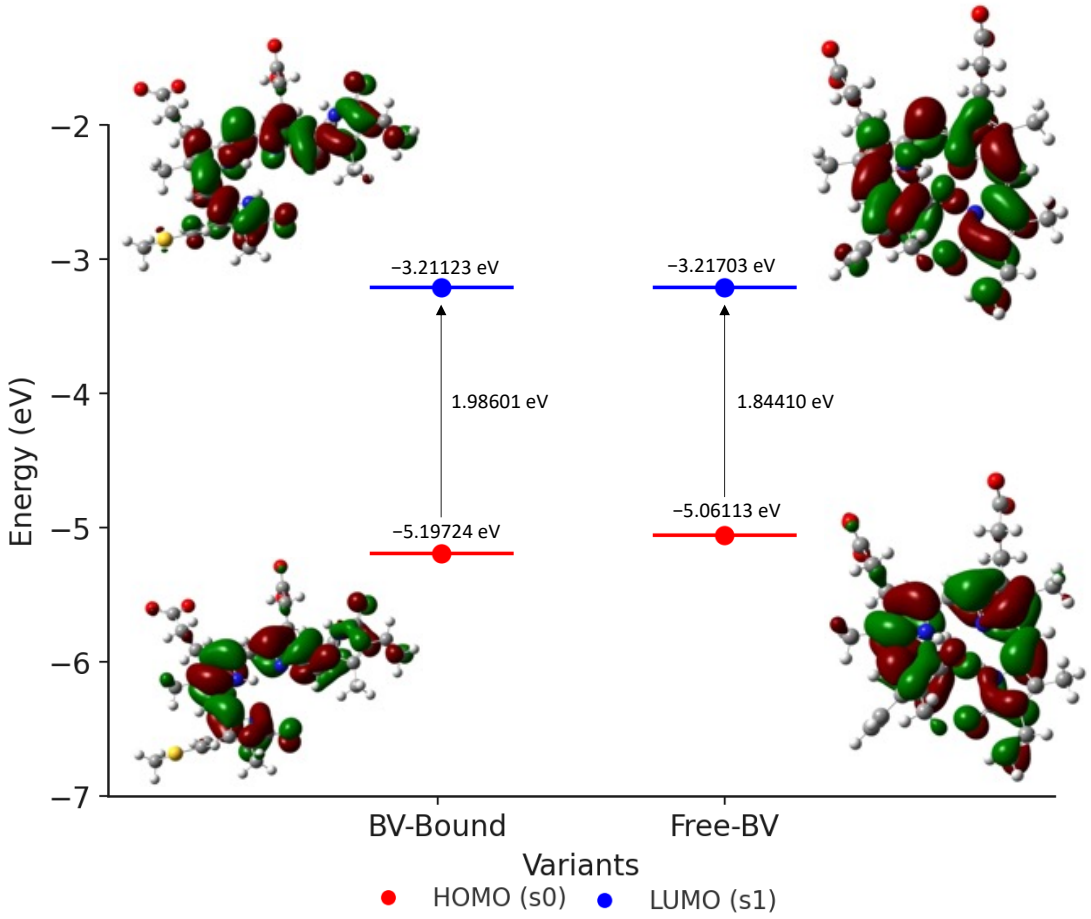


Figure 5.29: HOMO-LUMO energy levels and orbitals for BV-bound and free-BV model. Each energy level is labelled with the corresponding energy in eV, with the HOMO-LUMO gap labelled as eV. Red lines represent the ground state HOMO (S_0) level whereas blue lines show the excited state LUMO (S_1) level. The molecular orbitals of each state displaying the distribution of orbitals in HOMO (bottom) and LUMO (top).

The calculated HOMO energy for the BV-bound model is -5.19724 eV, while for free-BV, it is -5.0113 eV. Whereas the LUMO energy for both forms remain fairly consistent between the two (BV-bound; -3.21123 eV, free-BV; -3.21703 eV). The slight increase in HOMO energy in free-BV leads to a smaller HOMO-LUMO gap of 1.8441 eV, compared to 1.98601 eV for the BV-bound model. This suggests that the excitation energy for free-BV is reduced and aligns with a more redshifted profile, which has been observed in other literature for Pfr models (Modi et al. 2019). Therefore, the protein environment must alter

the ground state so that it is stable in a lower energy, compared to free-BV. However, DFT-calculated excitation energies do not fully capture the maximum absorption wavelengths as factors such as solvent interactions, excited state relaxation and the interactions between electronic and vibrational states, which will be mitigated by TD-DFT calculations in the next section (Jacquemin et al. 2011; Adamo and Jacquemin 2013; Escudero et al. 2017).

Despite the significant conformational changes observed between the two states, the molecular orbitals show minimal differences in their electron distribution, with the electron density primarily localised around the conjugated system for both the HOMO and LUMO. This suggests that the change in conformation does not considerably alter the electronic transitions within the chromophore itself. However, while the electron distributions might not show significant differences visually, small shifts in orbital energy levels can still influence the optical properties, as seen in the HOMO-LUMO gap.

5.2.4.2 Predicting absorbance profiles

Comparing the computational results with experimental spectra will reveal trends and differences that can refine our understanding of the chromophores behaviour in different conformations. The predicted UV-Vis spectra showed reasonable similarity with the experimental data with both models exhibiting two characteristic peaks: the blue-shifted Soret band and the red-shifted Q-band, commonly observed in BV-binding proteins (Polyakov et al. 2018; Ghosh et al. 2022). The predicted Q-band absorption maximum for the BV-bound form occurs at 690 nm, showing an 11 nm blue-shift compared to the experimental maximum at 701 nm (Figure 5.30). For the Soret band, the maximum for the BV-bound is at 460 nm, whereas the experimental value is at 395 nm. This overestimation of the Soret band wavelength is the only significant difference between the predicted and experimental data, likely due to the absence of the interacting residues in the simulation. These interacting residues may contribute to further stabilisation of the chromophores electronic transitions, particularly for the higher energy Soret band. The predicted free-BV shows a more prominent Soret band at 392 nm, which is relatively close to the experimental findings at 385 nm. This simulation also displays a broad weaker peak for the Q-band emulating what we observe in the laboratory.

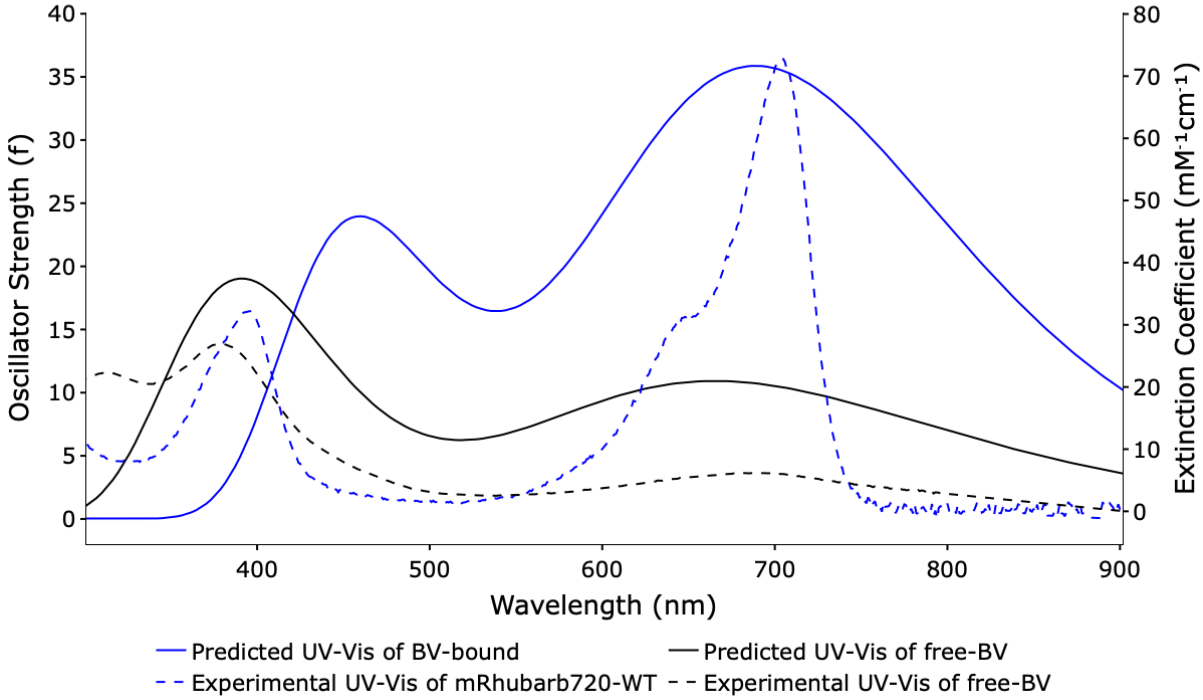


Figure 5.30: Predicted UV-Vis spectroscopy against experimental absorbances. TD-DFT calculation (nstates = 10) of each model; BV-bound (blue) and free-BV (black), overlayed onto the experiment mRhubarb720-WT data (blue dashed line) and experimental free-BV (black dashed line). Oscillator strength (f) is an arbitrary, dimensionless unit for the QM analysis, with excitation coefficient ($mM^{-1}cm^{-1}$) units for the experimental data.

The peak intensity for the models follows the trend of showing a more prominent Q-band when BV is in a Pr state (BV-bound). The opposite is observed for the Pfr state (Free-BV) with both the predicted and experimental data showing a higher intensity for the Soret band. This inversion of intensities between the Soret and Q-bands for Free-BV suggests a significant change in the chromophores absorption characteristics when unbound from the protein. The behaviour is typical of free chromophores exposed to aqueous environments, where conformational flexibility and solvent interactions can lead to a more dominant absorption in the higher energy Soret band, compared to the more stabilised red-shifted Q-band seen in the bound form (Maillard et al. 2021). Despite the deviation observed for the Soret band on the BV-bound model, the overall trends in the spectra are replicated between the predicted and experimental data, suggesting that the TD-DFT calculations are effective for capturing the optical behaviour of these systems at higher wavelengths.

5.2.4.3 Predicting Raman activity

The predicted Raman activity was examined to compare with experimental epr-SRS within the fingerprint region. These simulations will provide insight into which bonds are electronically enhanced by the protein and help identify additional peaks that may be susceptible to enhancement. In the experimental data, only the 1550-1950 cm^{-1} range of the fingerprint region was analysed, whereas the predicted spectra will cover the entire fingerprint region, offering a broader perspective on the vibrational modes of BV. To ensure consistency, the predicted Raman spectra was first compared specifically to the experimental range (Figure 5.31). The predictions show good agreement in terms of peak positions, with the mRhubarb720 WT experimental data showing enhancement at 1627-1652 cm^{-1}, with the simulated BV-bound Raman activity showing peaks at 1615-1655 cm^{-1}. The most notable finding was the peak intensities as the largest peak for the BV-bound form was at 1615 cm^{-1} with an intensity of ~12,000 a.u., while the predicted free-BV showed a maximum peak at ~1650 cm^{-1} with an intensity of ~3,000 a.u., representing a 4-fold difference between the two models. In the experimental data, the largest peak difference for mRhubarb720 WT and free-BV is a ~8-fold difference, with the largest WT peak at 1627 cm^{-1} displaying a 1 mV intensity and the largest free-BV peak at 1650 cm^{-1} maxing at 0.12 mV.

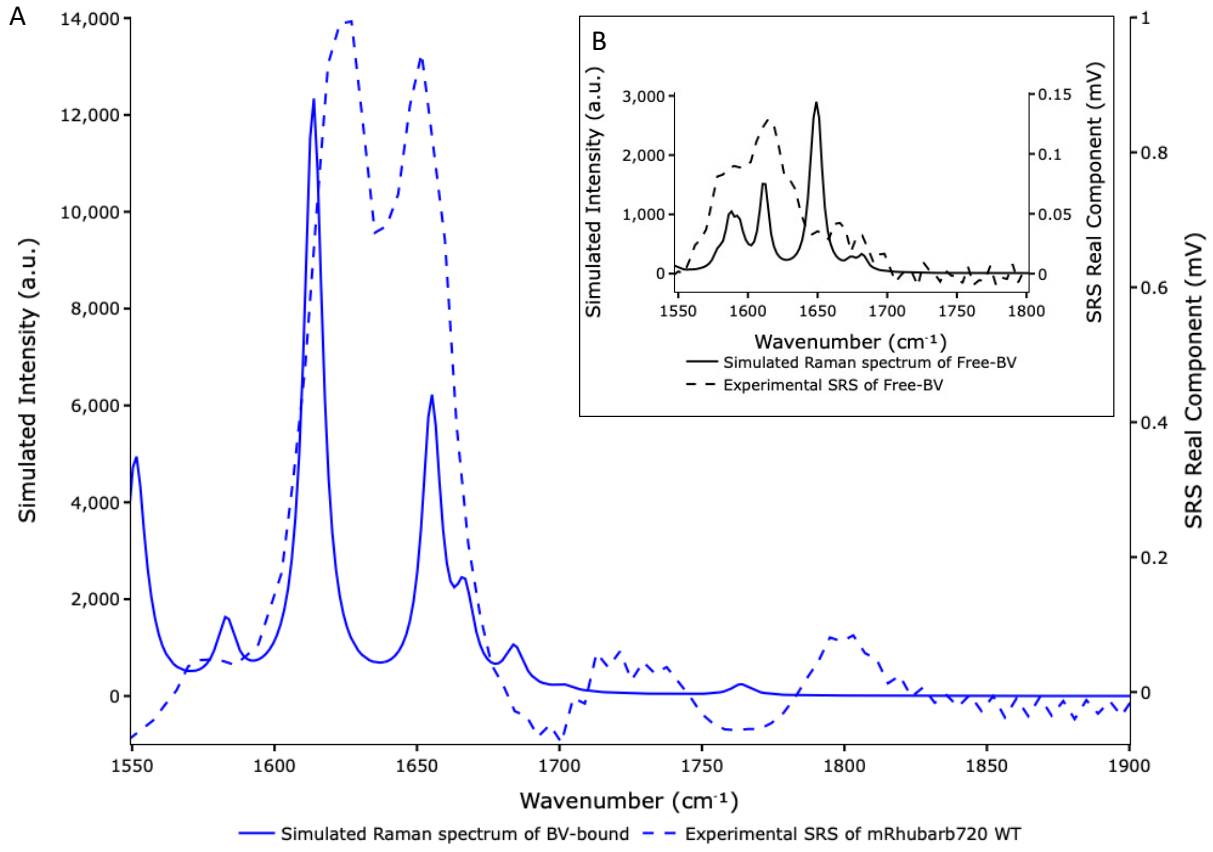


Figure 5.31: Comparison of predicted and experimental Raman spectroscopy. (A) The BV-bound simulated spectra (blue) is compared to the epr-SRS mRhubarb720 WT measurements (blue dashed line). (B) The free-BV simulated spectra (black) is compared to the epr-SRS of free-BV (black dashed line). A scaling factor of 0.97 was applied to the wavenumbers for the simulated spectra.

The 4-fold difference between the predicted Raman intensities for BV-bound and free-BV is significant since the measurements are of the same chemical bonds in different conformations. This variation suggests that the protein scaffold in mRhubarb720 significantly influences the vibrational behaviour of BV. The larger 8-fold difference observed in the experimental data can be attributed to the enhancement caused by epr-SRS, which is not captured by the computational models. Moreover, for the free-BV model, the flexibility of the chromophore in an aqueous environment likely reduces the coupling between its electronic and vibrational modes, explaining the lower Raman intensities observed. This flexibility is less evident in the BV-bound form due to the structural constraints, leading to more pronounced vibrational activity and therefore, higher intensities in both the experimental and predicted spectra.

Finally, the entire simulated fingerprint region was compared between the BV-bound and free-BV conformations in order to examine the differences in vibrational mode behaviours and identify the specific interactions that may be responsible for the enhanced Raman activity in the BV-bound conformation (Figure 5.32). Assigning specific vibrational modes to particular wavenumbers in Raman spectra can be ambiguous, as the literature often provides differing interpretations especially for complex systems like BV. Therefore, the following discussion presents an interpretation of the vibrational modes based on the current QM simulation, recognising that the exact assignments may vary depending on the chromophores environment and the specific parameters of the computational setup.

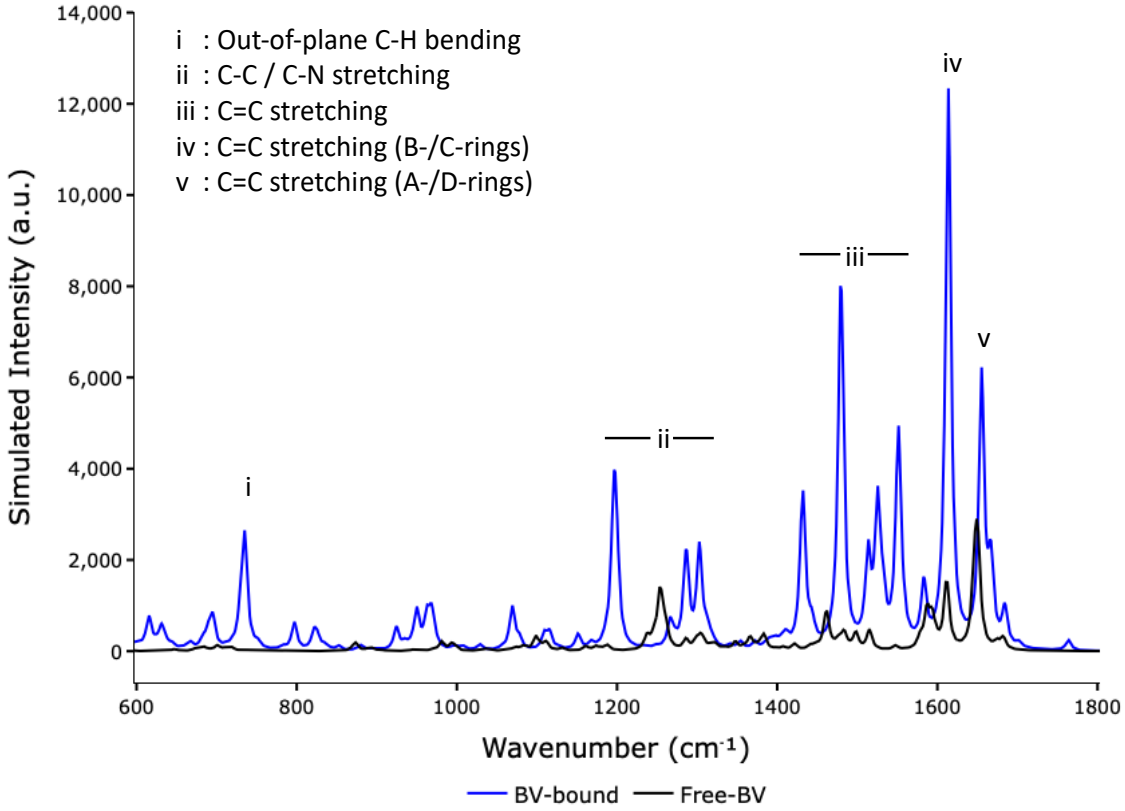


Figure 5.32: Full fingerprint region of predicted Raman spectroscopy. The simulated spectra of BV-bound (blue) is compared to free-BV (black). Selected peaks are labelled (i) Out-of-plane C-H bending, (ii) C-C/C-N stretching, (iii) C=C stretching, (iv) C=C stretching (B-/C-rings), (v) C=C stretching (A-/D-rings). A scaling factor of 0.97 was applied to the wavenumbers for the simulated spectra.

As previously mentioned, the BV-bound spectrum displays significantly higher intensities compared to free-BV, signifying the influence of the protein scaffold in

mRhubarb720. In the figure, the largest peaks are highlighted, with the predicted vibrational modes responsible for the Raman activity. The peak at 734 cm^{-1} (i) can be attributed to the out-of-plane bending of C-H bonds, which will most likely arise from the methine bridge that connects the pyrrole rings in BV (Mroginski et al. 2006; Yamada and Mizuno 2018). The range of peaks observed at 1196-1285 cm^{-1} (ii) corresponds to C-C and C-N stretching found in the pyrrole rings of BV (Hu et al. 2015; Sato and Martinho 2018). These stretches are enhanced due to their inclusion into the delocalised system, which increases their vibrational activity and also the presence of three protonated nitrogens on the pyrrole rings will increase this intensity. The higher wavenumber peaks correspond to C=C double bond stretching and display the highest intensities across the fingerprint region, which is expected due to the conjugated nature of BV. The 1432-1551 cm^{-1} (iii) wavenumber range can be attributed to C=C stretching, particular in the non-aromatic bonds in the methine bridge as well as the bonds extending away from the A- and D-rings (Edwards 2006). An important note for the wavenumber ranges at (ii) and (iii) is that the inclusion of R249 and H285 will increase the intensities at these peaks due to the increased number of bonds that contribute to these enhancements. The final labelled peaks at 1615 cm^{-1} (iv) and 1650 cm^{-1} (v) are indictive of aromatic C=C stretching, with the higher intensity peak (iv) likely be associated with highly conjugated regions in the B- and C-rings and the less intense peaks (v) linked to the less resonant (fewer double bonds) aromatics in A- and D-rings (Neugebauer et al. 2012; López et al. 2022). However, another published study has associated these peaks to C=C stretching between each ring in the methine bridge, contradicting the assignment of these vibrational modes to the aromatics (Mroginski et al. 2007). The predicted Raman spectra has outlined the importance of expanding the wavenumber range in the future to 1100-1700 cm^{-1} as this should capture other vibrational modes beyond the recognised double peak at 1627-1652 cm^{-1}. There are alternative vibrational modes that may offer new opportunities for enhancement and ultimately improving the specificity for BV imaging.

The QM computational methods employed here, using the B3LYP/6-31G+(d,p) basis set, allow for a realistic estimation of electronic properties by accounting for solvent effects using the PCM model. This approach provides predictive insights into dipole moments, excitation energies, absorption spectra and Raman profiles, with a level of accuracy that can help guide experimental work. The ability to simulate specific conditions, such as solvent interactions, adds a layer of flexibility to the analysis, helping isolate the effects of certain

environmental factors on the chromophores electronic behaviour. However, these methods also have inherent limitations. The PCM model provides an implicit solvent approximation, which may not capture the full range of solvent-protein interactions present in a real biological system. Additionally, QM approaches like B3LYP/6-31G+(d,p) are computationally demanding, particularly for large systems or complex chromophores. Such demands can require model truncations to balance feasibility with accuracy, potentially omitting parts of the chromophores environment that influence its electronic properties. Additionally, while B3LYP/6-31G+(d,p) is well-regarded and effective for conjugated and charged systems, DFT methods in general may sometimes underrepresent specific excited-state interactions, particularly when extreme charge delocalization or subtle resonance effects are involved.

5.3 Conclusion

The experimental analysis has demonstrated that through-space enhancement using pCNPhe is not feasible within the biologically silent region. The lack of Raman enhancement observed in this region indicates that the introduction of pCNPhe does not generate the desired vibrational coupling with BV, therefore limiting its use for enhancement-based imaging in this range when not integrating directly into the conjugated system. Furthermore, the incorporation of pCNPhe appears to reduce the efficiency of BV incorporation, as suggested by the mass spectroscopy and the spectral data via the decreased extinction coefficients for the mutants compared to the WT. This reduction likely stems from structural alterations introduced by pCNPhe, which hinder the BV interaction with the protein scaffold.

The MD simulations provided a comprehensive workflow to analyse WT in respect of the variants in this study, assessing structural metrics to understand each mutations impact. Although the control mutations (D177 and N224) proved less suitable than anticipated, this approach effectively identified structural variations, providing a foundation for assessing whether future mutations exhibit WT-like characteristics. The BV analysis in the pCNPhe mutants did not reveal significant instability of BV, indicating that the lack of BV incorporation experimentally is likely unrelated to BV stability. Instead, BV-binding issues are more likely tied to protein folding or incorporation mechanisms rather than BV instability once within the protein. Additionally, the Apo-proteins distinct structure underscores the role of BV-binding in maintaining the proteins conformation.

On the other hand, the QM simulations reveal a significant increase in vibrational mode intensity in the BV-bound form, without the pre-resonance enhancement provided by epr-SRS. This indicates that the WT protein-bound environment contributes significantly to the vibrational properties of BV, enhancing its Raman-active modes through structural stabilisation of the Pr state. The QM simulations also suggest that focusing on the 1100-1700 cm^{-1} range for future epr-RS and Raman imaging may be advantageous as multiple vibrational modes have the potential to be enhanced. This provides new opportunities for optimising the imaging, particularly when focussing on mRhubarb720, by targeting additional vibrational modes beyond the experimentally studied peaks.

6 Exploring the structural and spectral properties of incorporating non-natural amino acids into the mCherry chromophore.

6.1 Introduction

In the previous chapters, the focus was on mRhubarb720, where electronic pre-resonant stimulated Raman spectroscopy (epr-SRS) experiments were successfully conducted in the fingerprint region, but the desired enhancement in the biologically silent region was not achieved, as through-space enhancement via the chromophore was not possible. This was attempted through the incorporation of para-cyano-phenylalanine (pCNPhe), but the results were inconclusive at best. In contrast, this chapter shifts focus to exploring an alternative method by directly incorporating non-canonical amino acids (ncAAs) into the chromophore of a fluorescent protein (FP), specifically mCherry. Incorporating ncAAs directly into the chromophore presents an optimal strategy for introducing Raman-active bonds, as the chromophore undergoes a spontaneous post-translational maturation mechanism involving three contiguous residues. These residues undergo covalent rearrangement in the presence of O_2 to form the chromophore. By replacing one of the chromophore forming amino acids with ncAAs, it is possible to introduce novel vibrational modes directly in the electronically active chromophore, providing a promising approach for enhancing Raman scattering in both the fingerprint and biologically silent regions. This chapter will delve into the experimental and computational methods used to evaluate the potential of ncAA-incorporated mCherry variants to achieve these enhancements.

Most β-barrel FPs, including mCherry, possess an inherent chemistry in their amino acid sequences that allows formation of a chromophore with fluorescent properties in the visible region of the electromagnetic spectrum without external cofactors like biliverdin XI-alpha (BV). The chromophore typically forms from a tyrosine, glycine, and a variable third residue (XYG motif). While tyrosine can be changed to other aromatics, including ncAAs without affecting chromophore maturation, the glycine is strictly conserved (Tsien 1998; Wang et al. 2003; Reddington et al. 2015; Hartley et al. 2016). Maturation occurs in a three-step process: cyclisation of the peptide backbone (between the glycine's amine nitrogen and

the carbonyl carbon of the variable residue), dehydration to form a double bond, and an oxidation step that forms the β-methylene bridge linking the phenolic and imidazolinone rings (Subach et al. 2011; Grigorenko et al. 2017). Upon electronic excitation, the chromophore emits a photon when the excited electron returns to the ground state. The emitted photon is red-shifted, as it has lower energy than the excitation photon (Subach and Verkhusha 2012). The extent of this red-shift depends on the number of double bonds in the chromophore's conjugated system. For instance, in superfolder GFP (sfGFP), the chromophore (Thr-Tyr-Gly) has excitation and emission peaks at 485 nm and 510 nm, respectively. In contrast, DsRed has a second oxidation step that dehydrogenates the Cα-N bond of the residue at position 66, forming an extra double bond (N=C), extending the conjugation resulting in red-shifted excitation and emission peaks of 540 nm and 570 nm (Gross et al. 2000). This ability to manipulate chromophore composition directly affects its spectral properties, a principle that has been widely used to develop a palette of FPs, ranging from blue (e.g. EBFP via GFP-derived FPs) to red (e.g. mCherry via DsRed-derived FPs).

The mutational target for the mCherry chromophore will be Y67 in order to incorporate tyrosine/phenylalanine derivatives into the phenyl ring since the chemical modifications will be bonded directly to the conjugated system responsible for electronic transitions. The three ncAAs selected for this research is, pCNPhe (from Chapter 5), para-ethynyl-phenylalanine (pCCPhe) and meta-nitro-tyrosine (3-NO2-Tyr) (Stokes et al. 2000; Tucker et al. 2006; Ahsan 2013). Both pCNPhe (nitrile; -C≡N) and pCCPhe (ethynyl; -C≡CH) contain vibrational modes aimed at being detected within the biologically silent region, whereas 3-NO_2-Tyr (nitro; -NO_2) is detected within the fingerprint region. Since there is no nitro groups within the conjugated system of the chromophore, there should be significant detectable enhancement within the fingerprint region.

The epr-SRS measurements for mRhubarb720 in earlier chapters were taken at Cardiff University with a pump pulse set at λ=820 nm. Since mCherry is more blue-shifted compared to mRhubarb720, a different laser system was used for this chapter's experiments, provided by Dr. Andreas Zumbusch's team at the University of Konstanz, Germany, where the pump pulse was set to λ=640 nm. With this laser wavelength being closer to mCherry's electronic transition, a more significant enhancement is expected compared to what was observed in Chapter 4 with mCherry. This proximity to the

chromophores absorption peak should result in a more effective coupling, enhancing vibrational modes in the Raman spectrum and allowing for more prominent pre-resonant Raman effects. As a result, stronger signals in both the fingerprint and biologically silent regions are anticipated, improving the overall sensitivity and signal-to-noise ratio in these experiments.

Molecular dynamics (MD) analysis will also be employed in this chapter, to evaluate the structural properties of the mCherry chromophore variants incorporating ncAAs. These simulations are designed to provide insight into how each mutation affects the protein's conformational flexibility and overall structural integrity, similar to the approach outlined in Chapter 1.4.2. By monitoring key metrics such as root mean square deviation (RMSD), root mean square fluctuation (RMSF), solvent accessible surface area (SASA) and the radius of gyration (Rg), the MD simulations will assess how well the chromophore remains stable within the β-barrel under physiological conditions. Additionally, quantum mechanics (QM) calculations will be performed to investigate the spectral properties of these variants, including geometry optimisation, dipole moments, HOMO-LUMO gaps, predicted absorption wavelengths and Raman spectroscopic peaks. These calculations will use the same method and basis set (B3LYP/6-31G+(d,p)) as Chapter 5, and is discussed in detail in Chapter 1.4.3 and Chapter 2.7.

The main aim of this chapter is to evaluate the effects of incorporating ncAAs into the chromophore of mCherry, focusing on changes in their spectral properties such as absorbance and Raman activity. Computational objectives involve assessing the structural properties of these variants using MD simulations, while QM simulations will explore the optical properties of the chromophores.

6.2 Results and discussion

6.2.1 Spectral properties of incorporating ncAAs into the chromophore of mCherry

As revealed in the previous chapter, the use of through-space enhancement using ncAAs did not lead to enhancement of the Raman bonds in the biologically silent window. As has been shown before for chemical dye-based systems, ideally the Raman active bonds should be directly covalently coupled to the electronically excited system (Wei et al. 2017).

This is difficult for mRhubarb720 as the BV chromophore is a cofactor naturally present in cells. One solution would be chemically modifying the BV cofactor, but this also has limitations with the major one being the inability of incorporating BV into the cell.

The next logical step in this research is to focus on incorporating ncAAs directly into the chromophore of FPs in which chromophore synthesis is inherent to the protein scaffold and new chemistry can be incorporated directly into the chromophore. Here, the target will be the far-red FP, the mCherry. This protein is a β-barrel FP that contains the chromophore within its core (Figure 6.1.A). By integrating ncAAs into the electronically excited chromophore, it is hypothesised to directly alter the vibrational properties of the chromophore, potentially leading to enhanced Raman scattering within the biologically silent window. Three targets were selected to replace the tyrosine within the mCherry WT chromophore (CRO): pCNPhe, pCCPhe and 3-NO_2-Tyr (Figure 6.1.B-E). The first two ncAAs result in the hydroxyl group being replaced by the new chemical moiety whereas CRO-3-NO_2Tyr adds the nitrosyl group at the meta (or 3-) position of the aromatic ring so persevering the original hydroxyl group. The pCNPhe (2220-2260 cm^{-1}) and pCCPhe (2100-2150 cm^{-1}) mutations were constructed in order to observe sharp spectral peaks in the biologically silent region, whereas the 3-NO_2Tyr (1300-1400 cm^{-1} / 1510-1550 cm^{-1}) was included in attempt to see further enhancement within the fingerprint region.

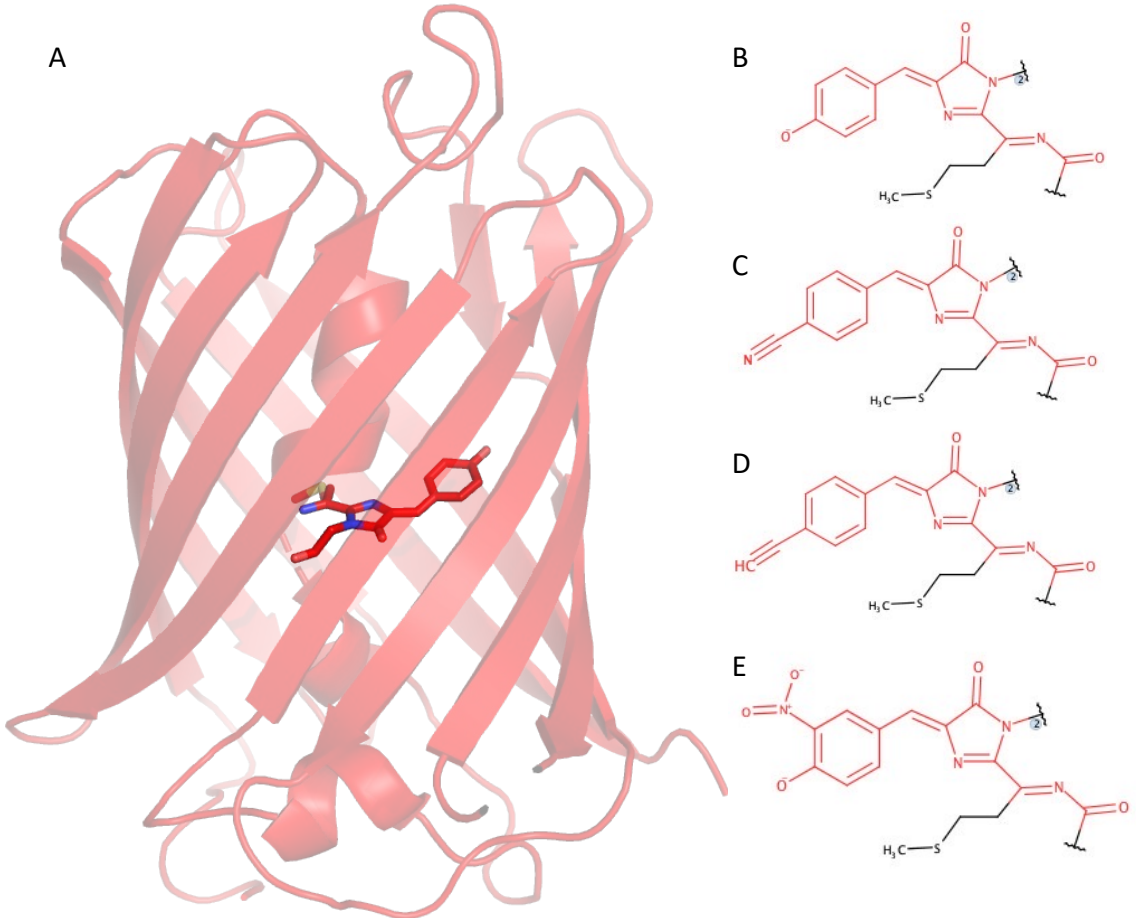


Figure 6.1: Structure of mCherry with the chromophore structures of WT and mutants. (A) mCherryWT crystal structure (PDB: 2H5Q), (B) WT chromophore and with Tyr67, (C) CRO-CN, (D) CRO-CC or (E) CRO-NO_2; CRO-NO_2-R, with the red highlight representing the delocalised system within the chromophore. The 3-NO_2-Tyr may also sample an isomer of the opposite meta-position, named CRO-NO_2-L.

The structure displayed here shows the β-barrel fold typical of fluorescent proteins, with the chromophore centred within the protein. The chemical structures, in Figure 6.1.B-E, depict variants of the chromophore with different substituents. The red-highlighted region illustrates the extended conjugated system responsible for the chromophores light absorption and fluorescence properties. However, these chemical modifications may not be fully conjugated with the core chromophore, for true conjugation, the groups would need resonance forms capable of electron sharing with the chromophores π-system; without this, their contributions to the conjugated system are limited. In mCherry WT and the mutants, the central ring system features a benzylidene-imidazolinone core (2-formyl-4-hydroxybenzylidene-3-methylimidazolinone; FHBMI) typical of red-fluorescent β-barrel FPs

(Addison et al. 2013). The CRO-CN and CRO-CC mutants incorporate a nitrile and ethynyl group, respectively, at the para-position of the aromatic ring. Whereas the CRO-NO_2 incorporates a nitro group at the meta-position of the aromatic ring, known for its strong electron withdrawing properties (along with nitriles), which could change the chromophores electronic structure and behaviour (Nishiwaki 2020).

These structural alterations in the chromophore are expected to influence the chromophores optical properties, including absorption and emission. However, despite the predicted shifts in the delocalised systems, experimental results revealed unexpected outcomes in terms of absorbance (Figure 6.2). Firstly, the CRO-NO_2 mutant failed to express altogether, likely due to interference with protein maturation. The addition of the bulky NO_2 group, replacing a hydrogen in the phenolic ring, may have introduced steric clashes within the central chromophore region of the β-barrel, disrupting the formation of the protein structure. Both the CRO-CN and CRO-CC mutants exhibited a major blue shift in absorbance, demonstrating that these modifications disrupted the chromophores ability to absorb light in this region. Furthermore, the CRO-CN and CRO-CC mutants also did not exhibit fluorescent properties when excited at their respective absorbance maximums, which further highlights that these modifications significantly altered the structure and conjugation of the chromophore.

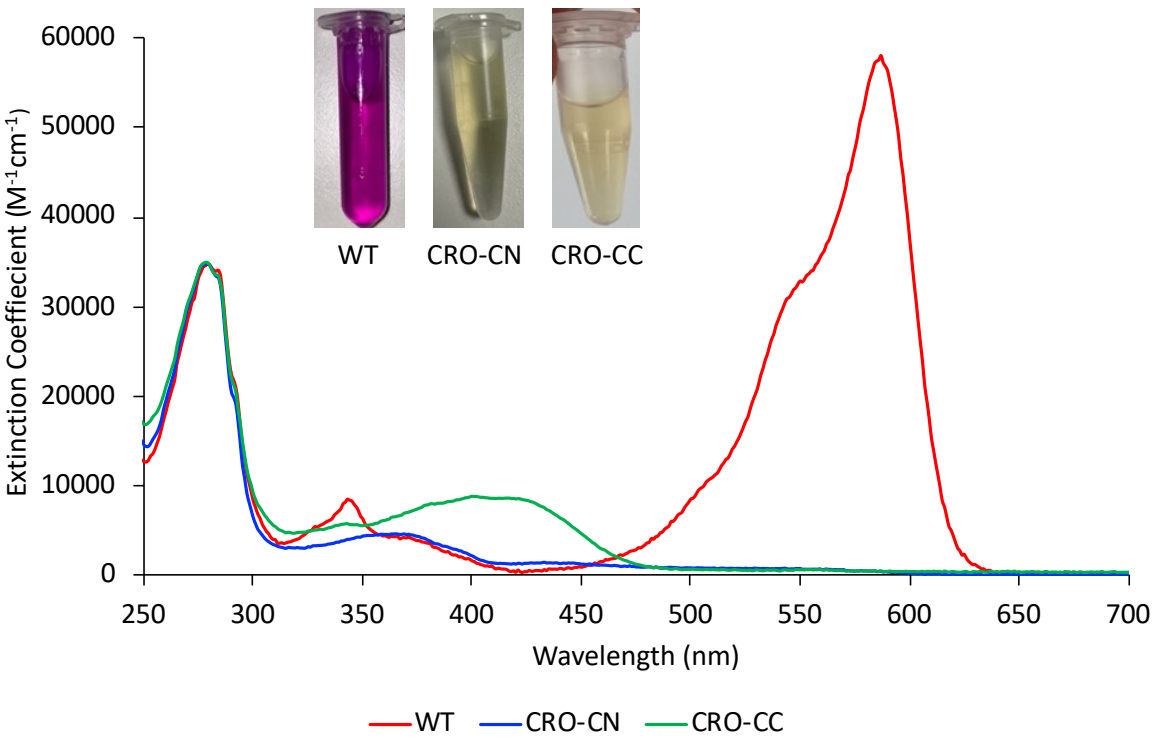


Figure 6.2: Absorbance spectra and colour of mCherry variants. WT (red), CRO-CN (blue) and CRO-CC (green) in respect to the proteins extinction coefficient. Images show the colour of each purified variant.

The mCherry WT spectrum shows a strong absorption peak at 587 nm, with an extinction coefficient of 58.1 $mM^{-1}cm^{-1}$, which correlates with the cherry-like colour observed in the first tube. The CRO-CN has the lowest and most blue-shifted absorbance with a broad peak at 367 nm (5.8 $mM^{-1}cm^{-1}$), whereas CRO-CC shows a broader peak at 405 nm (11.0 $mM^{-1}cm^{-1}$). Both the nitrile and ethynyl substitutions has been observed to disrupt the chromophores ability to absorb in the red region of the spectrum, as reflected in their respective protein colour. The inclusion of these chemical moieties may also interfere with chromophore maturation, as the hydroxyl group at the 4-position in mCherry WT participates in a hydrogen bond network (with S146) stabilising the intermediates during maturation (Strongin et al. 2007; Laurent et al. 2012).

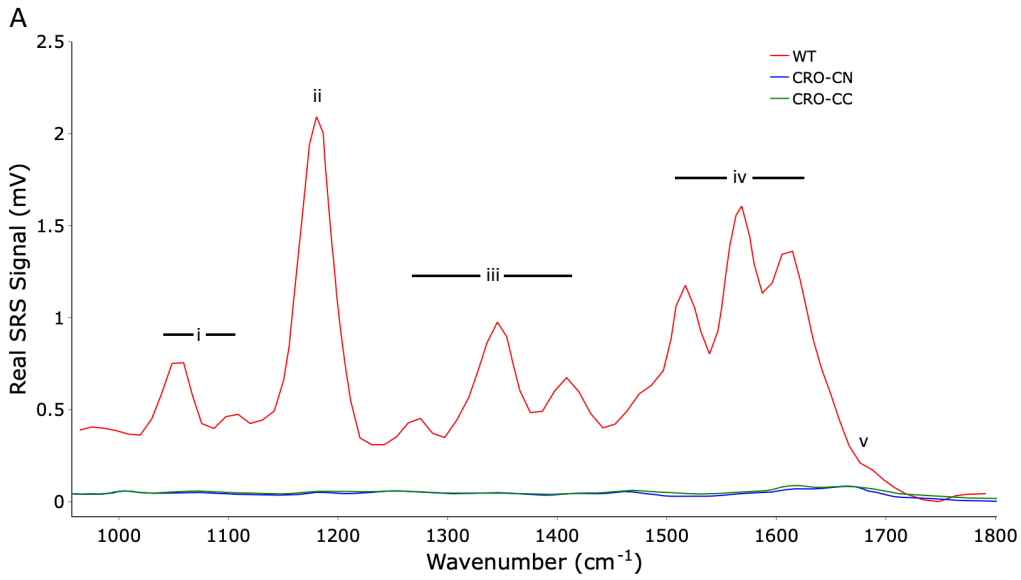
The observed spectral changes in the mutants, are not ideal for our current in-house Raman microscope setup at Cardiff University, which is optimised for IR absorbing probes. Given these challenges, all expressed mCherry variants (including mCherry WT; excluding CRO-NO_2) were sent to the University of Konstanz in Germany, where a more blue-shifted laser setup was available. Thanks to Dr Andreas Zumbusch, Dr Andrea Pruccoli and Dr

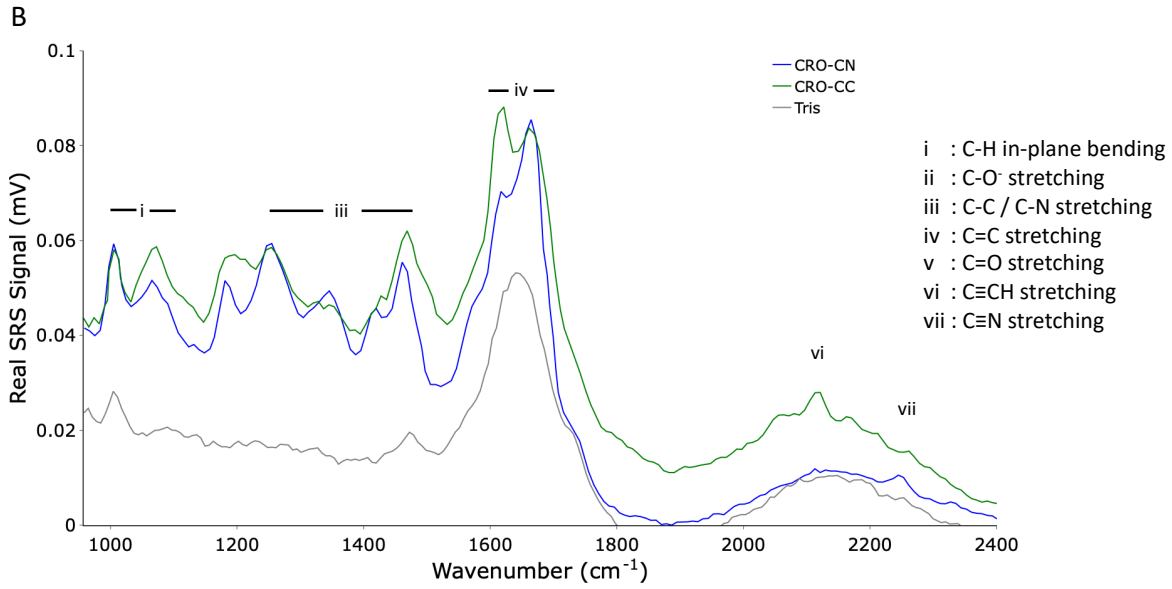
Franzika Rabold, the Raman profiles were measured on each variant. The pCNPhe and pCCPhe residues in solution were measured in Cardiff University by Dr David Regan as before.

The following epr-SRS measurements were conducted using a Pump excitation of 640 nm, which is much closer to the excitation maximum of mCherry WT (587 nm). The fingerprint region was initially examined to compare the mCherry WT chromophore with its chemically modified variants (Figure 6.3.A). The mCherry WT has very clear enhancement throughout this spectral range with various peaks from 1000-1700 cm^{-1}. The following Raman peak characteristics are sourced from '*The Handbook of Infrared and Raman Characteristic Frequencies of Organic Molecules*' (Lin-Vien et al. 1991).

The following peaks are attributed to Raman enhancement of the bonds within the mCherry chromophores. The first peak range (Figure 6.3.A-i) is attributed to the C-H in-plane bending, most likely from the aromatic phenolic ring in the chromophore that contain four C-H bonds. The conjugated bond connecting the aromatic rings also contains an in-plane C-H bond which will contribute to this Raman peak. The peak at 1180 cm^{-1}, with a maximum signal intensity of 2.1 mV, is associated with C-O^- stretching (ii) at the para-position of the phenolic ring in the chromophore. The negative charge indicates a larger electron density in this region, leading to a strong pre-resonant enhancement. The following range, 1250-1450 cm^{-1} (iii), is linked with C-C and C-N stretching. This vibration originates from the C-C and C-N bonds within the chromophores conjugated structure, which contribute strongly due to the nitrogens polar nature and therefore high polarisability. The intense peaks ranging between 1500-1650 cm^{-1} (iv) corresponds to C=C stretching, predominantly within the aromatic ring of the chromophore. This vibrational mode is characteristic of conjugated double bonds in aromatic systems and is further enhanced by the resonance within the ring structure, leading to a strong Raman signal. Finally, the peak at 1686 cm^{-1} (v), attributed to C=O stretching, is associated with the carbonyl groups within the chromophore. This peak is very weak, appearing with minimal intensity, and is barely visible in the German setup. In comparison, in-house measurements of mCherry WT (Chapter 4.2.1; Figure 4.2) also showed a low-intensity peak at 1619 cm^{-1} (~0.1 mV) alongside a peak at 1678 cm^{-1} (also at ~0.1 mV). However, in the German setup, using a 640 nm pump pulse, the 1619 cm^{-1} peak shifts slightly to 1614 cm^{-1} with a much higher intensity of 1.4 mV due to pre-resonance enhancement from the pump wavelength being closer to the chromophores absorption

maximum. This results in approximately a 14-fold increase in the Raman signal for this mode at 1614 cm^{-1}. The 1686 cm^{-1} peak, however, remains barely detectable in the German setup, appearing at an intensity of only 0.2 mV. This very faint peak likely reflects the selective enhancement of modes more resonant with the chromophores electronic structure under 640 nm excitation, leaving the 1686 cm^{-1} mode largely unenhanced with insignificant signal.


A
WT
CRO-CN
CRO-CC
Real SRS Signal (mV)
Wavenumber (cm-1)
i
ii
iii
iv
v


B
CRO-CN
CRO-CC
Tris
Real SRS Signal (mV)
Wavenumber (cm-1)
i
iii
iv
vi
vii
i : C-H in-plane bending
ii : C-O- stretching
iii : C-C / C-N stretching
iv : C=C stretching
v : C=O stretching
vi : C≡CH stretching
vii : C≡N stretching

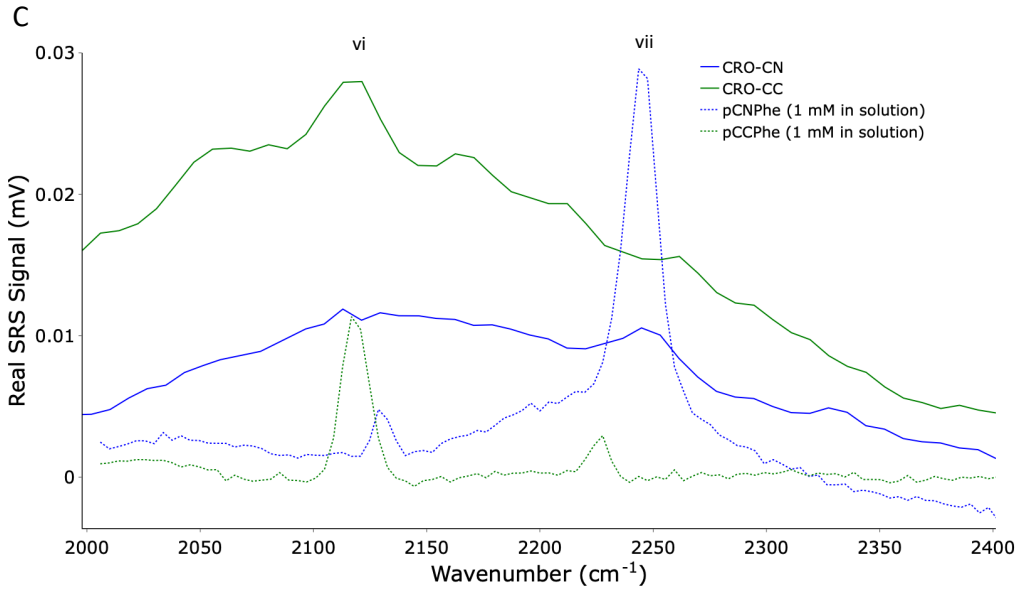

C
vi
vii
CRO-CN
CRO-CC
pCNPhe (1 mM in solution)
pCCPhe (1 mM in solution)
Real SRS Signal (mV)
Wavenumber (cm-1)

Figure 6.3: SRS measurements of mCherry variants. (A) Fingerprint spectra including WT (red) in comparison to CRO-CN (blue) and CRO-CC (green), where (B) displays the both the fingerprint and biologically silent window focussing on CRO-CN and CRO-CC mutants, against the Tris buffer solution. (C) Biological silent window scan of both CRO-CN and CRO-CC, compared to the pCNPhe in solution (blue dotted line) and pCCPhe in solution (green dotted line). All the proteins and ncAAs were measured at 1 mM. The proteins were in 50 mM Tris, and the ncAAs were in 50 mM Tris and 50 mM NaOH.

The epr-SRS signal from the CRO-CN and CRO-CC mutants shows significantly reduced signal throughout the fingerprint region, compared to mCherry WT. This lack of signal will likely be attributed to the blue-shifted absorbances observed, in Figure 6.1. This means that the chromophores excitation profiles are now too far removed from the optimal range of the Raman microscope setup, which is tuned closer to the excitation maxima of mCherry WT. However, in Figure 6.3.B, some Raman peaks within the fingerprint region can be revealed, against the Tris control SRS curve. Firstly, the C-O^- stretching Raman peak (ii) is absent in the mutants, emphasising the effect of replacement of the hydroxyl with the nitrile/ethynyl. Although the other peaks in the fingerprint region are labelled, these bands are most likely combined with the background from the whole protein and not fully enhanced from the chromophore itself. Nevertheless, this is conformation that the mCherry WT chromophore is pre-resonantly enhanced under these conditions (pump λ=640 nm). Finally, despite the low signal, the nitrile and ethynyl peaks are clearly discernible, verifying the successful incorporation of these ncAAs into the chromophore. These peaks are more prominent as they are situated within the biologically-silent region, removed from the background signals observed in the fingerprint region.

In Figure 6.3.C, a direct comparison between the mutants and the ncAAs in solution is presented, showing overlapping vibrational peaks. These signals are consistent with the characteristic vibrational frequencies of the nitrile at 2250 cm^{-1} and the ethynyl group at 2125 cm^{-1}, confirming the presence of these mutations within the chromophore structure. Despite this, the SRS signal for the chromophore-bound forms of CRO-CN and CRO-CC are relatively weak. As mentioned, this reduced intensity can be attributed to the significant blue shift in absorption wavelength, which pushes the chromophores absorption maxima below 450 nm. This shift diminishes the pre-resonant enhancement that would typically

amplify the Raman signals, leading to lower overall intensities. As a result, the signal strength for 1 mM solutions of CRO-CN and CRO-CC is comparable to that of a 10 µM solution of mCherry WT, highlighting the impact of the blue-shifted chromophore on the Raman enhancement.

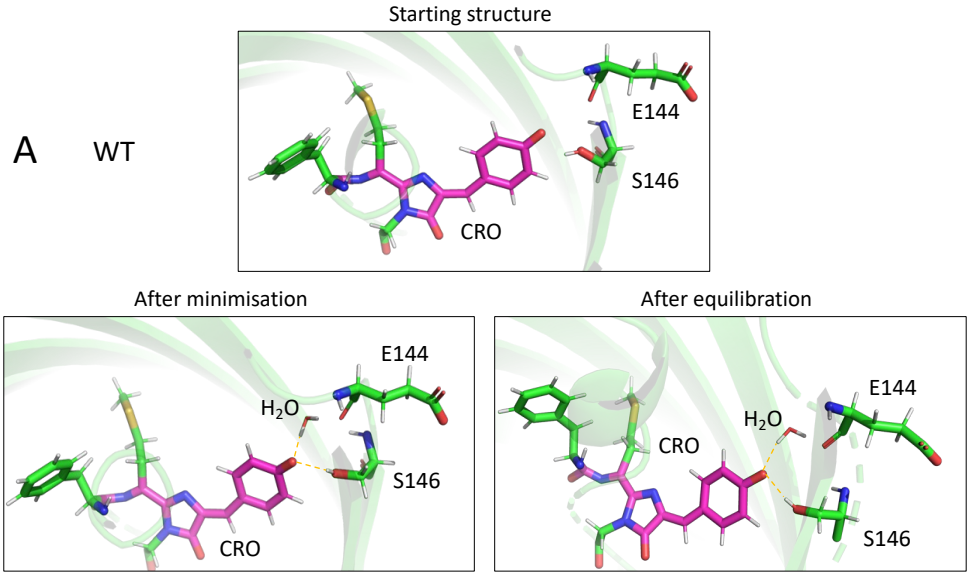
6.2.2 Molecular dynamics analysis of expressed mCherry variants

Building on the experimental data obtained from absorbance and SRS measurements, MD simulations were conducted to gain insight into the structural and dynamic effects of ncAA modifications on the mCherry chromophore. This section focuses on the expressed variants, mCherry WT, CRO-CN, and CRO-CC, while CRO-NO_2 is omitted as it did not express. Since it remains uncertain whether incorporation is feasible with the 3-NO_2 modification, this variant was considered redundant for the purposes of this analysis. Just as the previous chapter, three simulations of each variant at 500 ns will allow us to explore the potential structural arrangements and dynamic behaviours of the protein and chromophore over time, complementing the experimental data by understanding how these mutations influence overall stability.

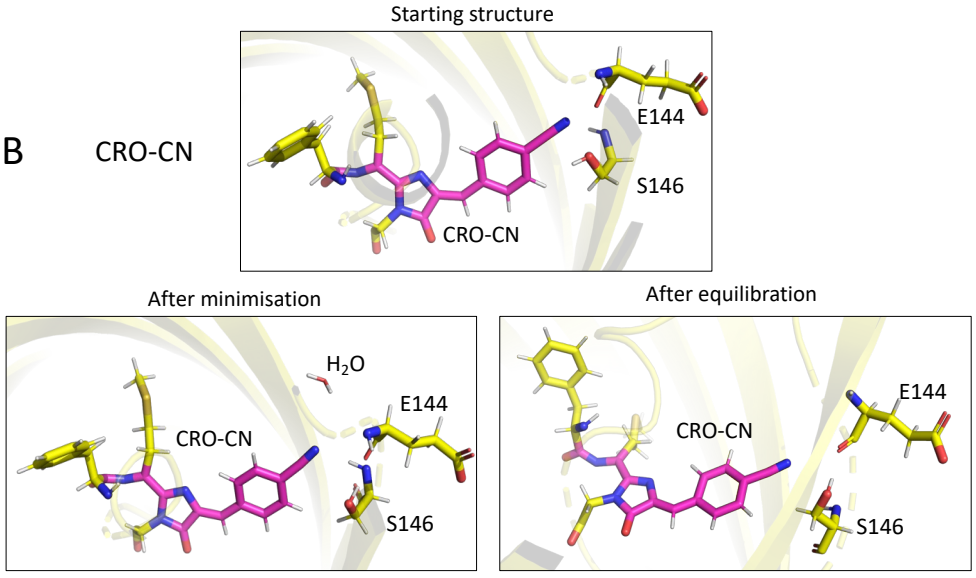
The models for the MD simulations were built using the crystal structure of mCherry WT (PDB: 2H5Q) as the starting template. The mCherry WT chromophore and F65 were parameterised as a single residue to accurately represent the conjugated system, which extends through the backbone of F65. Treating the chromophore and F65 as a unified system was essential to represent this conjugation in the simulations. For the CRO-CN and CRO-CC variants, the hydroxyl group was replaced with nitrile and ethynyl bonds, respectively, using the PyMOL builder function. Adding these modifications to the chromophore may have structural consequences, potentially impacting the local or overall configuration of the protein. To evaluate these effects, the structure of each variant was monitored across three stages: the initial starting structure, after energy minimisation, and after equilibration (Figure 6.4). In the starting structure, the chromophore modifications were introduced without any observed steric clashes, indicating a stable initial configuration. During energy minimisation, the system was refined to alleviate any unfavourable interactions, promoting local structural adjustments. Equilibration then introduced

temperature, pressure, and density to the simulation, allowing the system to reach a stable state under physiological-like conditions.

A WT



B CRO-CN



C CRO-CC

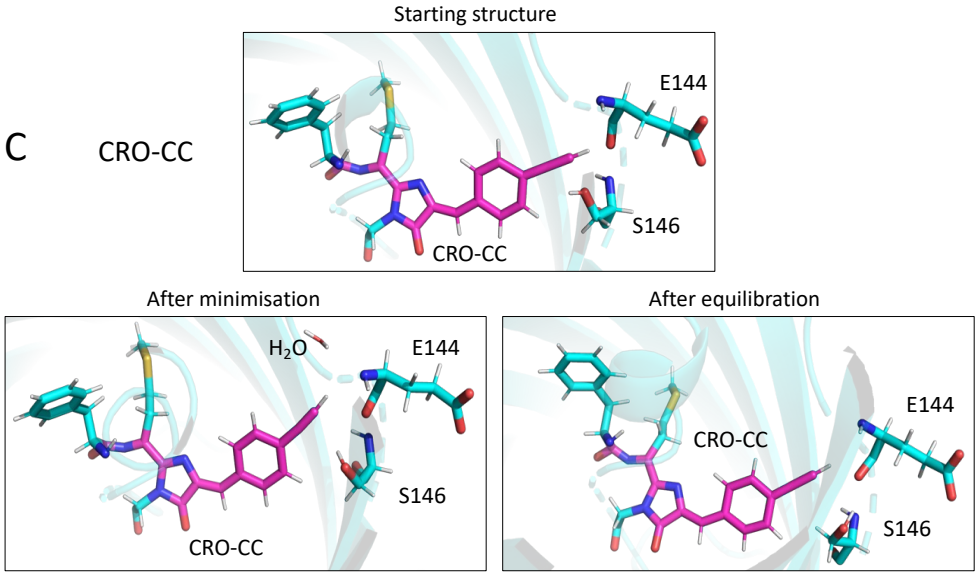


Figure 6.4: Structural evolution of chromophores and selected residues throughout initial MD simulation phases. The starting structures, after energy minimisation and after equilibration of each variant. (A) WT, (B) CRO-CN, (C) CRO-CC structures with labelled interacting residues (E144 and S169) with also a H_2O molecule labelled where applicable. The yellow lines represent the hydrogen bonding. The magenta for each chromophore highlights the conjugated system with unique colours for each variant (WT: green, CRO-CN: yellow, CRO-CC: cyan).

The interacting residues, E144 and S146, were selected due to their proximity to the chromophores hydroxyl group, which is known to form a hydrogen bond with S146. In turn, E144, hydrogen bonds with a nearby water molecule that also interacts with this hydroxyl (Mukherjee et al. 2023). In the mCherry WT structure, these interacting residues remain relatively fixed throughout the simulation stages. Following both minimisation and equilibration, a water molecule remains in the vicinity, forming hydrogen bonds with both the chromophores hydroxyl and E144, which supports the stability and consistency of these interactions in the mCherry WT variant. In contrast, the CRO-CN and CRO-CC variants exhibit significant shifts in these interacting residues, particularly in the orientation of S146. The R-group of S146 moves from within the β-barrel, away from the chromophore's immediate environment. Additionally, the closest water molecule after minimisation is located outside the β-barrel, with no water molecules observed near the nitrile or ethynyl groups. This loss of proximal water and displacement of S146 suggest a disruption in the hydrogen-bonding network, potentially impacting the stability and optical properties of the FPs.

Given that the concepts of RMSD, RMSF, SASA and Rg were thoroughly discussed in the context of mRhubarb720 in Chapter 5, here the focus will be directly on the results of these analyses for mCherry WT compared to the mutant chromophores. The goal is to identify any major changes in stability, flexibility, or compactness caused by the chemical modifications, with an emphasis on how these alterations influence the inherent interactions and structural dynamics of the chromophore throughout the simulations. Firstly, the protein backbone RMSD was investigated to reveal any clear differences in the structural stability of each system (Figure 6.5). For the mCherry WT, the RMSD stabilises at 0.138 ± 0.028 nm on average across all three simulations, indicating that the protein backbone maintains a stable structure throughout each simulation.

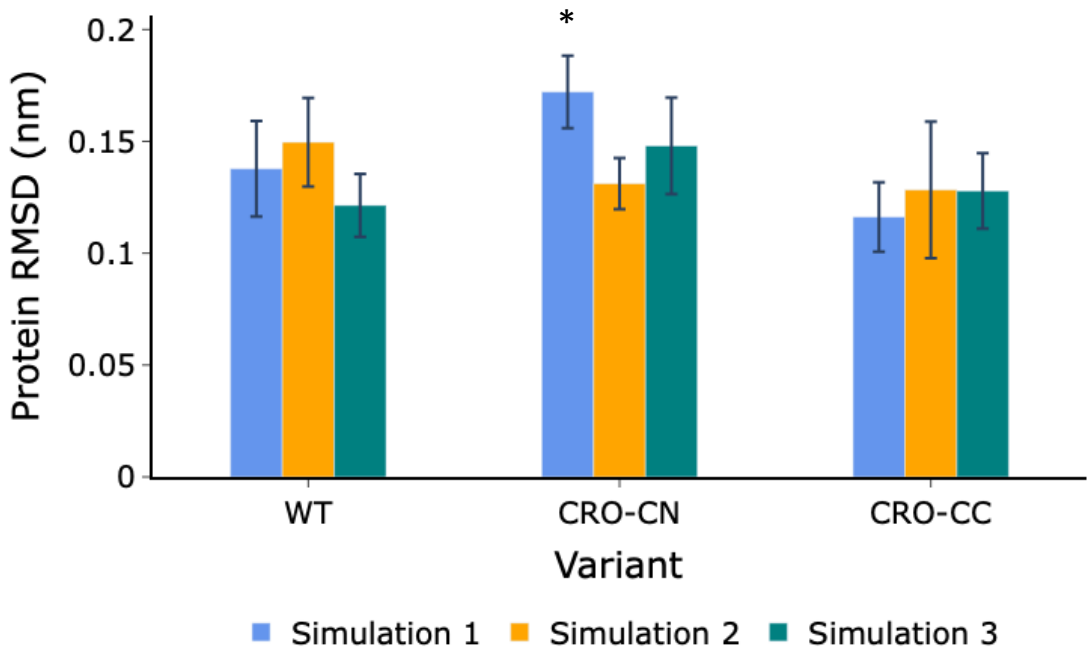


Figure 6.5: Average RMSD of WT, CRO-CN and CRO-CC calculated over 50-500 ns time frame. All three simulations are represented by their median values and the error bars represent the standard deviation of the averaged RMSD values for each simulation. Kruskal-Wallis test was performed on each simulations to compare against each WT simulation, with * denoting significance ($p < 0.0001$). All individual RMSD graphs can be seen in Figures S52-54.

The results show that CRO-CC has comparable RMSD to mCherry WT with relatively similar values (0.132 ± 0.027 nm), suggesting comparable stability in these structures over the simulation time frame. In contrast, the CRO-CN variant exhibits a slightly higher RMSD (0.147 ± 0.028), with Simulation 1 significantly larger deviation from mCherry WT. The significant difference observed for CRO-CN suggests that the nitrile modification may introduce structural changes that reduce stability compared to mCherry WT. As mentioned in Chapter 5.2.3.3, the Kruskal-Wallis test was applied on each mutant simulation, comparing the median values against all three mCherry WT simulations. If there was a non-significant p value associated with the test, then that particular simulation will be treated as such. However, this statistical application to large MD datasets is limited due to the high variability in protein flexibility and increased number of data points, which may complicate the interpretation of the results.

The delocalised system of the chromophore, comprising the key chemical bonds and atoms responsible for electronic excitation, was also analysed in terms of the RMSD (Figure 6.6.A). This analysis is critical as these elements directly influence the chromophores optical behaviour and its ability to undergo electronic transitions. Notably, the data reveals distinct differences between the mCherry WT and the mutants, particularly regarding their ability to revert back to the native conformation after significant deviations.

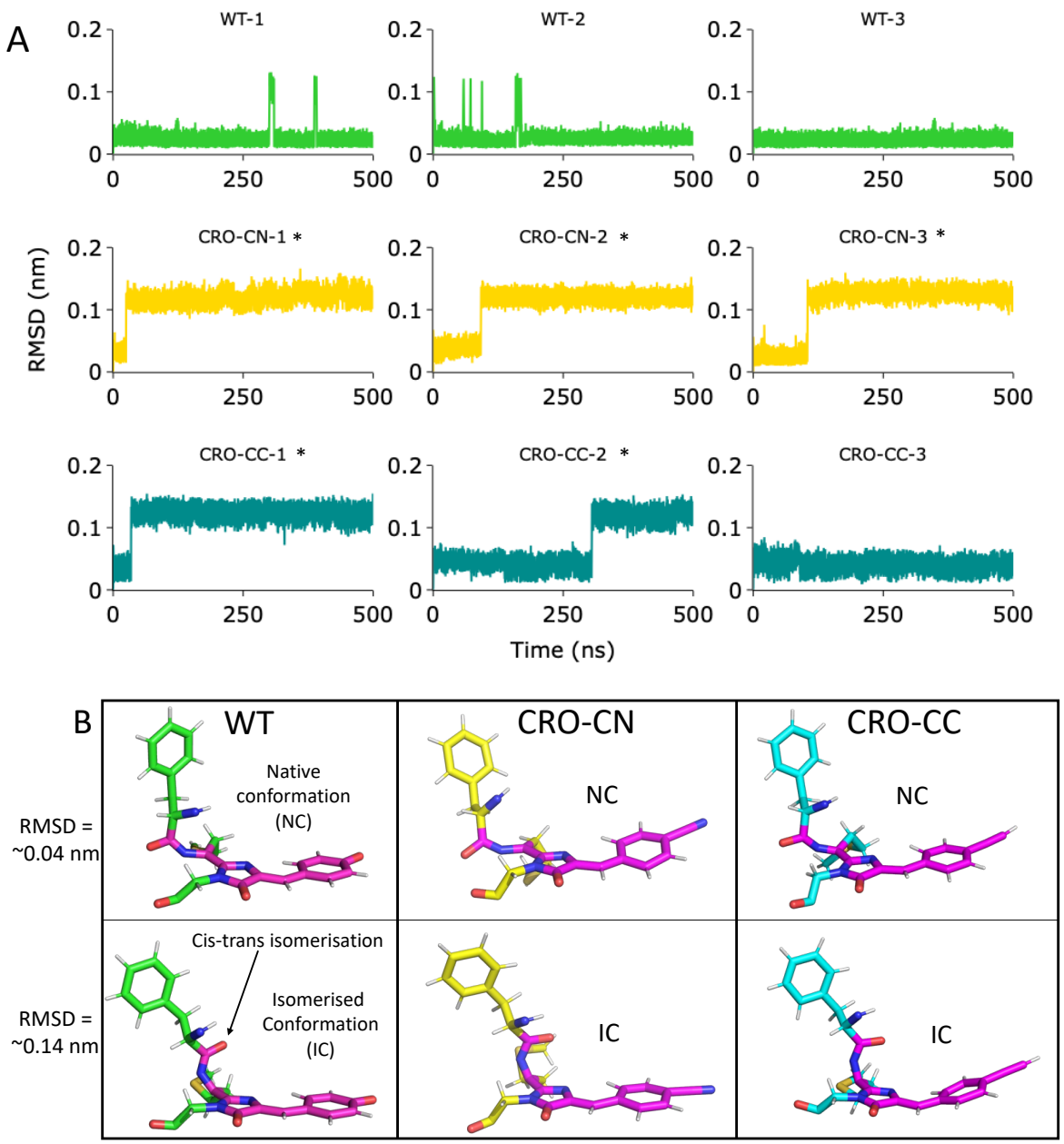


Figure 6.6: RMSD plots of the delocalised system in the mCherry chromophores and their respective conformations. (A) Individual RMSD plots of each simulation across all mCherry variants. Kruskal-Wallis test was performed on each simulations to compare against each WT simulation, with * denoting significance ($p < 0.0001$). (B) Structures of the native conformations (NC) and isomerised conformations (IC) observed in the simulations, the magenta atoms represent the delocalised system of the chromophore

The simulations present two alternative conformations for mCherry WT, CRO-CN and CRO-CC, where the initial chromophore presented as a native conformation and an cis-trans isomerised conformation occurring at the backbone of F65, also observed (Figure 6.6.B). The

mCherry WT simulations remain stable at a baseline RMSD of ~0.04 nm, with brief fluctuations up to 0.14 nm as the chromophore briefly transitions into the isomerised conformation, where the oxygen that forms the backbone of F65 undergoes a cis-trans isomerisation. The mCherry WT chromophore has the ability to revert back to the native conformation within at least 5 ns, demonstrating its conformational flexibility.

In all three 500 ns simulations, the mCherry WT chromophore is in the isomerised state for less than 1% of the time. This stable conformational behaviour is consistent with the expectation that the mCherry chromophore occupies a dominant structural state in the ground state. However, fluorescence lifetime studies have shown that many fluorescent proteins, including mCherry, exhibit heterogeneous excited-state behaviour (Walther et al. 2011). This paper demonstrated that the majority of FPs show that a heterogeneous population of excited states can be resolved, when sufficient photon counts are detected. For mCherry, they reported an approximate 80:20 population split between two distinct emissive states, as inferred from biexponential fluorescence decay fitting. It was reported also reported for mCherry that a strong biexponential decay was attributed to the excited states with distinct transition dipole moment orientations, differing by as much as 20° within the β-barrel (Masters et al. 2013). While this research was not the objective of this thesis, this could reflect the presence of multiple conformers, where the MD simulations show that the WT chromophore remains in the native configuration for >99% of the time, with only brief fluctuations and no persistent alternative conformations. Therefore, the biexponential fluorescence likely reflects excited-state dynamics or restricted relaxation pathways rather than two well-populated ground-state conformations. However, these photophysical phenomena will not be captured at the classical MD level, where electronic excitation and polarisation effects are absent.

In contrast, the CRO-CN mutant exhibits a notable shift to the isomerised conformation in all three simulations. The RMSD increases from the baseline at 0.04 nm, to a higher value of 0.14 nm, where it stabilises in the new conformation without reverting. Similarly for Simulation 1 and 2 of CRO-CC, once the conformational change takes place, there is also an inability to revert back to the native state. Both CRO-CN and CRO-CC, introduce larger chemical groups at the para-position, replacing the deprotonated oxygen typically present in this location. This substitution likely forces the chromophore into a shifted position, preventing it from reverting to its native conformation.

The structural differences between both conformations indicate significant distortions within the chromophore. In the mCherry WT, the isomerised state is reverted back, enabling the chromophore back to its functional state. However, the mutants inability to revert back to the native conformation points to a restrictive environment around the chromophore. When the chromophore is unable to regain its stable conformation, the integrity of its electronic transitions is compromised, while potentially reducing the dipole moment (explored further in Section 6.2.3.1). The isomerised, non-planar conformation may also disrupt the chromophores resonance by shifting it out of alignment with the surrounding structure. Additionally, the removal of the hydroxyl group, replaced by nitrile or ethynyl groups, further limits resonance stabilisation around the 4-position.

In addition to analysing the conformational changes within the chromophore, the overall flexibility of the protein was assessed through RMSF calculations. This approach will allow for a deeper understanding of how the mutations, along with the distinct forms of the chromophore, will impact the dynamic behaviour of the entire protein (Figure 6.7.A). The RMSF profiles indicate that the overall flexibility of the protein scaffold is largely unaffected by the mutations, with minimal deviations observed between mCherry WT, CRO-CN, and CRO-CC variants. In Figure 6.7.B, a schematic of the protein is presented, outlining the positions of each beta-sheet and alpha-helix and their respective residue numbering, in order to identify the location of each data point on the RMSF plot.

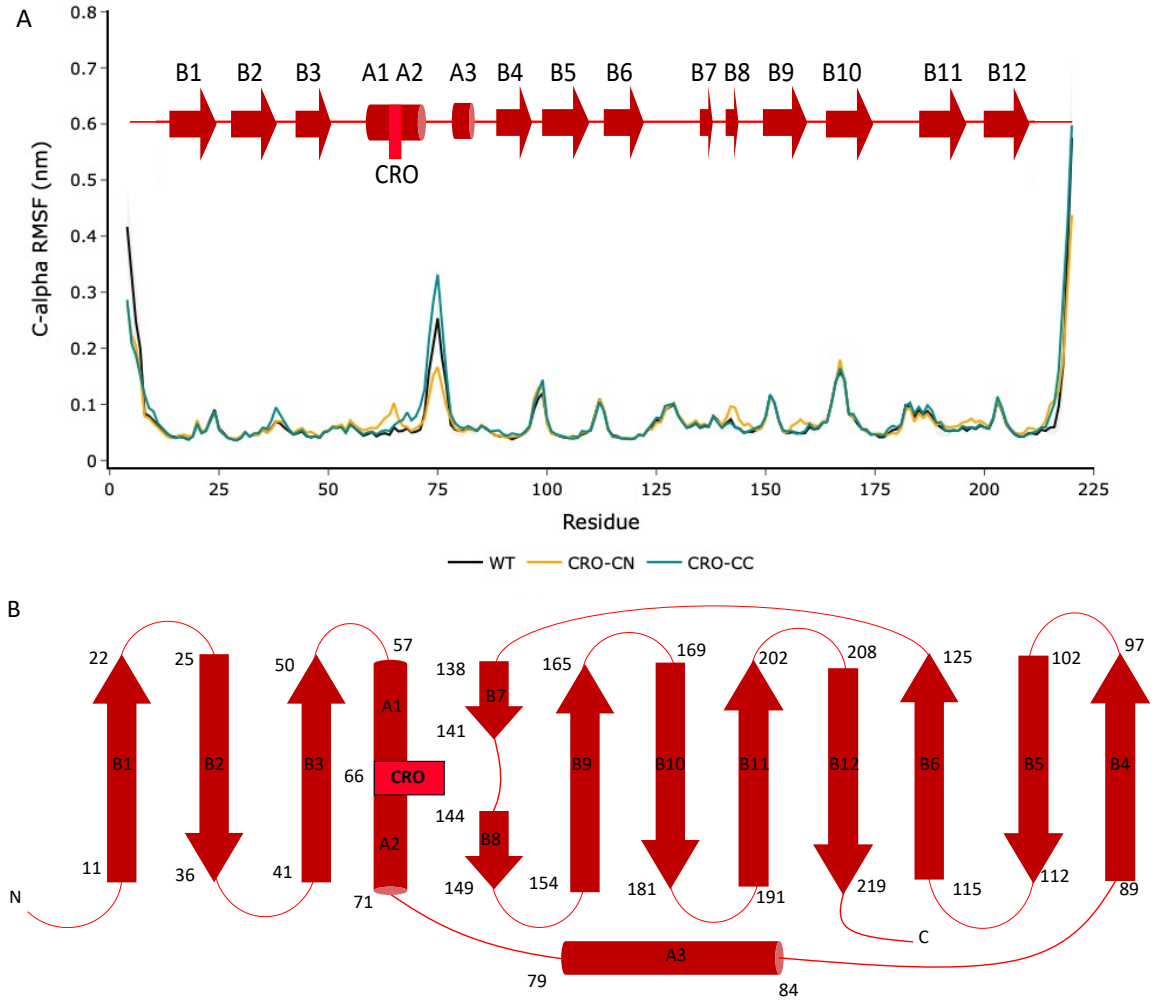


Figure 6.7: Protein scaffold C-α RMSF of variants with structural schematic of the mCherry scaffold. (A) RMSF of WT (black), CRO-CN (yellow) and CRO-CC (cyan) with a linear schematic of the secondary structure of mCherry is depicted at the top with arrows representing beta-sheets and cylinders representing alpha-helices.(B) Representation of secondary structure of mCherry showing the arrangement of beta-sheets (arrows; B1 to B12) and alpha-helices (cylinders; A1 to A3) and the chromophore depicted in bright red. All individual RMSF graphs can be seen in Figures S55-57.

As expected, the terminal regions exhibit the highest flexibility, consistent across all variants. As this flexibility is generally inherent to these termini, the increased deviation observed in the mutants can be expected and is likely to have little impact the overall structural integrity of the protein. Residue 75, located between helices A2 and A3 at the top of the protein, demonstrates the greatest flexibility outside of the terminal regions, which is a structural feature rather than an effect of mutation. In CRO-CN, at residue 64, adjacent to

the CRO, shows a slight increase in flexibility compared to the mCherry WT and CRO-CC variants. This suggests a localised effect of the mutation in this region without propagating significant changes across the protein structure. These observations suggest that the introduced mutations have a negligible impact on the protein's structural dynamics, maintaining the stability and rigidity of the β-barrel scaffold and the surrounding regions.

An all-atom RMSF analysis for the chromophore itself was also conducted, focussing on the delocalised system (Figure 6.8). The key observation from this analysis is that the highest flexibility observed, was the oxygen (atom 2) undergoing the cis-trans isomerisation, as previous described in the chromophore RMSD analysis.

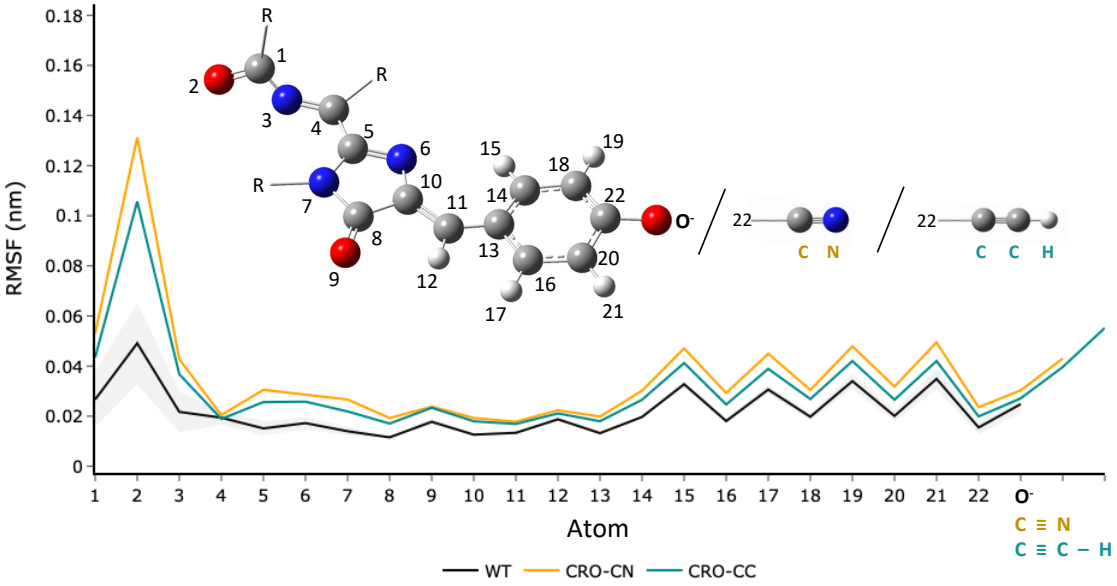


Figure 6.8: All-atom RMSF analysis of the delocalised system of the chromophores. WT (black), CRO-CN (yellow) and CRO-CC (cyan) are plotted with the standard deviation of WT values (grey) presented. The structural diagram of the chromophore highlights the atom numbering corresponding to the RMSF plot, with the hydroxyl, nitrile and ethynyl labelled by their atomic symbols.

Both CRO-CN and CRO-CC exhibit consistently higher RMSF values across the chromophore compared to mCherry WT, indicating an overall increase in flexibility. Notably, the modifications at the 4-position of the phenolic ring, lead to a progressive increase in flexibility as additional atoms are introduced. This increased flexibility causes the chromophore to adopt a more non-planar conformation, which may influence its electronic and optical properties.

The analysis of the Rg and SASA are presented together as they follow similar trends across the variants (Figure 6.9). For the protein analysis, both Rg and SASA remain relatively unchanged in the mutants compared to mCherry WT, reflecting the stability of the protein scaffold irrespective of the chromophore modifications. In contrast, the chromophore shows notable differences in both Rg and SASA, driven by the structural modifications and isomerisation observed in the mutants.

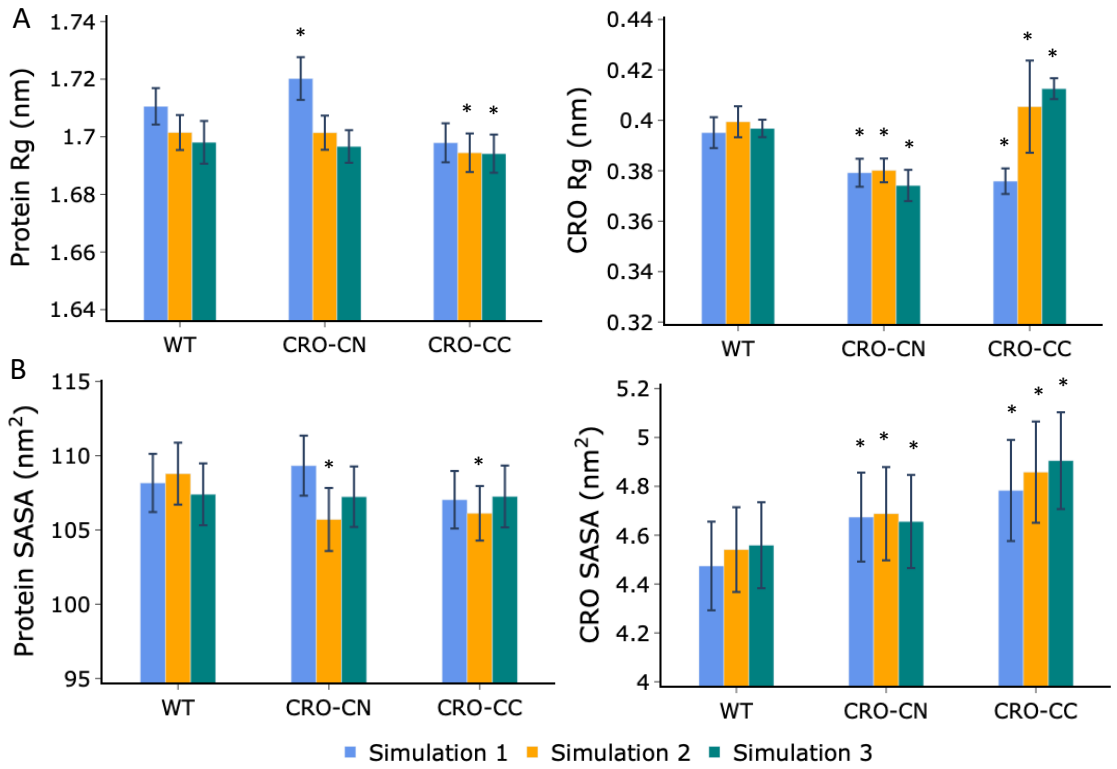


Figure 6.9: Radius of gyration (Rg) and solvent-accessible surface area (SASA) of all simulations in the variants calculated over 50-500 ns time frame. (A) Rg of the protein (left) and chromophore (CRO; right) and (B) SASA of the protein (left) and chromophore (CRO; right). Kruskal-Wallis test was performed on each simulations to compare against each WT simulation, with * denoting significance ($p < 0.0001$). All individual graphs can be seen in Figures S61-72.

The differences in Rg among the mutant chromophores are explained by their conformational states. Simulation 3 of CRO-CC remains in the native conformation throughout, with the additional atoms from the modification contributing to the observed higher Rg. In Simulation 2 of CRO-CC, the chromophore isomerised at approximately 300 ns, meaning over half the simulation remained in the native conformation. This results in a larger standard deviation (as indicated by the error bars) due to the later transition between

conformational states. In contrast, the other four mutant simulations underwent isomerisation before 120 ns, meaning the majority of their trajectories were spent in a more compact form, explaining the lower Rg observed for these simulations compared to mCherry WT. As expected, the increase in SASA for the chromophore mutants is solely due to the additional atoms introduced compared to the mCherry WT, with CRO-CC showing the largest SASA, followed by CRO-CN, and then mCherry WT.

To investigate the interactions between the chromophore and its surrounding protein and solvent environment, the analysis focused on the number of hydrogen bonds formed by specific functional groups in each chromophore variant (Figure 6.10). For mCherry WT, the deprotonated oxygen at the para-position of the phenolic ring was selected, as it is reported in the literature to form two hydrogen bonds in mCherry WT, with S146 and a water molecule (Shu et al. 2006; Mukherjee et al. 2023). For CRO-CN, the nitrile group was selected, and for CRO-CC, the ethynyl group was examined. Although other hydrogen-bonding interactions involving the chromophore, such as E215 and the imidazolinone ring contributing to the overall stability, this investigation focuses specifically on the modifications at the phenolic ring. This analysis serves two purposes: first, to validate whether the hydrogen bonding interactions observed in mCherry WT simulations align with previous studies; and second, to assess how these interactions are altered in the mutant chromophores due to the introduced modifications. These findings are critical for constructing accurate models for subsequent QM simulations, as the hydrogen bonding environment should influence the chromophores electronic structure and optical properties.

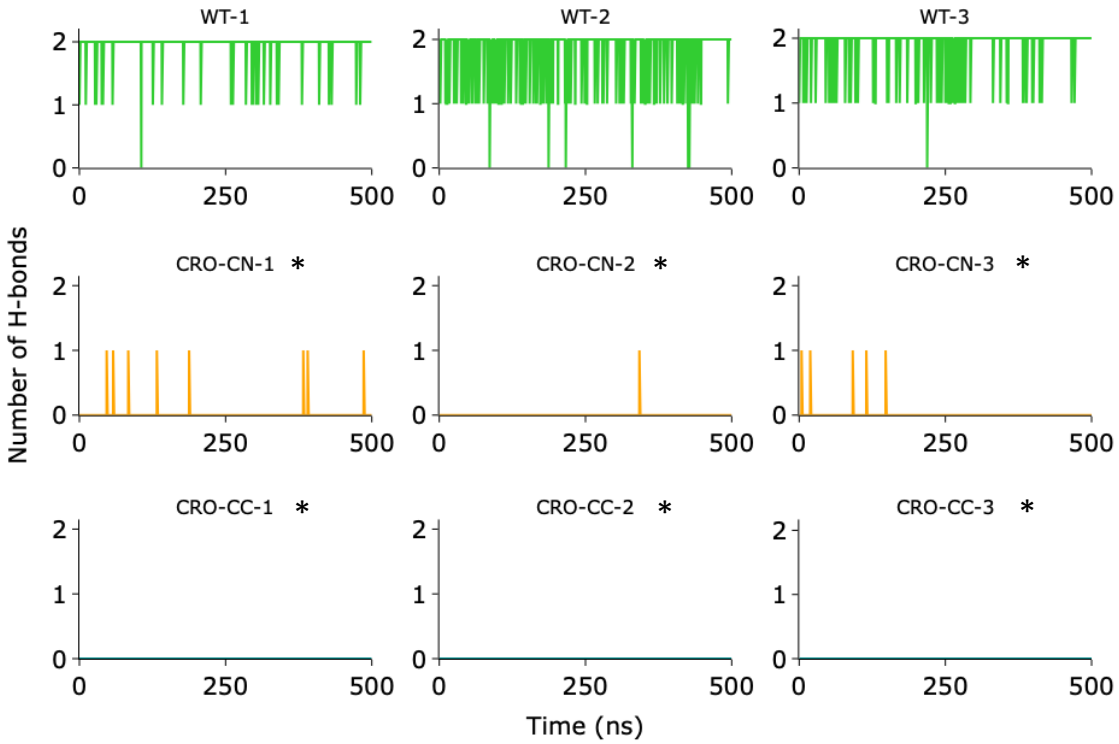


Figure 6.10: Number of H-bonds between the phenolic functional groups of the mCherry chromophore and the surrounding protein and solvent environment. For mCherry WT, the deprotonated oxygen was selected for analysis, where the nitrile for CRO-CN and ethynyl for CRO-CC was analysed. Kruskal-Wallis test was performed on each simulations to compare against each WT simulation, with * denoting significance ($p < 0.0001$). Please note that each data point represents 10 ps (50,000 data points per graph), and the line widths visually overrepresent the frequency of hydrogen bonds, creating the impression of higher occurrences than indicated by the true percentage values below.

The number of hydrogen bonds formed between the chromophore and its surrounding protein or solvent environment reveals significant differences between mCherry WT and the mutant chromophores. In mCherry WT, the deprotonated oxygen maintains a stable polar network, forming two hydrogen bonds 88.81% of the time and one hydrogen bond 10.95% of the time, with almost no instances (0.24%) of no hydrogen bonds across three simulations. These results are consistent with the known interactions of the mCherry WT chromophore as reported in the literature. In contrast, CRO-CN and CRO-CC show

drastically reduced hydrogen bonding. For CRO-CN, the nitrile group forms one hydrogen bond only 1.42% of the time, with 98.58% of the simulations showing no hydrogen bonds. Similarly, CRO-CC exhibits no hydrogen bonds throughout each simulation, as the ethynyl hydrogen does not interact with the protein or solvent environment. These findings highlight the inability of the modified functional groups to establish stable hydrogen-bonding interactions, which could influence the stability and electronic properties of the chromophore. These results will directly inform the setup of QM simulations, where the absence of hydrogen bonding in the mutants will need to be considered in modelling their electronic and optical properties.

6.2.3 Quantum mechanics simulation set up and geometry optimisation of chromophore variants

To further explore the electronic properties that influence the behaviour of the mCherry variants, a QM approach was employed, building upon the results gained from the experimental and MD analysis. The primary goal of this QM analysis is to investigate key electronic factors such as dipole moments, excitation energies, predicted absorbances and vibrational modes, which are central to understanding the chromophores optical behaviour and the potential enhancements observed in Raman spectroscopy. The B3LYP/6-31G+(d,p) level of theory was chosen for the following results due to its reliable accuracy in predicting electronic properties for organic molecules, balanced with a moderate computational cost. The Polarisable Continuum Model (PCM) using water was also included to mimic biological and cellular environments, as it approximates solvent effects and better reflects the chromophores behaviour in aqueous conditions (further discussed in detail in Methods section 2.7).

For the mCherry WT QM analysis, two structures were utilised to examine different aspects of the chromophore's electronic properties. A basic chromophore-only model was used to investigate the intrinsic properties of the delocalised electron system, focusing on dipole moments and orbital maps without the influence of external interactions (Figure 6.11.A). In contrast, a larger, extended model was employed for analysing predicted excitation energies, TD-DFT calculations, and Raman spectra. This extended model incorporates two key interacting residues, E144 and S146, as well as a water molecule

positioned between the deprotonated oxygen of the chromophore and the interacting residues, allowing for a more comprehensive assessment of the polar networks global effects. This water molecule, together with E144 and S146, form a polar network with the deprotonated oxygen, a network further stabilised by interactions with additional residues such as E148, Q163, and L199 (Shu et al. 2006; Topol et al. 2011). Furthermore, E215 forms a hydrogen bonds with the imidazolinone ring but the purpose of this experiment, focus is solely on the modifications introduced at the para-position on the phenolic ring. The exclusion of the other additional residues from the extended QM model was primarily driven by computational cost. For DFT calculations using the B3LYP or Hartree-Fock functionals, the computational time typically scales as an order of N^4, where N is the number of atoms in the system (Goletto et al. 2021). For example, if the number of atoms doubled, the computation time would increase by a factor of 16. To maintain a balance between accuracy and feasibility, only the nearest interacting residues, E144 and S146, were included in the model.

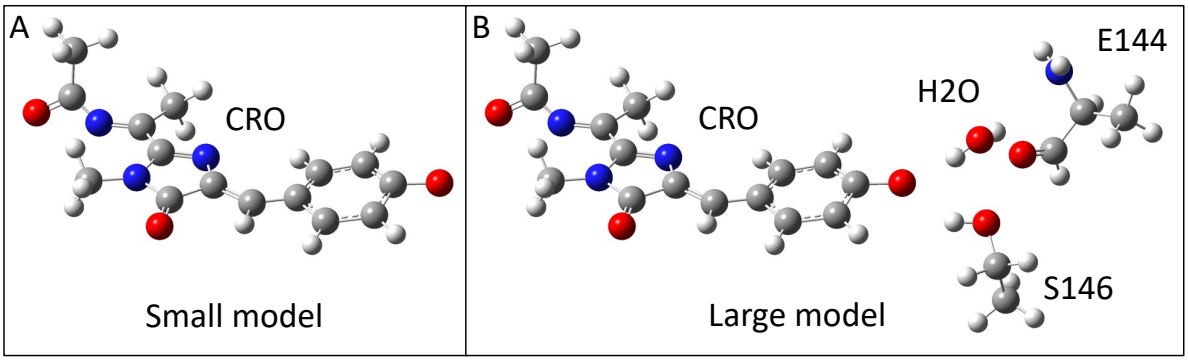


Figure 6.11: WT QM models. (A) Basic model of mCherry WT to determine dipole magnitude and direction. (B) Extended model of mCherry WT to predict excitation energies, TD-DFT and Raman spectra.

The cis-trans isomerised mCherry WT chromophore was not considered, as it is assumed that the native conformation is stabilised in the absence of mutations. This assumption allows us to focus on the primary functional state of the chromophore and its interactions with surrounding residues. In both models, the chromophore is truncated and replaced with methyl groups, to isolate the delocalised system for simulation. This methyl truncation approach is a standard practice in QM simulations, simplifying the model and

reducing computational costs while preserving the essential electronic properties of the chromophore.

For the mutant QM analysis, two models were generated for each mutant: the native conformation (NC) and an isomerised conformation (IC) (Figure 6.12). These models were derived from snapshots obtained during the MD simulations and underwent the same iterative geometry optimisation process as the mCherry WT. However, unlike the mCherry WT analysis, the polar network was excluded from these models, as the MD simulations indicated the absence of a polar network at the para-position in the phenolic ring of the mutants. One of the primary aims of this QM analysis was to assess the impact of the isomerisation on the chromophores electronic properties and to determine whether the computational results aligned with experimental observations.

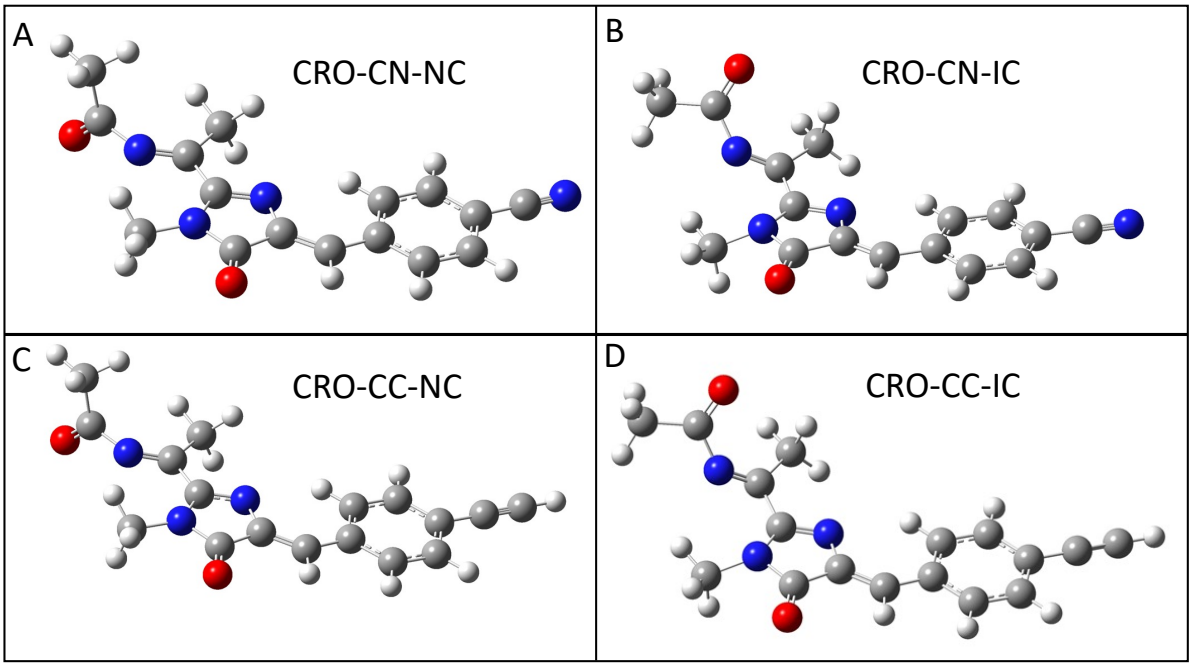


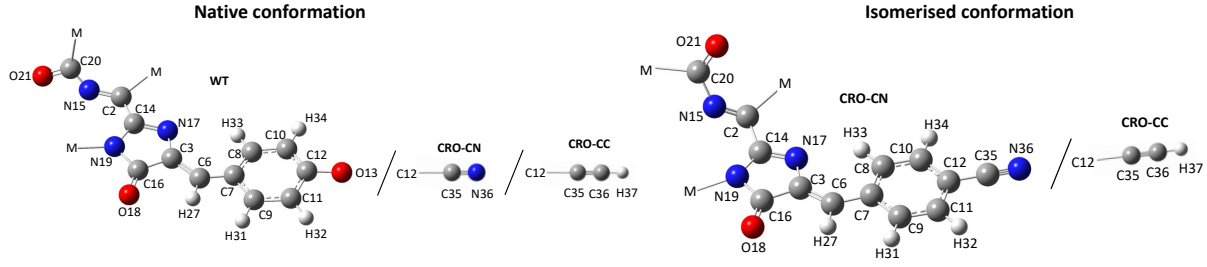
Figure 6.12: Mutant QM models. (A) CRO-CN Native Conformation (NC), (B) CRO-CN Isomerised Conformation (IC), (C) CRO-CC-NC, (D) CRO-CC-IC.

Similar to the Chapter 5 QM analysis, multiple rounds of geometry optimisation were performed on each of the chromophore model. Similar to the mCherry WT, the mutant chromophores were truncated and replaced with methyl groups. These methyl groups were fixed in position, and for mCherry WT, the interacting residues were also fixed in position to focus the optimisation on the delocalised electron system. Additionally, for the mCherry WT,

the oxygen atom in the H_2O molecule was also fixed, while the polarising hydrogens, including the hydrogen on S148, were allowed to optimise freely. The comparison of bond lengths between the B3LYP/6-31G+(d,p) optimised geometry of all models simulated and the crystal structure, as well as the comparison of the bond angles within the delocalised system, has been provided in Table 6.1. The schematic of the atom labels for both the native conformation and isomerised conformation has also been provided.

Table 6.1: Geometry optimisation of bond lengths and bond angles. The table compares data from QM simulations to the crystal structure of the chromophore (PDB: 2H5Q, CH6), with the % difference provided in the brackets where applicable. Schematic of the atom labels are also shown below, M represents the methyl groups incorporated.

Bond	Bond Length (Å) (% difference from Crystal Structure, where applicable)					
	Native Conformation (NC)				Isomerised Conformation (IC)	
	Crystal Structure	WT	CRO-CN-NC	CRO-CC-NC	CRO-CN-IC	CRO-CC-IC
C2-C14	1.48254	1.49566 (+0.22%)	1.50518 (+0.38%)	1.50379 (+0.36%)	1.49698 (+0.24%)	1.49456 (+0.20%)
C2-N15	1.30481	1.29116 (-0.26%)	1.27373 (-0.60%)	1.27479 (-0.58%)	1.28112 (-0.46%)	1.28210 (-0.44%)
C3-C6	1.37348	1.33999 (-0.62%)	1.35604 (-0.32%)	1.35845 (-0.28%)	1.35609 (-0.32%)	1.35878 (-0.27%)
C3-C16	1.46642	1.49509 (+0.48%)	1.49053 (+0.41%)	1.48627 (+0.34%)	1.49100 (+0.42%)	1.48739 (+0.35%)
C6-C7	1.38140	1.44001 (+1.04%)	1.45200 (+1.25%)	1.44823 (+1.18%)	1.45172 (+1.24%)	1.44763 (+1.17%)
C6-H27	1.09034	1.08989 (-0.01%)	1.08634 (-0.09%)	1.08672 (-0.08%)	1.08637 (-0.09%)	1.08675 (-0.08%)
C7-C8	1.42315	1.39145 (-0.56%)	1.41088 (-0.22%)	1.41171 (-0.20%)	1.41089 (-0.22%)	1.41184 (-0.20%)
C7-C9	1.42882	1.40364 (-0.44%)	1.40989 (-0.33%)	1.41067 (-0.32%)	1.40991 (-0.33%)	1.41081 (-0.32%)
C9-C11	1.36116	1.39997 (+0.70%)	1.38558 (+0.44%)	1.38535 (+0.44%)	1.38554 (+0.44%)	1.38520 (+0.44%)
C9-H31	1.09018	1.09065 (+0.01%)	1.08373 (-0.15%)	1.08426 (-0.14%)	1.08372 (-0.15%)	1.08425 (-0.14%)
C10-C12	1.45926	1.38714 (-1.27%)	1.40546 (-0.94%)	1.40911 (-0.87%)	1.40550 (-0.94%)	1.40923 (-0.87%)
C10-H34	1.09054	1.08965 (-0.02%)	1.08273 (-0.18%)	1.08317 (-0.17%)	1.08269 (-0.18%)	1.08314 (-0.17%)
C11-C12	1.45146	1.40261 (-0.86%)	1.40304 (-0.85%)	1.40634 (-0.79%)	1.40298 (-0.85%)	1.40635 (-0.79%)
C11-H32	1.08999	1.08985 (0.00%)	1.08248 (-0.17%)	1.08290 (-0.16%)	1.08245 (-0.17%)	1.08287 (-0.16%)
C14-N17	1.36142	1.30789 (-1.00%)	1.30093 (-1.14%)	1.30404 (-1.08%)	1.30255 (-1.10%)	1.30278 (-1.10%)
C14-N19	1.43547	1.41291 (-0.40%)	1.41667 (-0.33%)	1.41719 (-0.32%)	1.42221 (-0.23%)	1.42328 (-0.21%)
N15-C20	1.17794	1.36358 (+3.65%)	1.38674 (+4.07%)	1.38702 (+4.08%)	1.37418 (+3.84%)	1.37274 (+3.82%)
C16-O18	1.21631	1.22711 (+0.22%)	1.22157 (+0.11%)	1.22349 (+0.15%)	1.22071 (+0.09%)	1.22251 (+0.13%)
C16-N19	1.42509	1.40102 (-0.43%)	1.39105 (-0.60%)	1.39140 (-0.60%)	1.39490 (-0.54%)	1.39559 (-0.52%)
C20-O21	1.23819	1.22985 (-0.17%)	1.22089 (0.35%)	1.22128 (-0.34%)	1.22855 (-0.20%)	1.22927 (-0.18%)
C12-O13	1.25982	1.35402 (+1.80%)	N/A	N/A	N/A	N/A
C12-C35	N/A	N/A	1.42900	1.42601	1.42906	1.42587
C35-N36	N/A	N/A	1.15663	N/A	1.15654	N/A
C35-C36	N/A	N/A	N/A	1.20636	N/A	1.20635
C36-H37	N/A	N/A	N/A	1.06446	N/A	1.06449
Selected Angles	**Bond Angle (°) (% difference from Crystal Structure, where applicable)**					
	Native Conformation (NC)				Isomerised Conformation (IC)	
	Crystal Structure	WT	CRO-CN-NC	CRO-CC-NC	CRO-CN-IC	CRO-CC-IC
O21-C20-N15	109.069	119.002 (+2.18%)	121.606 (+2.72%)	121.577 (+2.71%)	120.861 (+2.56%)	121.020 (+2.60%)
C20-N15-C2	132.058	138.784 (+1.24%)	135.884 (+0.71%)	136.129 (+0.76%)	133.228 (+0.22%)	133.223 (+0.22%)
N15-C2-C14	126.439	116.885 (-1.96%)	115.911 (-2.17%)	116.212 (-2.11%)	111.660 (-3.10%)	111.899 (-3.05%)
C2-C14-N17	124.840	120.409 (-0.90%)	119.105 (-1.18%)	119.119 (-1.17%)	119.080 (-1.18%)	119.196 (-1.16%)
C14-N17-C3	105.921	106.917 (0.23%)	107.112 (+0.28%)	107.161 (+0.29%)	106.892 (+0.23%)	106.944 (+0.24%)
N17-C3-C6	131.897	130.786 (-0.21%)	128.635 (-0.63%)	128.530 (-0.65%)	128.576 (-0.64%)	128.458 (-0.66%)
C3-C6-C7	134.307	130.932 (-0.64%)	129.923 (-0.83%)	130.299 (-0.76%)	130.020 (-0.81%)	130.381 (-0.74%)
C6-C7-C8	121.724	123.296 (+0.32%)	123.678 (+0.40%)	123.909 (+0.44%)	123.717 (+0.41%)	123.942 (+0.45%)
C6-C7-C9	121.334	119.370 (-0.41%)	117.781 (-0.74%)	117.924 (-0.71%)	117.723 (-0.76%)	117.872 (-0.72%)
C7-C8-C10	123.457	123.181 (-0.06%)	120.602 (-0.58%)	120.661 (-0.57%)	120.591 (-0.59%)	120.650 (-0.57%)
C7-C9-C11	121.066	121.922 (+0.18%)	121.230 (+0.03%)	121.310 (+0.05%)	121.219 (+0.03%)	121.299 (+0.05%)
C8-C10-C12	119.775	121.381 (+0.33%)	120.048 (+0.06%)	120.778 (+0.21%)	120.037 (+0.05%)	120.767 (+0.21%)
C10-C12-C11	116.056	118.836 (+0.59%)	120.118 (+0.86%)	118.907 (+0.61%)	120.139 (+0.86%)	118.931 (+0.61%)
C9-C11-C12	122.026	119.084 (-0.61%)	119.462 (-0.53%)	120.177 (-0.38%)	119.454 (-0.53%)	120.167 (-0.38%)
C11-C12-O13	125.564	121.527 (-0.82%)	N/A	N/A	N/A	N/A
C10-C12-O13	118.237	122.611 (+0.91%)	N/A	N/A	N/A	N/A
C11-C12-C35	N/A	N/A	119.920	120.567	119.944	120.571
C10-C12-C35	N/A	N/A	119.962	120.526	119.917	120.498
C12-C35-N36	N/A	N/A	179.947	N/A	179.957	N/A
C12-C35-C36	N/A	N/A	N/A	179.966	N/A	179.945
C35-C36-H37	N/A	N/A	N/A	179.934	N/A	179.996
Dihedral angle of O21 – C20 – N15 – C2: Crystal structure: 175.488°, WT: 89.293°, CRO-CN-NC: 87.702°, CRO-CC-NC: 86.471°, CRO-CN-IC: -36.636°, CRO-CC-IC: -35.986°						



Overall, the QM-optimised bond lengths are consistent with the experimental crystallographic data, with an average difference of approximately 0.034 Å across all bonds in the mCherry WT. The largest bond length deviation between the crystal structure and geometry optimised mCherry WT datasets is observed for the N15-C20 bond, with a difference of +0.14564 Å (+3.65% difference). This large difference is most likely due to the way the crystallographic structure processes the chromophore. Atoms O21, C20 and N15 form part of the backbone of residue F65 in mCherry, despite being part of the delocalised system within the mCherry chromophore. In the PDB structure, these atoms are still treated as part of F65 and therefore adopt the average bond lengths and angles of a typical phenylalanine residue. Similarly, the biggest difference in angle (C20-N15-C2; -9.726°) is also located in this region, which connects F65 to the chromophore. Another point of interest is the C12-O13 bond, which is significantly longer in the QM model (1.35402 Å) compared to the crystal structure (1.25982 Å). This is most likely due to the crystal structure treating this as a double bond, when in fact it's a deprotonated oxygen within the resonant bond structure. Given the precision of QM optimised geometry simulations, the computed structure gives a more accurate representation of the overall molecule, accounting for the resonance and deprotonation states.

A comparison of dihedral angles further highlights structural differences between the crystal structure, mCherry WT, and mutant chromophores. For the O21–C20–N15–C2 dihedral angle, the crystal structure reports a value of 175.488°, while the QM-optimised mCherry WT model yields 89.293°. As previously mentioned, the crystal structure treats this dihedral as a typical amino acid backbone, adopting the average bond lengths and angles, whereas the QM model accounts for the delocalised electronic system and resonance effects, providing a more accurate representation. The mutants display similar dihedral angles in their native conformations: CRO-CN-NC (87.702°) and CRO-CC-NC (86.471°). As expected, the isomerised conformations exhibit different values: CRO-CN-IC (-36.636°) and CRO-CC-IC (-35.986°). The negative dihedral angles in the ICs indicate a counter-clockwise rotation relative to the plane, with values falling within the range of -180° to +180°. Despite these dihedral differences in the ICs, the bond lengths and angles remain consistent between the mCherry WT and NC-mutant models. This similarity underscores the robustness of the chromophores delocalised electronic structure across different

conformations and variants, supporting the idea that isomerisation primarily impacts the spatial arrangement rather than the intrinsic bonding properties of the chromophore.

6.2.3.1 Dipole moments

The dipole moments for the mCherry WT and the variants reveal substantial differences across the mutations, which will likely impact the chromophores electronic properties and behaviour in Raman spectroscopy (Figure 6.13). The mCherry WT exhibits a large dipole moment of 11.6 Debye (D), indicating a significant separation of charge. For reference, a typical $HBDI^-$ molecule, found in most β-barrel FPs, was calculated to have 1.7 D, using the same B3LYP method (Filippi et al. 2012). However, the QM simulations suggest that the overall polarity of this system is predominantly influenced by the presence of the deprotonated oxygen, which plays a key role in driving the chromophores polarisation. The CRO-CN and CRO-CC variants, which lack this deprotonated oxygen, exhibit reduced dipole magnitudes compared to the mCherry WT. Where CRO-CN-NC has a dipole moment of 4.59 D, CRO-CN-IC 7.06 D, CRO-CC-NC 5.01 D, and CRO-CC-IC 4.54 D, these finding underscores the significant contribution of the deprotonated oxygen to the charge separation and dipole direction in the mCherry WT chromophore. The lower dipole moments in the variants suggest weaker polar interactions, likely leading to diminished Raman enhancement. A stronger dipole moments result in larger polarisability changes during molecular vibrations, enhancing Raman signal intensity (Atkins 2022). Consequently, the diminished dipole moments in the variants likely reduce their Raman intensity compared to mCherry WT.

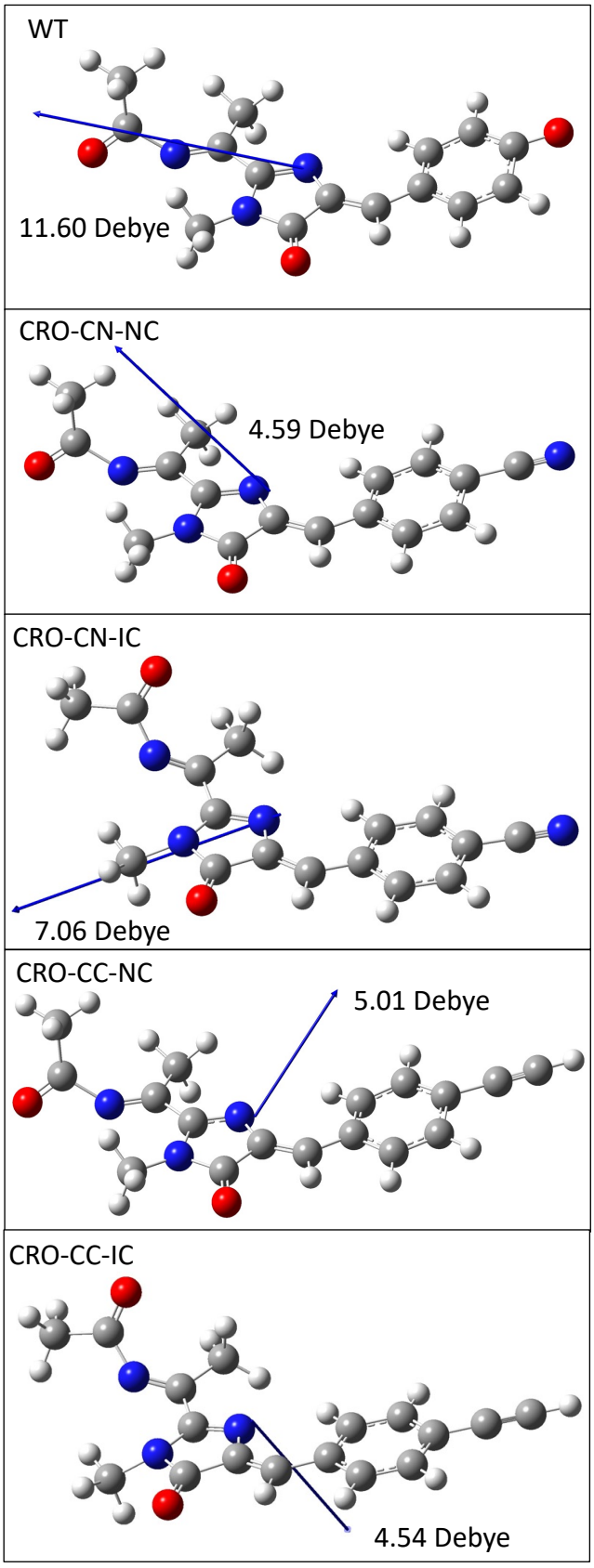


Figure 6.13: Dipole moments of WT and the mutations. Each box contains the dipole magnitude (in Debye; D) and the direction of the moment (blue arrow), which originates from the centre of electronic charge.

The direction of the dipole moment plays a critical role in how the chromophore interacts with external electromagnetic fields, particularly in Raman spectroscopy. Alignment of the dipole moment with the incident electric field can enhance or suppress vibrational modes, directly affecting the Raman signal intensity observed for certain transitions (Jones et al. 2019). If the dipole moment direction aligns favourably with the light's electric field, stronger Raman signals may be observed for certain vibrational transitions. Conversely, a misalignment or altered dipole direction, as seen in CRO-CN and CRO-CC, can reduce the efficiency of this coupling. This leads to weaker Raman intensities, as reflected in the experimental data (Figure 6.3).

Additionally, the differences between native and isomerised conformations in both CRO-CN and CRO-CC further impact dipole moment direction and magnitude. The IC models show a pronounced shift in dipole direction compared to the NC models. This arises because isomerisation alters the spatial distribution of charge within the chromophore by reorienting a key functional group comprising the backbone of F65 and potentially disrupting the polarisation of the chromophore. Such changes in charge distribution modify the dipole moment vector, influencing how the chromophore interacts with incident light.

6.2.3.2 HOMO/LUMO gap, predicted excitation energies and TD-DFT

The HOMO (Highest Occupied Molecular Orbital) and LUMO (Lowest Unoccupied Molecular Orbital) represent the orbital maps of a molecule and are critical in determining its electronic properties. The HOMO is the highest energy orbital that contains electrons, while the LUMO is the lowest energy orbital that can accept electrons. The energy difference between these two orbitals, known as the HOMO-LUMO gap, is a key factor in estimating the molecules excitation energy, which is related to its ability to absorb light and undergo electronic transitions (Huang et al. 2017). However, it must be noted it does not account for certain effects that influence absorption, such as vibronic coupling (interaction between electronic and vibrational states), transition dipole moments and solvent interactions (Lupo Pasini et al. 2023). Therefore, the energy difference between ground to excited states is an approximation and does not capture the absorption profile (as it is only looking at one transition). Nevertheless, TD-DFT calculations mitigate these limitations by considering

electron-electron interactions, excited-state relaxation, and vibronic coupling (Jacquemin et al. 2011; Adamo and Jacquemin 2013; Escudero et al. 2017).

Investigating the HOMO/LUMO orbitals of the mCherry variants, can provide information into how structural modifications can influence their electronic behaviour. This analysis will help predict how each variant interacts with light, which in turn will affect optical properties such as absorbance and Raman. For the mCherry WT orbital maps, the basic model from Figure 6.11.A, was implemented in order to exclude external influences on the chromophore's electronic configuration. Using the extended model would introduce bias in favour of the interacting residues, and since this model does not have all of the residues that interact with the chromophore, the decision was made to analyse the chromophore only. The extended model for mCherry WT was still used to measure HOMO-LUMO energy gap and to calculate TD-DFT. Each HOMO/LUMO orbital map is presented in Figure 6.14. In the mCherry WT chromophore, the electron density is evenly distributed across both the HOMO and LUMO orbitals, reflecting efficient delocalisation and a stable electronic system within the chromophores conjugated system. The positive (green) and negative (red) lobes represent the distribution of electron density, showing the regions where electron transitions are most likely to occur. This balanced delocalisation across the chromophore suggests optimal conditions for light absorption and efficient electronic transitions, key factors in maintaining its optical properties.

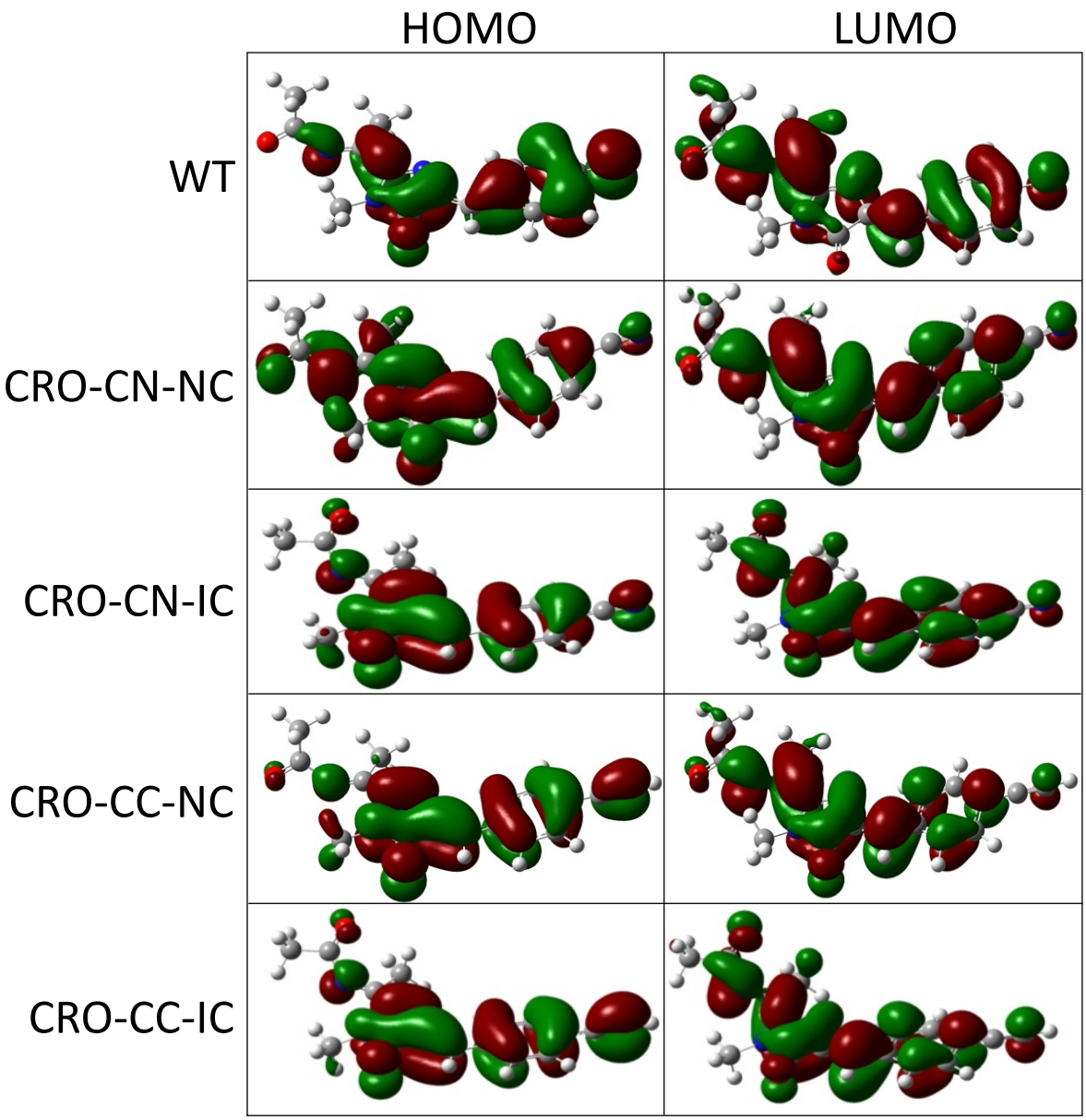


Figure 6.14: HOMO and LUMO orbitals for each mCherry variant. B3LYP/6-31G+(d,p) method and basis set displaying the distribution of orbitals in HOMO (left) and LUMO (right).

For CRO-CN and CRO-CC, both native and isomerised conformations exhibit similar HOMO and LUMO distributions, with no significant shifts in electron density compared to the mCherry WT. This consistency suggests that the mutations and isomerisation does not drastically alter the fundamental electronic structure or the delocalisation of the conjugated system. The lack of noticeable differences in orbital distributions between the NC and IC conformations implies that the isomerisation predominantly affects the spatial arrangement of the chromophore rather than significantly altering its intrinsic electronic properties. This

observation aligns with the geometry optimisation and dipole moment analysis, where changes in dipole direction and magnitude were more prominent than alterations in the electronic structure (i.e. no change in the bond order).

To determine the predicted excitation energies, the energy gap between the ground state (HOMO) and the excited state (LUMO) levels were calculated. The energy difference corresponds to the energy absorbed by the chromophore during an electronic transition, which dictates the wavelength of light absorbed. The differences can be observed in Figure 6.15. The mCherry WT chromophore was predicted to have an excitation energy of 2.13951 eV, which demonstrates the computational model's ability to accurately capture the electronic properties of the chromophore. The slight deviation from experimental measurements, which report an excitation energy equivalent to a wavelength of 587 nm (~2.11 eV), can likely be attributed to the exclusion of key interacting residues (E148, Q163, and L199) that form part of the polar network near the deprotonated oxygen, and also E215 which interacts with the imidazolinone ring. Considering these exclusions, the QM method was effective in capturing the chromophores fundamental properties. Including more interacting residues or applying a more comprehensive method could potentially refine the accuracy of the prediction further. The accuracy of these simulations is consistent with prior TD-DFT benchmarks on FP chromophores and literature values for similar red FPs. For example, the monomeric absorbs at ~555 nm (2.23 eV) experimentally (predicted 551 nm in cluster calculations), slightly blue-shifted relative to mCherry due to its modestly different chromophore structure (Khrenova et al. 2015). Another study also characterised mCherry's transition dipole in crystals, confirming a well-defined excitation vector, which aligns with our finding of a strong $S_0 \rightarrow S_1$ transition (543 – 594 nm) using an experimental procedure (Myšková et al. 2020).

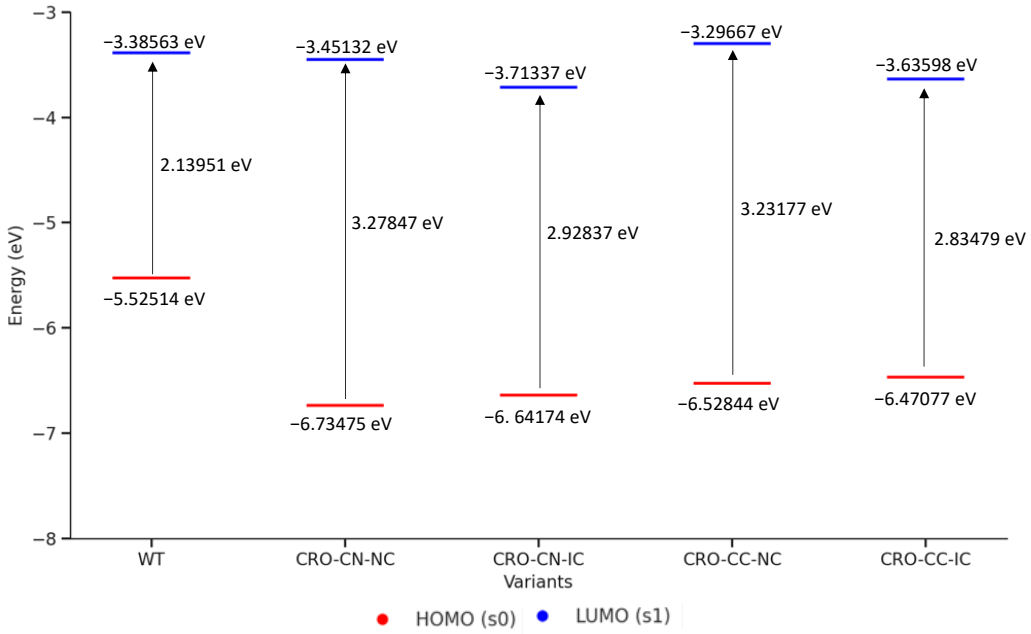
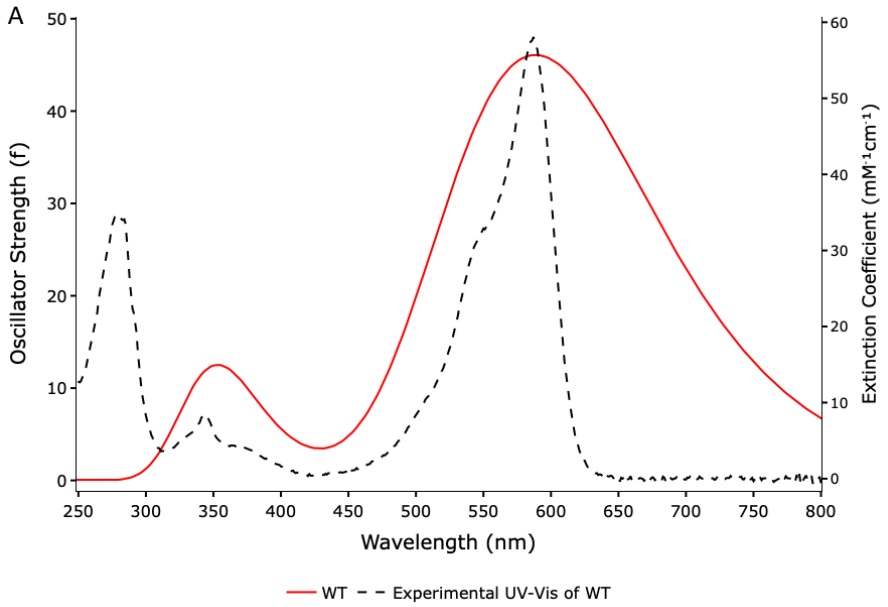
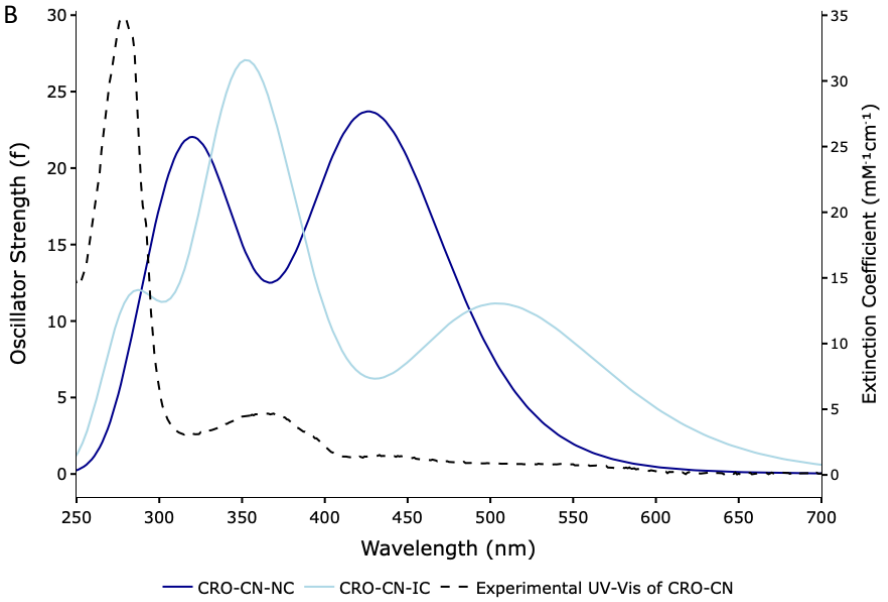


Figure 6.15: HOMO-LUMO energy levels for each mCherry variant. Each energy level is labelled with the corresponding energy in eV, with the HOMO-LUMO gap labelled as eV. Red lines represent the ground state HOMO (S_0) level whereas blue lines show the excited state LUMO (S_1) level.

The excitation energies for the CRO-CN and CRO-CC variants are consistently higher than the mCherry WT, indicating that these mutations result in a blue-shifted electronic transition. For the native conformations, CRO-CN-NC and CRO-CC-NC show excitation energies of 3.27847 eV and 3.23177 eV, respectively, while their isomerised counterparts have slightly lower excitation energies, with CRO-CN-IC at 2.92837 eV and CRO-CC-IC at 2.83479 eV. The higher excitation energies observed in all the mutant states suggest that their electronic structures require more energy for the HOMO-LUMO transition, likely due to the removal of the deprotonated oxygen, which plays a critical role in polarising the chromophores electronic system. The same principle was observed in another computational study where deprotonating the deprotonating the p-hydroxybenzylidene-imidazolinone (creating an anion) red-shifts the first absorption by ~70 nm, so losing that negative charge induces a sizable blue-shift (Ansbacher et al. 2012). The WT RFPs with an extended π-system (e.g. DsRed) typically absorb around 550 – 570 nm (2.1 – 2.3 eV), whereas the phenolate-lacking mutants absorb in the 320 – 425 nm (2.9 – 3.9 eV) range. This ~100+ nm blue-shift underscores how crucial the anionic phenolate and extended conjugation are for lowering the $S_o \rightarrow S_1$ energy.

The HOMO-LUMO gap provides an approximate measure of the absorption energy, while TD-DFT offers a more refined prediction of UV-Vis absorption peaks by incorporating additional interactions, as previously mentioned. In the predicted UV-Vis spectra, the mutants show alternative absorption maximums depending on the conformation, demonstrating the effects of ncAA incorporation and chromophore conformation on the electronic properties (Figure 6.16). The mCherry WT spectrum peaks at 588 nm, closely matching the experimental value (587 nm), confirming the accuracy of the TD-DFT predictions. However, the theoretical spectrum shows excessive broadening beyond 800 nm due to the default Gaussian broadening factor (σ = 0.4 eV), overestimating absorption at longer wavelengths where the experimental data cuts off at ~620 nm. At room temperature, higher vibrational levels of the ground state are sparsely populated (Boltzmann distribution), limiting transitions to higher vibrational levels of the excited state. Reducing the broadening factor or using explicit vibronic coupling could better align the theoretical spectrum with experimental results.


A
Oscillator Strength (f)
Extinction Coefficient (mM-1cm-1)
Wavelength (nm)
WT
Experimental UV-Vis of WT


B
Oscillator Strength (f)
Extinction Coefficient (mM-1cm-1)
Wavelength (nm)
CRO-CN-NC
CRO-CN-IC
Experimental UV-Vis of CRO-CN

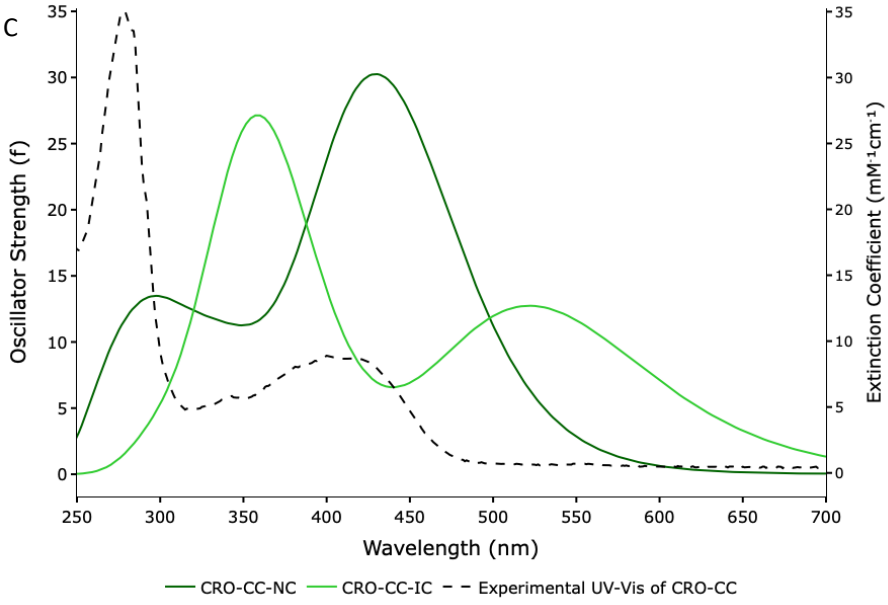

C
Oscillator Strength (f)
Extinction Coefficient (mM-1cm-1)
Wavelength (nm)
CRO-CC-NC
CRO-CC-IC
Experimental UV-Vis of CRO-CC

Figure 6.16: Predicted UV-Vis spectroscopy of mCherry variants against their experimental absorbances. TD-DFT B3LYP/6-31G+(d,p) calculation (nstates = 10) of each variant; (A) WT, (B) CRO-CN and (C) CRO-CC overlayed onto their respective experimental absorbance data (black dashed lines). Dark coloured lines for simulated spectrums represents NC, whereas light coloured lines represents IC for the respective mutants. Oscillator strength (f) is an arbitrary, dimensionless unit.

The experimental absorption maximum for CRO-CN is observed at 367 nm. In the TD-DFT predictions, the native conformation shows two primary absorption peaks at 320 nm and 425 nm, while the isomerised conformation has three peaks at 285 nm, 352 nm and 502 nm. These results suggest that both conformations contribute to the experimental spectrum, with the experimental peak at 367 nm potentially representing an average of the chromophores mixed population. For CRO-CC, the experimental spectrum shows a broad absorption maximum at 405 nm, spanning a range from 350 nm to 450 nm. The TD-DFT predictions for the native conformation identify peaks at 298 nm and 427 nm, while the isomerised conformation shows peaks at 360 nm and 522 nm. The experimental broad peak also likely arises from a combination of contributions from both conformations, with the 405 nm feature reflecting the overlap of multiple transitions.

In Section 6.2.2, the MD simulations reveal that the transition from the NC to the IC is irreversible under the simulated condition, suggesting that IC is more energetically favourable after the transition. However, the computational setup may not fully replicate the experimental conditions, potentially allowing some NC to persist in the sample, or to allow the reversible isomerisation back to the native form. Consequently, the observed experimental spectra likely represent an averaged signal from a mixed population of conformations, with the relative contributions of NC and IC depending on their respective stabilities under experimental conditions. Furthermore, both conformations exhibit a blue-shifted absorption compared to the mCherry WT chromophore. This shift is primarily attributed to the absence of the deprotonated oxygen, which likely plays a critical role in the mCherry WT chromophores electronic properties. The exclusion of the polar network in the mutant models, which would stabilise the deprotonated oxygen in mCherry WT, further supports this observation.

6.2.3.4 Predicting Raman activity

Having explored the geometry optimisation, dipole moments, and electronic transitions of the chromophore variants, the final step is to predict their Raman spectra and compare these predictions with experimental data. Predicted Raman spectroscopy provides critical insights into the vibrational modes of the chromophore, which are directly influenced by the molecular geometry and electron distribution identified in earlier analyses. By using the QM models developed so far, it is possible to simulate the vibrational frequencies and intensities associated with key chemical bonds in each chromophore variant. These predictions will then be compared to experimental Raman spectra, allowing for a deeper understanding of what vibrational modes correspond to certain peaks, as well as how each mutation alters the chromophores vibrational behaviour and Raman activity.

In epr-SRS, the excitation wavelength is chosen near the chromophore's electronic absorption, which enhances the Raman signals for specific vibrational modes. The proximity of the excitation energy to the HOMO-LUMO gap makes it possible to selectively amplify certain vibrational modes that couple most strongly with the chromophore's electronic transitions. This is particularly important for understanding how mutations like those in the mutated chromophores alter not only their absorption properties but also their Raman activity. The following section explores the predicted Raman spectra for each chromophore variant, comparing these theoretical predictions to experimental epr-SRS measurements to assess how well the QM models capture these vibrational dynamics.

The first prediction was performed on the mCherry WT, within the fingerprint region, with most of the key peaks aligning well between the experimental data and the prediction (Figure 6.17). These slight shifts in wavenumber are expected due to the simplified QM model which do not fully account for other nearby interacting residues present. As before, the following Raman peak assignments are based on references from Lin-Vien et al. (1991) Raman characteristics handbook. The most intense peak is observed at 1160-1180 cm^{-1}, corresponding to C-O^- stretching (ii) at the para-position of the phenolic ring within the chromophore. As discussed in Section 6.2.1, the negative charge in this region results in a higher electron density, leading to significant pre-resonant enhancement. In 1250-1450 cm^{-1} range (iii), the peaks are associated with C–C and C–N stretching vibrations, originating from bonds within the chromophore's conjugated structure. Finally, the pronounced peaks

between 1500-1650 cm^{-1} (iv) correspond to C=C stretching, predominantly within the aromatic ring of the chromophore. This mode is characteristic of conjugated double bonds in aromatic systems and is strongly enhanced by resonance within the ring structure, contributing to the intense Raman signal in this region. Discrepancies between the simulated and experimental frequencies are likely due to the absence of interacting residues in the QM model, which could influence the vibrational properties. The following spectra also has a scaling factor of 0.97 applied to the wavenumbers as B3LYP has been reported to slightly overestimate the raw frequencies Click or tap here to enter text.(Sinha et al. 2004).

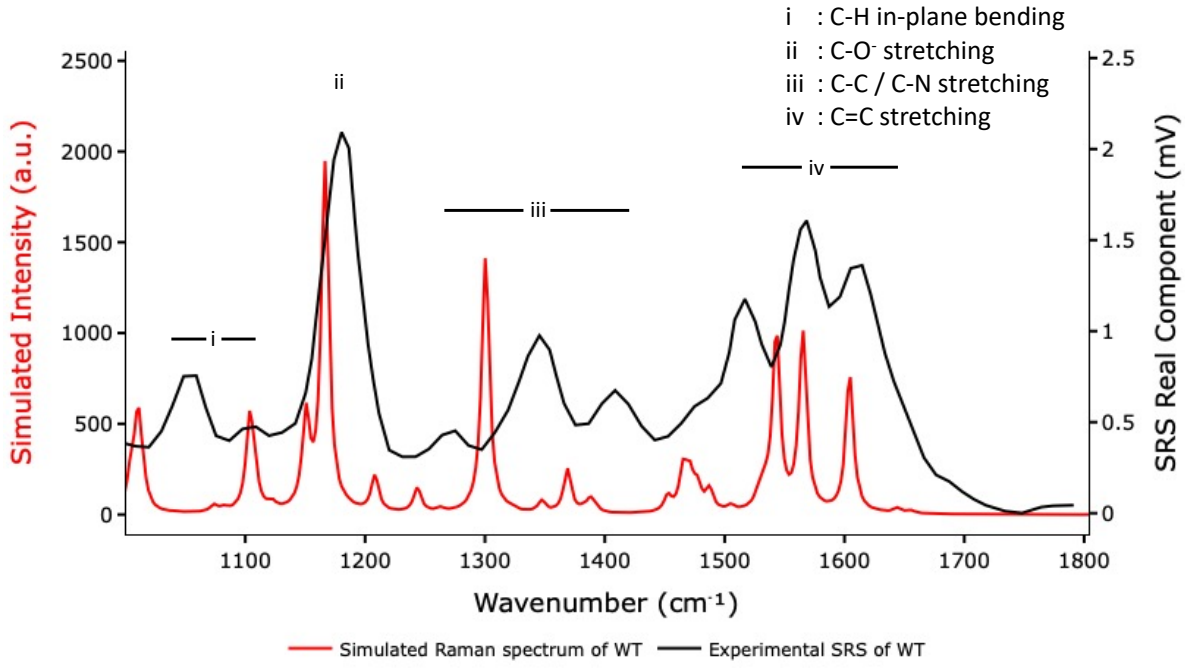


Figure 6.17: Predicted vs. experimental Raman spectroscopy of mCherry WT. The simulated spectra (black) is compared to the epr-SRS measurements (red) with selected peaks labelled. (i) C-H in-plane bending, (ii) $C\text{-}O^-$ stretching, (iii) C-C / C-N stretching and (iv) C=C stretching. A scaling factor of 0.97 was applied to the wavenumbers for the simulated spectra.

The epr-SRS data presents peaks of enhanced vibrational modes within the chromophores double conjugated system, whereas the predicted Raman spectrum generated includes all bond vibrations within the chromophore as well as bonds within the interacting residues in the mCherry WT. As a result, not all predicted peaks will overlap with the experimental data. Furthermore, the interpretation of specific vibrational modes at specific wavenumbers is often vague across literature, with conflicting reports on assigning certain bonds to different frequencies. Therefore, the predicted results present an

estimation of the enhanced bonds involved, while acknowledging that certain peaks may differ based on the methodology and the system submitted to the QM simulation.

The CRO-CN and CRO-CC mutants, along with their alternative conformations, were compared to experimental measurements, as shown in Figure 6.18. Notably, the positions of the nitrile and ethynyl peaks in the simulated Raman spectra closely align with the experimental data, confirming the successful incorporation of these mutations into the mCherry chromophore. The mutations, which introduce bulky groups into the chromophore environment, disrupt the polar network observed in the mCherry WT protein. This disruption results in a shift toward two alternative states for the chromophore, as indicated by the experimental absorbance data and TD-DFT calculations. These states are likely shared in the experimental sample, as suggested by the close agreement between the predicted absorbances for the native and isomerised conformations. The blue-shift in absorbance caused by the mutations results in weaker pre-resonance effects, as the absorption maxima are further away from the 640 nm pump wavelength used in SRS measurements. This absorbance shift explains the weak SRS signals observed experimentally.

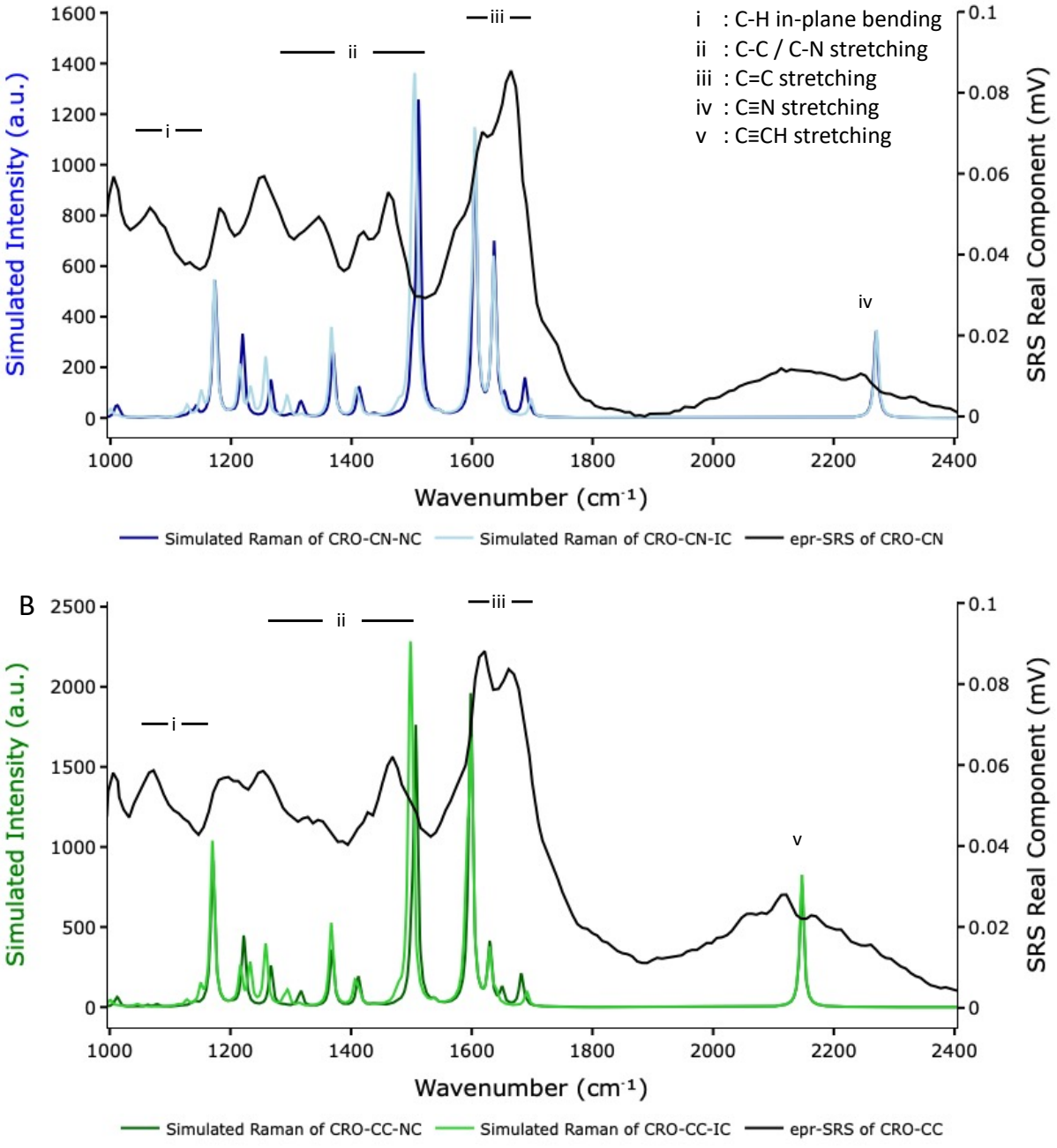


Figure 6.18: Predicted vs. experimental Raman spectroscopy of mutants. The epr-SRS spectra (black) is compared to the predicted Raman measurements of (A) CRO-CN and (B) CRO-CC with selected peaks labelled. (i) C-H in-plane bending, (ii) C-C / C-N stretching, (iii) C=C stretching, (iv) C≡N stretching and (v) C≡C stretching. A scaling factor of 0.97 was applied to the wavenumbers for the simulated spectra.

One peak is not present in these variants, compared to mCherry WT, due to the placement of the chemical modifications, and that is the C-O^- stretching at 1180 cm^{-1}, which is expected since both the nitrile and ethynyl replace the deprotonated oxygen on the

phenyl ring. Both mutants exhibit good alignment between the predicted vibrational frequencies and experimental peak positions, particularly for key modes such as C-C / C-N stretching (ii) and C=C stretching (iii). However, discrepancies are evident, particularly in the CRO-CC experimental epr-SRS spectrum, which shows a distinct double peak in the 1621–1661 cm^{-1} range, whereas the simulated spectra predict only one prominent peak, with a smaller secondary peak that is less pronounced compared to the experimental data. This difference likely arises from the simplified conditions in the QM simulations, which do not account for all environmental and solvation effects present in the experimental system. In particular, the interaction of the chromophore with the protein scaffold and solvent could enhance the secondary peak in the experimental data, contributing to the observed double peak. Minimal differences are observed between the simulated spectra of the NC and IC for both mutants. The vibrational frequencies and intensities are nearly identical, indicating that the structural changes associated with the NC to IC transition have little impact on the chromophores intrinsic vibrational properties. This suggests that the electronic properties driving these vibrations remain consistent across both conformations. Consequently, the experimental spectra may exhibit an average of contributions from both NC and IC conformations, as their vibrational fingerprints are indistinguishable within the resolution of the SRS measurements.

There is potential for multiple peaks to be more prominent and better resolved in the experimental spectra if the pump laser is tuned closer to the absorption maxima of the mutants. This adjustment could enhance resonance effects and improve the intensity of specific vibrational modes, revealing additional features in the spectra. However, increasing the pump laser energy poses a trade-off, as higher energy could lead to photobleaching of the samples, particularly under prolonged exposure.

The QM approaches employed in this study provided detailed insights into the chromophores geometry, electronic properties, optical transitions, and vibrational characteristics, aligning well with the experimental data. The geometry optimisation calculations effectively captured the chromophores structural features, including bond lengths, bond angles, and dihedral angles, which closely matched the experimental crystallographic data for the mCherry WT chromophore (except for the F65 component). The analysis of intrinsic electronic properties, such as dipole moments and orbital distributions, highlighted the effects of the mutations on the chromophores electronic structure. Whereas,

TD-DFT calculations provided a more refined prediction of the chromophores optical properties, with the predicted absorption maxima for the mCherry WT closely aligning with experimental measurements. For the mutants, TD-DFT revealed blue-shifted absorbance maxima that reflected the disrupted polar network and altered conformations. Finally, the stimulated Raman spectra further validated the computational models, accurately predicting key vibrational frequencies for both mCherry WT and mutant chromophores. The computational approaches demonstrate significant strengths in capturing the chromophores' intrinsic properties and predicting optical and vibrational behaviour. However, limitations such as the exclusion of the protein environment and explicit solvent effects introduce discrepancies, particularly in relative peak intensities of each spectroscopic method.

Despite these challenges, the predictions provide a robust framework for understanding the electronic and vibrational characteristics of chromophore variants. These insights can guide future efforts to engineer chromophores with tailored optical properties, while also offering practical guidance for experimental design, such as optimising excitation wavelengths for epr-SRS or identifying specific structural features for further mutagenesis via ncAA incorporation. Overall, the integration of computational and experimental data proves invaluable for advancing our understanding and application of engineered chromophores.

6.3 Conclusion

This chapter demonstrated a number of key experimental and computational outcomes, contributing valuable insights into the understanding and prediction of chromophore behaviour. The successful epr-SRS of the mCherry WT chromophore using a 640 nm Pump laser revealed multiple measurable peaks, which confirms the method's sensitivity to vibrational modes in pre-resonant conditions. Additionally, the detection of CRO-CN and CRO-CC signals in the biologically silent region, despite the chromophores sampling two potential alternative conformations, highlights the structural and optical plasticity of the chromophores.

The MD and QM analyses were critical in reinforcing and explaining the experimental findings. The MD simulations revealed that the CRO can potentially sample two alternative conformations through isomerisation around the C=O bond in the backbone of F65 forming

the IC state. This is a novel observation that to my knowledge has not been observed in any crystal structures to date. Interestingly, the MD suggests that while the alternative IC state is only transiently sampled in mCherry WT, when the phenol oxygen is replaced by CN or CC, it can become trapped in this alternative form, at least over the 500 ns duration of the simulation. This does imply that the IC state becomes more energetically favourable under the simulation conditions. However, the TD-DFT predicted absorbances for NC and IC are sufficiently close to suggest that both states could coexist in the experimental system. This reinforces the idea that experimental spectra represent a collective average of contributions from both conformations, with their similar vibrational fingerprints making it difficult to distinguish them based on absorbance and Raman data alone. QM simulations also captured key vibrational features, particularly for the CRO-CN and CRO-CC mutants, whose predicted spectra aligned well with experimental frequencies, providing a framework for assigning peaks to specific vibrational modes.

One major insight from this work is the role of the polar network in maintaining the mCherry WT chromophore's electronic and vibrational properties. The mutations at the para-position of the phenolic ring, disrupt this polar network, leading to blue-shifted absorbance maxima. To mitigate these effects, future work should focus on incorporating ncAAs at the meta-position of the phenolic ring. This approach would leave the deprotonated oxygen intact, preserving the polar network and stabilising the chromophores electronic environment. Such modifications could maintain red-shifted absorbance and enhance the chromophores overall stability and functionality, while still introducing novel vibrational signatures.

Moving forward, expanding the computational framework to include more environmental interactions, such as additional residues near the chromophore or more expansive methods/basis-sets, could further refine predictions of vibrational and electronic properties. This would provide deeper insight into the interplay between structure, conformation, and vibrational behaviour, especially for ncAA-modified chromophores. The integration of computational and experimental methods offers a powerful toolkit for designing and characterising chromophores with tailored optical and vibrational properties. These efforts will be critical for advancing applications in molecular imaging and biosensing, ultimately enabling the rational design of chromophores that balance optical performance with stability and tunability.

7 Discussion

7.1 General overview

This thesis explored the development, characterisation and application of genetically encoded Raman-active probes via the integration with advanced spectroscopic techniques, highlighting the cooperation between experimental and computational approaches to advance molecular imaging. The work covered a multidisciplinary pathway, encompassing protein engineering, spectroscopy, molecular dynamics (MD), and quantum mechanics (QM), to deliver a comprehensive framework for designing and characterising chromophores tailored for Raman-based imaging. By combining these methodologies, the thesis achieved several key milestones, linking structural and spectroscopic properties to enable biomolecular applications in cellular imaging.

The selection and characterisation of mRhubarb720 emerged as a cornerstone of this research. Chapter 3 demonstrated that among selected near-IR fluorescent proteins (FPs), mRhubarb720 holds superior attributes for Raman-based imaging due to its red-shifted spectral properties and high molar extinction coefficient. These features make it well-suited for electronic pre-resonant stimulated Raman spectroscopy (epr-SRS), where red-shifted absorbance facilitates resonance conditions under the excitation parameters of the setup. The successful determination of the mRhubarb720 crystal structure provided a valuable foundation for understanding the BV-binding pocket. Chapter 4 extended these findings to cell imaging applications, showcasing the viability of genetically encoded Raman-active probes for microscopy. The incorporation of mRhubarb720 into mammalian and bacterial cells resulted in significantly enhanced epr-SRS signals, particularly in the fingerprint region. The work demonstrated the critical role of red-shifted near-IR FPs in amplifying Raman signals under pre-resonant conditions.

Chapters 5 and 6 advanced the thesis by integrating computational and experimental methodologies to refine chromophore design and explore new vibrational signatures. The work in Chapter 5 evaluated the incorporation of pCNPhe for through-space enhancement within the biologically silent region. While this approach did not yield Raman enhancement in the biologically silent region, it revealed structural challenges associated with

incorporating non-canonical amino acids (ncAAs) into the ligand binding sites. MD simulations elucidated the structural impacts of mutations, identifying mechanisms underlying reduced BV incorporation efficiency. Importantly, the simulations indicated that BV stability within the protein environment was not compromised, shifting the focus toward folding and binding dynamics as key factors influencing incorporation efficiency. QM simulations further demonstrated the role of the WT protein in enhancing the BVs vibrational properties through structural stabilisation, particularly in the 1100–1700 cm^{-1} range. These findings inform future probe designs by highlighting vibrational modes that can be targeted for enhancement. Chapter 6 built upon this foundation by providing detailed insights into chromophore behaviour and vibrational properties via incorporation of ncAA's directly into the chromophore of mCherry. The experimental and computational analyses converged to characterise WT and mutant chromophores, revealing how structural modifications, particularly in disrupting a polar network, influence vibrational and electronic properties. The QM predictions captured key vibrational features, aligning well with experimental data, and provided strategies for maintaining chromophore stability and red-shifted absorbance.

This thesis represents a significant step toward the rational design of genetically encoded Raman-active probes, combining structural insights, advanced spectroscopic techniques, and computational predictions. Probe developments should focus on incorporating electronically coupled vibrational bonds and expanding the genetic code to access a wider array of ncAA's with tailored vibrational properties. The experimental validation of these probes in live-cell imaging, combined with computational predictions of their vibrational spectra, demonstrates the utility of integrating diverse methodologies for advancing spectroscopic techniques. This comprehensive approach provides a blueprint for future research, balancing optical performance with protein/chromophore stability and functionality to unlock new possibilities in molecular imaging and beyond.

7.2 Development of genetically encoded Raman-active probes

The development of genetically encoded Raman-active tags represents a transformative approach in molecular imaging, offering a pathway to overcome the limitations of traditional fluorescence microscopy. The selection of mRhubarb720 as a

Raman-active probe was a critical milestone. Previous studies have demonstrated that GFP-like FPs hold significant potential for Raman spectroscopy due to their unique spectral signatures (Yuan et al. 2018). However, the structural and vibrational properties of near-IR FPs have remained underexplored, particularly in the context of epr-RS signal amplification and cell imaging. The work in Chapter 3 addressed this gap by combining spectral characterisation with crystallographic analysis to establish mRhubarb720 as an optimal candidate for Raman-based imaging.

Key advancements include the determination of mRhubarb720's crystal structure, which revealed the spatial organisation of the BV chromophore within the protein scaffold. This provided insights into how the protein environment stabilises BV's Pr conformation, enhancing its Raman scattering cross-section. Unlike Free-BV, which lacks the necessary absorptive properties for pre-resonant excitation, the BV-bound form in mRhubarb720 exhibited significant vibrational mode amplification via a substantially increased molar extinction coefficient. These findings align with prior studies suggesting that the protein matrix plays a pivotal role in modulating chromophore properties (Spezia et al. 2003; Stepanenko et al. 2011; Ghosh et al. 2022). The identification of specific interactions, such as hydrogen bonding networks (via H285) and the orientation of BV within the binding pocket, has laid the groundwork for targeted mutagenesis to further enhance Raman activity.

The ability to genetically encode Raman-active probes represents a significant advantage over traditional dye-based labelling methods as these engineered FPs offer a streamlined, highly specific approach to tagging cellular structures. Unlike synthetic dyes that often require secondary antibody conjugates or cell fixation for effective labelling, genetically encoded FPs can be directly expressed in living cells or organisms, maintaining physiological conditions, and reducing experimental complexity (Kim et al. 2021). This makes FPs ideal for dynamic, real-time imaging applications. The specificity is further enhanced by the ability of FPs to be fused to any protein of interest without significantly disrupting its function, allowing researchers to track molecular interactions and processes with minimal interference. For example, mRhubarb720s utility as a Raman-active probe was demonstrated in Chapter 4 through its incorporation into Histone 2B (H2B) for chromatin imaging in mammalian cells, where it generated a significantly enhanced SRS signal in the fingerprint region. These results highlight the potential of genetically encoded tags for real-time, *in vivo* imaging.

The development of genetically-encoded Raman-active probes also opens the door for engineering proteins with tailored vibrational properties. As demonstrated in Chapter 6, the structural environment provided by the protein scaffold plays a crucial role in modulating chromophore vibrational modes. The mutagenesis and computational analyses in this thesis revealed that introducing key bonds in the chromophore could enhance Raman scattering cross-sections or introduce novel vibrational signatures. This ability to tune these properties, combined with the incorporation of ncAAs through reprogramming the genetic code, positions these probes as versatile tools for studying complex biological systems in great detail.

7.3 Application of electronic pre-resonant Raman spectroscopy

Epr-RS has emerged as a powerful tool for enhancing vibrational signals, enabling detailed molecular imaging at bond specific sensitivity. This thesis significantly advances the application of epr-RS by leveraging near-IR FPs and genetically-encoded, post-translationally developed chromophores, focusing on their ability to exponentially increase Raman scattering under pre-resonant conditions. Chapter 4 demonstrated that mRhubarb720 outperforms other near-IR FPs due to its red-shifted absorbance and high molar extinction coefficient, which align well with the pre-resonance conditions set by the pump laser at 820 nm. This property enabled significant signal amplification in the fingerprint region of the Raman spectrum with the C=C bond stretching vibrational modes.

The imaging of H2B-mRhubarb720 fusions in mammalian cells demonstrated that epr-SRS could achieve results comparable to traditional fluorescence microscopy, with the added advantage of bond-specific sensitivity. By targeting the chromatin-associated H2B, the Raman-active mRhubarb720 probe provided sufficient signal amplification under pre-resonant conditions to visualise cellular structures. The ability of epr-SRS to distinguish these molecular features underscores its suitability for detailed molecular imaging. A key factor in achieving these results was the condensation of H2B within the chromatin, bringing the mRhubarb720 probes into close proximity. This arrangement amplified the Raman signal, as the concentration of Raman-active probes within the focal volume directly influences signal intensity (Parchaňský et al. 2014). Despite requiring this spatial clustering, the SRS signal was robust and provided clear imaging of the nucleus in HeLa cells, demonstrating the feasibility

of epr-SRS for live-cell studies. These results highlight how genetically encoded Raman-active probes can rival fluorescence microscopy in sensitivity while offering distinct advantages such as sharp spectral resolution and bond-specific detection.

The biologically silent region offers a unique spectral window for Raman imaging, as cellular components typically do not exhibit significant vibrational signals in this range. Chapters 5 and 6 applied epr-SRS to investigate the potential of accessing this region using genetically encoded systems. While the incorporation of pCNPhe did not yield the desired enhancement via through-space enhancement, the detection of nitrile (C≡N) and ethynyl (C≡C) signals in the mCherry chromophores demonstrates the sensitivity of epr-SRS to subtle vibrational features. Compared to fluorescence microscopy, which cannot utilise this spectral range, epr-SRS provides a transformative tool for expanding multiplexing capabilities and reducing non-specific background signals (Wei et al. 2017; Vardaki et al. 2024).

The alignment of pump and Stokes laser wavelengths with the chromophore's pre-resonant excitation range is critical for optimising epr-RS signal amplification. The near-IR FP work in this thesis employed an in-house setup using an 820 nm pump wavelength, which matched well with the absorption maxima of mRhubarb720. This alignment facilitated strong pre-resonance enhancement and clear Raman signal detection in the fingerprint region. However, the spectral range of the other near-IR FPs, emiRFP670 and miRFP670nano3, varies compared to mRhubarb720, requiring adjustments in laser configurations for optimal performance. Comparative work using external setups, such as the one used at Konstanz University (with a 640 nm pump wavelength), highlighted the importance of alternative pump-Stokes configurations. This setup allowed for more precise tuning of excitation conditions for far-red FPs, improving the Raman signal for the mCherry (by 14-fold increase in signal compared to the in-house set-up) with absorption maxima closer to the visible spectrum.

7.4 The impact of computational approaches

The integration of computational approaches in this thesis has been pivotal in providing detailed insights into the structural and electronic properties of chromophores, all while complementing the experimental findings. The work in Chapters 5 and 6 highlights how MD simulations and QM calculations were utilised to explain experimental results,

validate hypotheses, and inform future designs. Using computational tools as a secondary step demonstrated their value in uncovering details about chromophore behaviour and interactions that were not immediately evident from experimental data alone.

MD simulations played a critical role in understanding the structural dynamics of chromophores and their interactions with the protein scaffold. In Chapter 5, MD was used to evaluate the structural impacts of incorporating pCNPhe into mRhubarb720. The simulations revealed that while pCNPhe did not destabilise BV within the protein, therefore, the reduced BV incorporation efficiency was more likely due to either protein folding mechanisms and/or binding dynamics. Additionally, MD simulations shed light on the distinct structural characteristics of apo-proteins, emphasising the role of BV in maintaining the WT protein conformation. For the mCherry variants explored in Chapter 6, these simulations highlighted the isomerisation of the F65 backbone that forms part of the delocalised bond system in the chromophore. Furthermore, the modified chromophores displayed an inability to return to the native conformation which is stabilised in WT. Another key finding of these particular simulations was the loss of the polar network for the mCherry mutants, with previous studies showing that the deprotonated oxygen on the mCherry WT chromophore is involved with two hydrogen bonds (Shu et al. 2006; Pletnev et al. 2010; Mukherjee et al. 2023). This thesis confirms that the deprotonated oxygen in WT is involved with two hydrogen bond interactions, with S146 and a water molecule, throughout the simulations. The simulations for the mutants exhibited a lack of these interactions as S146 is forced away from the internal environment of the β-barrel, as well as the water molecule not being present after equilibration of the system. This was due to the placement of the modification at the para-position of the phenolic ring which occupies a larger area than the replaced deprotonated oxygen and therefore, forces a localised shift of the placement of these interacting molecules. These MD simulations were critical in providing the starting structures for the QM calculations, specifically for the mCherry calculations as the isomerised conformations and loss of the polar network was integrated into the analysis.

In Chapter 5, QM simulations provided critical insights into the vibrational modes of BV-bound and Free-BV, revealing how the protein-bound environment enhances Raman-active modes. The ability to predict vibrational features in the 1100–1700 cm^{-1} range allowed for the identification of spectral regions that could be targeted for enhancement in future experiments. In Chapter 6, the QM calculations also supported the experimental

detection of WT, CRO-CN and CRO-CC signals in both the fingerprint and biologically silent region, with predicted and experimental Raman peaks closely aligning. The mCherry QM analysis did however contradict the MD simulations as the predicted absorbances from the TD-DFT suggest there could be a mixed population of native and isomerised conformations in the mutants, whereas the MD simulations presented an inability to revert back to the native conformation for CRO-CN and CRO-CC. One of the most significant insights from the QM analysis was the role of the polar network in maintaining the electronic and vibrational properties of the mCherry chromophore, with the lack of hydrogen bonding at the para-position significantly blue-shifting the absorbances of the mutants, which is also observed in the experimental data. Identifying how mutations disrupted this network allowed for the simulations to provide a clear pathway for future designs. For example, incorporating ncAAs at alternative meta-/ortho-positions to preserve the polar network could maintain red-shifted absorbance while introducing novel vibrational signatures.

A defining feature of this thesis was the decision to conduct experimental work first, followed by computational analyses to interpret the results. This approach allowed for the generation of robust experimental data, which served as a foundation for validating and refining computational models. While this iterative approach proved effective, it also highlighted inefficiencies in the workflow. Conducting experimental work first required substantial resources and time, particularly for variants that did not yield promising results. In future studies, reversing this workflow by performing computational analyses first could streamline the process. For example, MD simulations could identify mutations that are likely to not disturb the stability of the protein scaffold or chromophore before experimental validation, reducing the number of variants that need to be synthesised and tested. Similarly, QM calculations could predict spectral properties and guide the selection of chromophore modifications, prioritising those with the highest potential for Raman enhancement.

This thesis has demonstrated how each methodology can inform and enhance the other by emphasising the complementarity of experimental and computational approaches. The insights gained through this process not only advance the understanding of chromophore behaviour but also provide a blueprint for optimising workflows in future research. The ability to combine experimental robustness with computational precision

holds great potential for the rational design of molecular imaging tools, paving the way for more efficient and effective development of Raman-active probes.

7.5 Future work

The findings presented in this thesis underscore the transformative potential of epr-RS for molecular imaging, while also highlighting key challenges and opportunities for further advancement. One major limitation identified was the inability of pCNPhe to enhance Raman signals in the biologically silent region via through-space enhancement. This outcome emphasises the challenges associated with incorporating Raman-active targets outside of the chromophore, which often lack the structural integration necessary for consistent vibrational coupling. Future efforts should therefore prioritise covalent integration of vibrational bonds into chromophores, as this approach is more likely to achieve robust enhancement under pre-resonant conditions. For mCherry, a promising avenue involves introducing meta-/ortho-position ncAA tyrosine derivatives that preserve the deprotonated oxygen in the chromophore. This preservation would maintain the chromophore's polar network, stabilising its electronic environment and ensuring far-red absorption. This strategy may require targeting mutational sites within the chromophore environment to optimise chromophore orientation and interactions, guided by structural and computational analyses. For example, the incorporation of 3-NO_2 into mCherry chromophores was found to inhibit protein maturation, likely due to steric clashes caused by the bulky nitro-group. Altering the chromophore environment to provide additional space could enable the incorporation of such derivatives, as well as other nitrile and ethynyl groups, broadening the range of usable Raman-active tags.

An exciting frontier for future work is the development of multiplexing approaches using multiple genetically encoded Raman-active tags. For instance, introducing two Raman-active proteins, such as TagRFP and mCherry, could enable the spatial tracking of interacting proteins. While fluorescence microscopy will induce fluorescence resonance energy transfer (FRET) when these particular FPs are within 57 Å of each other, FRET inherently skews fluorescence signals, complicating spatial tracking (Wu and Brand 1994; Shcherbo et al. 2009). However, the Raman-active FPs should feature distinct vibrational peaks due to slight differences in chromophores structure and environment, enabling simultaneous detection of

multiple proteins without overlap. This represents a key advantage over fluorescence-based methods, expanding the capabilities of molecular imaging for complex cellular systems.

Increasing the repertoire of genetically-encoded Raman probes requires careful consideration of the accessible spectral range for each epr-RS setup. As demonstrated in this thesis, the choice of pump wavelength (e.g. 640 nm vs. 820 nm) significantly impacts the signal intensity of chromophores, such as mCherry WT. Developing chromophores with tailored excitation properties or adapting the pump and Stokes lasers to accommodate more blue-shifted near-IR FPs could enhance the flexibility of epr-SRS. However, such adaptations must balance increased spectral coverage with the risk of photobleaching, a persistent challenge for blue-shifted chromophores in standard fluorescence microscopy (Pruccoli et al. 2023) Overcoming these limitations could greatly broaden the potential of epr-SRS for use with a wider range of Raman-active proteins.

Future workflows should incorporate pre-experimental screening using MD and QM simulations to prioritise chromophore modifications and protein variants with the highest likelihood of success. This approach would streamline the experimental process, conserving time and resources by focusing efforts on promising candidates. MD simulations could predict the effects of mutations on protein/chromophore stability, while QM calculations could refine spectral and vibrational predictions to guide experimental design. Expanding computational frameworks to model additional residues near the chromophore could yield deeper insights into the interplay between structure, environment, and vibrational behaviour. Benchmarking these simulations with more complex methods and basis sets (e.g. B3LYP/6-311G++(d,p)) would further enhance their predictive power.

Ultimately, the genetically encoded nature of Raman-active probes marks a paradigm shift in molecular imaging, allowing researchers to move beyond traditional dyes and fluorescent microscopy, towards a more precise, multiplexed, and dynamic approach. Expanding the genetic code to introduce novel ncAAs with unique vibrational signatures, or engineering FPs with even greater Raman signal enhancement, represents a key area for future innovation. These advancements will solidify the role of genetically encoded Raman-active probes as essential tools for next-generation molecular imaging.

8 Bibliography

Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B. and Lindahl, E. 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2, pp. 19–25. doi: 10.1016/j.softx.2015.06.001.

Adamo, C. and Jacquemin, D. 2013. The calculations of excited-state properties with Time-Dependent Density Functional Theory. *Chem. Soc. Rev.* 42(3), pp. 845–856. doi: 10.1039/C2CS35394F.

Addison, K., Heisler, I.A., Conyard, J., Dixon, T., Bulman Page, P.C. and Meech, S.R. 2013. Ultrafast excited state dynamics of the green fluorescent protein chromophore and its kindling fluorescent protein analogue. *Faraday Discussions* 163, p. 277. doi: 10.1039/c3fd00019b.

Ahsan, H. 2013. 3-Nitrotyrosine: A biomarker of nitrogen free radical species modified proteins in systemic autoimmunogenic conditions. *Human Immunology* 74(10), pp. 1392–1399. doi: 10.1016/j.humimm.2013.06.009.

Ansari, A.M. et al. 2016. Cellular GFP Toxicity and Immunogenicity: Potential Confounders in in Vivo Cell Tracking Experiments. *Stem Cell Reviews and Reports* 12(5), pp. 553–559. doi: 10.1007/s12015-016-9670-8.

Ansbacher, T., Srivastava, H.K., Stein, T., Baer, R., Merkx, M. and Shurki, A. 2012. Calculation of transition dipole moment in fluorescent proteins—towards efficient energy transfer. *Physical Chemistry Chemical Physics* 14(12), p. 4109. doi: 10.1039/c2cp23351g.

Atkins, P. 2022. *Concepts in Physical Chemistry*. Royal Society of Chemistry. doi: 10.1039/9781837674244.

Auldridge, M.E. and Forest, K.T. 2011. Bacterial phytochromes: More than meets the light. *Critical Reviews in Biochemistry and Molecular Biology* 46(1), pp. 67–88. doi: 10.3109/10409238.2010.546389.

Badar, M.S., Shamsi, S., Ahmed, J. and Alam, Md.A. 2022. Molecular Dynamics Simulations: Concept, Methods, and Applications. pp. 131–151. doi: 10.1007/978-3-030-94651-7_7.

Bagheri, M. and Komsa, H.-P. 2023. High-throughput computation of Raman spectra from first principles. *Scientific Data* 10(1), p. 80. doi: 10.1038/s41597-023-01988-5.

Baloban, M., Shcherbakova, D.M., Pletnev, S., Pletnev, V.Z., Lagarias, J.C. and Verkhusha, V. V. 2017. Designing brighter near-infrared fluorescent proteins: insights from structural and biochemical studies. *Chemical Science* 8(6), pp. 4546–4557. doi: 10.1039/C7SC00855D.

Barros, E.B. and Dresselhaus, M.S. 2014. Theory of Raman enhancement by two-dimensional materials: Applications for graphene-enhanced Raman spectroscopy. *Physical Review B* 90(3), p. 035443. doi: 10.1103/PhysRevB.90.035443.

Bartoschek, M.D., Ugur, E., Nguyen, T.-A., Rodschinka, G., Wierer, M., Lang, K. and Bultmann, S. 2021. Identification of permissive amber suppression sites for efficient non-canonical amino acid incorporation in mammalian cells. *Nucleic Acids Research* 49(11), pp. e62–e62. doi: 10.1093/nar/gkab132.

Bechstedt, F. 2015. Exchange and Correlation. pp. 29–47. doi: 10.1007/978-3-662-44593-8_3.

Becke, A.D. 1993. Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics* 98(7), pp. 5648–5652. doi: 10.1063/1.464913.

Bellini, D. and Papiz, M.Z. 2012. Dimerization properties of the *Rp* BphP2 chromophore-binding domain crystallized by homologue-directed mutagenesis. *Acta Crystallographica Section D Biological Crystallography* 68(8), pp. 1058–1066. doi: 10.1107/S0907444912020537.

Belyy, A., Merino, F., Sitsel, O. and Raunser, S. 2020. Structure of the Lifeact–F-actin complex. *PLOS Biology* 18(11), p. e3000925. doi: 10.1371/journal.pbio.3000925.

Beyer, J.N. et al. 2020. Overcoming Near-Cognate Suppression in a Release Factor 1-Deficient Host with an Improved Nitro-Tyrosine tRNA Synthetase. *Journal of Molecular Biology* 432(16), pp. 4690–4704. doi: 10.1016/j.jmb.2020.06.014.

Binkley, J.S., Pople, J.A. and Hehre, W.J. 1980. Self-consistent molecular orbital methods. 21. Small split-valence basis sets for first-row elements. *Journal of the American Chemical Society* 102(3), pp. 939–947. doi: 10.1021/ja00523a008.

Bjelkmar, P., Larsson, P., Cuendet, M.A., Hess, B. and Lindahl, E. 2010. Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *Journal of Chemical Theory and Computation* 6(2), pp. 459–466. doi: 10.1021/ct900549r.

Blacker, T.S., Mann, Z.F., Gale, J.E., Ziegler, M., Bain, A.J., Szabadkai, G. and Duchen, M.R. 2014. Separating NADH and NADPH fluorescence in live cells and tissues using FLIM. *Nature Communications* 5(1), p. 3936. doi: 10.1038/ncomms4936.

Bogaard, M.P. and Haines, R. 1980. Raman intensities and cartesian polarizability derivatives. *Molecular Physics* 41(6), pp. 1281–1289. doi: 10.1080/00268978000103541.

Bohr, N. 1913. I. *On the constitution of atoms and molecules*. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 26(151), pp. 1–25. doi: 10.1080/14786441308634955.

Born, M. and Oppenheimer, R. 1927. Zur Quantentheorie der Molekeln. *Annalen der Physik* 389(20), pp. 457–484. doi: 10.1002/andp.19273892002.

Brasselet, S. 2011. Polarization-resolved nonlinear microscopy: application to structural molecular and biological imaging. *Advances in Optics and Photonics* 3(3), p. 205. doi: 10.1364/AOP.3.000205.

Brooks, B.R. et al. 2009. CHARMM: The Biomolecular Simulation Program. *Journal of Computational Chemistry* 30(10), pp. 1545–1614. Available at: <Go to ISI>://WOS:000267269600001.

Caligiuri, V., Nucera, A., Patra, A., Castriota, M. and De Luca, A. 2024. Raman Scattering Enhancement through Pseudo-Cavity Modes. *Nanomaterials* 14(10), p. 875. doi: 10.3390/nano14100875.

Campanella, B., Palleschi, V. and Legnaioli, S. 2021. Introduction to vibrational spectroscopies. *ChemTexts* 7(1), p. 5. doi: 10.1007/s40828-020-00129-4.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W. and Prasher, D.C. 1994. Green Fluorescent Protein as a Marker for Gene Expression. *Science* 263(5148), pp. 802–805. doi: 10.1126/science.8303295.

Chen, C. et al. 2023. Structural Characterization of Fluorescent Proteins Using Tunable Femtosecond Stimulated Raman Spectroscopy. *International Journal of Molecular Sciences* 24(15), p. 11991. doi: 10.3390/ijms241511991.

Chen, R., Jäättelä, M. and Liu, B. 2020. Lysosome as a Central Hub for Rewiring PH Homeostasis in Tumors. *Cancers* 12(9), p. 2437. doi: 10.3390/cancers12092437.

Cheng, J.-X., Book, L.D. and Xie, X.S. 2001. Polarization coherent anti-Stokes Raman scattering microscopy. *Optics Letters* 26(17), p. 1341. doi: 10.1364/OL.26.001341.

Choudhary, V., Bhatt, A., Dash, D. and Sharma, N. 2019. DFT calculations on molecular structures, HOMO–LUMO study, reactivity descriptors and spectral analyses of newly synthesized diorganotin(IV) 2-chloridophenylacetohydroxamate complexes. *Journal of Computational Chemistry* 40(27), pp. 2354–2363. doi: 10.1002/jcc.26012.

Chudakov, D.M., Matz, M. V., Lukyanov, S. and Lukyanov, K.A. 2010. Fluorescent Proteins and Their Applications in Imaging Living Cells and Tissues. *Physiological Reviews* 90(3), pp. 1103–1163. doi: 10.1152/physrev.00038.2009.

Clark, T., Chandrasekhar, J., Spitznagel, G.W. and Schleyer, P.V.R. 1983a. Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements, Li–F. *Journal of Computational Chemistry* 4(3), pp. 294–301. doi: 10.1002/jcc.540040303.

Clark, T., Chandrasekhar, J., Spitznagel, G.W. and Schleyer, P.V.R. 1983b. Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements, Li–F. *Journal of Computational Chemistry* 4(3), pp. 294–301. doi: 10.1002/jcc.540040303.

Cranfill, P.J. et al. 2016. Quantitative assessment of fluorescent proteins. *Nature Methods* 13(7), pp. 557–562. doi: 10.1038/nmeth.3891.

Czamara, K., Adamczyk, A., Stojak, M., Radwan, B. and Baranska, M. 2021. Astaxanthin as a new Raman probe for biosensing of specific subcellular lipidic structures: can we detect

lipids in cells under resonance conditions? *Cellular and Molecular Life Sciences* 78(7), pp. 3477–3484. doi: 10.1007/s00018-020-03718-1.

Dammeyer, T. and Frankenberg-Dinkel, N. 2008. Function and distribution of bilinbiosynthesis enzymes in photosynthetic organisms. *Photochemical & Photobiological Sciences* 7(10), pp. 1121–1130. doi: 10.1039/b807209b.

Datta, R., Heaster, T.M., Sharick, J.T., Gillette, A.A. and Skala, M.C. 2020. Fluorescence lifetime imaging microscopy: fundamentals and advances in instrumentation, analysis, and applications. *Journal of Biomedical Optics* 25(07), p. 1. doi: 10.1117/1.JBO.25.7.071203.

Dennington, R., Keith, T.A. and Millam, J.M. 2016. GaussView, Version 6. *Semichem Inc., Shawnee Mission*.

Dezhurov, S. V., Volkova, I.Y. and Wakstein, M.S. 2011. FRET-Based Biosensor for Oleic Acid in Nanomolar Range with Quantum Dots As an Energy Donor. *Bioconjugate Chemistry* 22(3), pp. 338–345. doi: 10.1021/bc100133u.

Dietzek, B., Cialla, D., Schmitt, M. and Popp, J. 2018. Introduction to the Fundamentals of Raman Spectroscopy. pp. 47–68. doi: 10.1007/978-3-319-75380-5_3.

Ditchfield, R., Hehre, W.J. and Pople, J.A. 1971. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *The Journal of Chemical Physics* 54(2), pp. 724–728. doi: 10.1063/1.1674902.

Dunning, Thom.H. and Hay, P.J. 1977. Gaussian Basis Sets for Molecular Calculations. In: *Methods of Electronic Structure Theory*. Boston, MA: Springer US, pp. 1–27. doi: 10.1007/978-1-4757-0887-5_1.

Dyall, K.G. 2016. Relativistic double-zeta, triple-zeta, and quadruple-zeta basis sets for the light elements H–Ar. *Theoretical Chemistry Accounts* 135(5), p. 128. doi: 10.1007/s00214-016-1884-y.

Edwards, H.G.M. 2006. Spectra–Structure Correlations in <scp>R</scp> aman Spectroscopy. In: Chalmers, J. M. ed. *Handbook of Vibrational Spectroscopy*. Wiley. doi: 10.1002/0470027320.s4103.

Einstein, A. 1905. Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. *Annalen der Physik* 322(6), pp. 132–148. doi: 10.1002/andp.19053220607.

El-Saady, A.A., Roushdy, N., Farag, A.A.M., El-Nahass, M.M. and Abdel Basset, D.M. 2023. Exploring the molecular spectroscopic and electronic characterization of nanocrystalline Metal-free phthalocyanine: a DFT investigation. *Optical and Quantum Electronics* 55(7), p. 662. doi: 10.1007/s11082-023-04877-8.

Escudero, D., Laurent, A.D. and Jacquemin, D. 2017. Time-Dependent Density Functional Theory: A Tool to Explore Excited States. In: *Handbook of Computational Chemistry*. Cham: Springer International Publishing, pp. 927–961. doi: 10.1007/978-3-319-27282-5_43.

Eskelinen, E.-L. 2006. Roles of LAMP-1 and LAMP-2 in lysosome biogenesis and autophagy. *Molecular Aspects of Medicine* 27(5–6), pp. 495–502. doi: 10.1016/j.mam.2006.08.005.

Eskelinen, E.-L. and Saftig, P. 2009. Autophagy: A lysosomal degradation pathway with a central role in health and disease. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1793(4), pp. 664–673. doi: 10.1016/j.bbamcr.2008.07.014.

Evans, C.L. and Xie, X.S. 2008. Coherent Anti-Stokes Raman Scattering Microscopy: Chemical Imaging for Biology and Medicine. *Annual Review of Analytical Chemistry* 1(1), pp. 883–909. doi: 10.1146/annurev.anchem.1.031207.112754.

Eynaudi, A., Díaz-Castro, F., Bórquez, J.C., Bravo-Sagua, R., Parra, V. and Troncoso, R. 2021. Differential Effects of Oleic and Palmitic Acids on Lipid Droplet-Mitochondria Interaction in the Hepatic Cell Line HepG2. *Frontiers in Nutrition* 8. doi: 10.3389/fnut.2021.775382.

Filippi, C., Buda, F., Guidoni, L. and Sinicropi, A. 2012. Bathochromic Shift in Green Fluorescent Protein: A Puzzle for QM/MM Approaches. *Journal of Chemical Theory and Computation* 8(1), pp. 112–124. doi: 10.1021/ct200704k.

Fleming, P. 2024. *QUANTUM CHEMISTRY WITH APPLICATIONS IN SPECTROSCOPY*.

FRANKLIN, K.A. and WHITELAM, G.C. 2005. Phytochromes and Shade-avoidance Responses in Plants. *Annals of Botany* 96(2), pp. 169–175. doi: 10.1093/aob/mci165.

Gfeller, D., Michielin, O. and Zoete, V. 2012. Expanding molecular modeling and design tools to non-natural sidechains. *Journal of Computational Chemistry* 33(18), pp. 1525–1535. doi: 10.1002/jcc.22982.

Ghosh, S., Mondal, S., Yadav, K., Aggarwal, S., Schaefer, W.F., Narayana, C. and Ramaswamy, S. 2021. Structural heterogeneity in biliverdin modulates spectral properties of Sandercyanin fluorescent protein. doi: 10.1101/2021.04.02.438172.

Ghosh, S., Mondal, S., Yadav, K., Aggarwal, S., Schaefer, W.F., Narayana, C. and Subramanian, R. 2022a. Modulation of biliverdin dynamics and spectral properties by Sandercyanin. *RSC Advances* 12(31), pp. 20296–20304. doi: 10.1039/D2RA02880H.

Ghosh, S., Mondal, S., Yadav, K., Aggarwal, S., Schaefer, W.F., Narayana, C. and Subramanian, R. 2022b. Modulation of biliverdin dynamics and spectral properties by Sandercyanin. *RSC Advances* 12(31), pp. 20296–20304. doi: 10.1039/D2RA02880H.

Gilmore, J. and McKenzie, R.H. 2006. Quantum dynamics of electronic excitations in biomolecular chromophores: role of the protein environment and solvent.

Goletto, L., Kjønstad, E.F., Folkestad, S.D., Høyvik, I.-M. and Koch, H. 2021. Linear-Scaling Implementation of Multilevel Hartree–Fock Theory. *Journal of Chemical Theory and Computation* 17(12), pp. 7416–7427. doi: 10.1021/acs.jctc.1c00299.

Griffiths, D.J. and Schroeter, D.F. 2018. *Introduction to Quantum Mechanics*. Cambridge University Press. doi: 10.1017/9781316995433.

Grigorenko, B.L., Krylov, A.I. and Nemukhin, A. V. 2017. Molecular Modeling Clarifies the Mechanism of Chromophore Maturation in the Green Fluorescent Protein. *Journal of the American Chemical Society* 139(30), pp. 10239–10249. Available at: <Go to ISI>://WOS:000407089500018.

Gross, L.A., Baird, G.S., Hoffman, R.C., Baldridge, K.K. and Tsien, R.Y. 2000. The structure of the chromophore within DsRed, a red fluorescent protein from coral. *Proceedings of the National Academy of Sciences* 97(22), pp. 11990–11995. doi: 10.1073/pnas.97.22.11990.

Hariharan, P.C. and Pople, J.A. 1973. The influence of polarization functions on molecular orbital hydrogenation energies. *Theoretica Chimica Acta* 28(3), pp. 213–222. doi: 10.1007/BF00533485.

Hartley, A.M., Worthy, H.L., Reddington, S.C., Rizkallah, P.J. and Jones, D.D. 2016. Molecular basis for functional switching of GFP by two disparate non-native post-translational modifications of a phenyl azide reaction handle. *Chemical Science* 7(10), pp. 6484–6491. doi: 10.1039/C6SC00944A.

He, K., Xu, Z., Zhang, X., Li, Q. and Wang, F. 2023. Influence of the apparatus on the intensity ratio of the measured Raman peaks. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 289, p. 122249. doi: 10.1016/j.saa.2022.122249.

Hehre, W.J., Stewart, R.F. and Pople, J.A. 1969. Self-Consistent Molecular-Orbital Methods. I. Use of Gaussian Expansions of Slater-Type Atomic Orbitals. *The Journal of Chemical Physics* 51(6), pp. 2657–2664. doi: 10.1063/1.1672392.

Heisenberg, W. 1927. □ber den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift f□r Physik* 43(3–4), pp. 172–198. doi: 10.1007/BF01397280.

Herbert, J.M. 2024. Visualizing and characterizing excited states from time-dependent density functional theory. *Physical Chemistry Chemical Physics* 26(5), pp. 3755–3794. doi: 10.1039/D3CP04226J.

Hickey, S.M. et al. 2021. Fluorescence Microscopy—An Outline of Hardware, Biological Handling, and Fluorophore Considerations. *Cells* 11(1), p. 35. doi: 10.3390/cells11010035.

Hohenberg, P. and Kohn, W. 1964. Inhomogeneous Electron Gas. *Physical Review* 136(3B), pp. B864–B871. doi: 10.1103/PhysRev.136.B864.

Housecroft, C.E. and Sharpe, A.G. 2018. *Inorganic chemistry*. Fifth. Harlow, England: Pearson Education.

Hu, Z., Wang, X., Wang, W., Zhang, Z., Gao, H. and Mao, Y. 2015. Raman spectroscopy for detecting supported planar lipid bilayers composed of ganglioside-GM1/sphingomyelin/cholesterol in the presence of amyloid-β. *Physical Chemistry Chemical Physics* 17(35), pp. 22711–22720. doi: 10.1039/C5CP02366A.

Huang, Y., Rong, C., Zhang, R. and Liu, S. 2017. Evaluating frontier orbital energy and HOMO/LUMO gap with descriptors from density functional reactivity theory. *Journal of Molecular Modeling* 23(1), p. 3. doi: 10.1007/s00894-016-3175-x.

Hughes, J., Lamparter, T., Mittmann, F., Hartmann, E., Gärtner, W., Wilde, A. and Börner, T. 1997. A prokaryotic phytochrome. *Nature* 386(6626), pp. 663–663. doi: 10.1038/386663a0.

Jacquemin, D., Mennucci, B. and Adamo, C. 2011. Excited-state calculations with TD-DFT: from benchmarks to simulations in complex environments. *Physical Chemistry Chemical Physics* 13(38), p. 16987. doi: 10.1039/c1cp22144b.

Jones, R.R., Hooper, D.C., Zhang, L., Wolverson, D. and Valev, V.K. 2019. Raman Techniques: Fundamentals and Frontiers. *Nanoscale Research Letters* 14(1), p. 231. doi: 10.1186/s11671-019-3039-2.

Katchalski, E. 1966. Novel Techniques for the Synthesis of Linear and Cyclic Peptides. In: *Hypotensive Peptides*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 55–62. doi: 10.1007/978-3-642-94965-4_7.

Kaufmann, T. et al. 2020. Direct measurement of protein–protein interactions by FLIM-FRET at UV laser-induced DNA damage sites in living cells. *Nucleic Acids Research* 48(21), pp. e122–e122. doi: 10.1093/nar/gkaa859.

Keresztury, G. 2006. Raman Spectroscopy: Theory. *Handbook of Vibrational Spectroscopy*.

Ketterson, J.B. 2016. Hartree–Fock and Density Functional Theory. In: *The Physics of Solids*. Oxford University PressOxford, pp. 257–294. doi: 10.1093/acprof:oso/9780198742906.003.0015.

Khrenova, M., Topol, I., Collins, J. and Nemukhin, A. 2015. Estimating Orientation Factors in the FRET Theory of Fluorescent Proteins: The TagRFP-KFP Pair and Beyond. *Biophysical Journal* 108(1), pp. 126–132. doi: 10.1016/j.bpj.2014.11.1859.

Kim, H., Ju, J., Lee, H.N., Chun, H. and Seong, J. 2021. Genetically Encoded Biosensors Based on Fluorescent Proteins. *Sensors* 21(3), p. 795. doi: 10.3390/s21030795.

Kohn, W. and Sham, L.J. 1965. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* 140(4A), pp. A1133–A1138. doi: 10.1103/PhysRev.140.A1133.

Kolesov, B. 2021. Hydrogen Bonds: Raman Spectroscopic Study. *International Journal of Molecular Sciences* 22(10), p. 5380. doi: 10.3390/ijms22105380.

Kruskal, W.H. and Wallis, W.A. 1952. Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association* 47(260), pp. 583–621. doi: 10.1080/01621459.1952.10483441.

Kumarapperuma, I., Tom, I.P., Bandara, S., Montano, S. and Yang, X. 2023. Mode of autophosphorylation in bacteriophytochromes RpBphP2 and RpBphP3. *Photochemical & Photobiological Sciences*. doi: 10.1007/s43630-023-00366-9.

Lai, W.-Q., Chang, Y.-F., Chou, F.-N. and Yang, D.-M. 2022. Portable FRET-Based Biosensor Device for On-Site Lead Detection. *Biosensors* 12(3), p. 157. doi: 10.3390/bios12030157.

Laurent, A.D., Mironov, V.A., Chapagain, P.P., Nemukhin, A. V. and Krylov, A.I. 2012. Exploring Structural and Optical Properties of Fluorescent Proteins by Squeezing: Modeling High-Pressure Effects on the mStrawberry and mCherry Red Fluorescent Proteins. *The Journal of Physical Chemistry B* 116(41), pp. 12426–12440. doi: 10.1021/jp3060944.

Lee, C., Yang, W. and Parr, R.G. 1988a. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B* 37(2), pp. 785–789. doi: 10.1103/PhysRevB.37.785.

Lee, C., Yang, W. and Parr, R.G. 1988b. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B* 37(2), pp. 785–789. doi: 10.1103/PhysRevB.37.785.

Li, H., Zhang, J., Vierstra, R.D. and Li, H. 2010. Quaternary organization of a phytochrome dimer as revealed by cryoelectron microscopy. *Proceedings of the National Academy of Sciences* 107(24), pp. 10872–10877. doi: 10.1073/pnas.1001908107.

Li, J., Li, G., Wang, H. and Wang Deng, X. 2011. Phytochrome Signaling Mechanisms. *The Arabidopsis Book* 9, p. e0148. doi: 10.1199/tab.0148.

Li, S., Li, Y., Yi, R., Liu, L. and Qu, J. 2020. Coherent Anti-Stokes Raman Scattering Microscopy and Its Applications. *Frontiers in Physics* 8. doi: 10.3389/fphy.2020.598420.

Li, X. and Frisch, M.J. 2006. Energy-Represented Direct Inversion in the Iterative Subspace within a Hybrid Geometry Optimization Method. *Journal of Chemical Theory and Computation* 2(3), pp. 835–839. doi: 10.1021/ct050275a.

Lichtman, J.W. and Conchello, J.-A. 2005. Fluorescence microscopy. *Nature Methods* 2(12), pp. 910–919. doi: 10.1038/nmeth817.

Lin-Vien, D., Colthup, N.B., Fateley, W.G. and Grasselli, J.G. 1991. *The Handbook of Infrared and Raman Characteristic Frequencies of Organic Molecules*. Elsevier. doi: 10.1016/B978-0-08-057116-4.50023-7.

Long, D.A. 2002. The Polarizability Tensor. In: *The Raman Effect*. John Wiley & Sons, Ltd, pp. 471–495. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1002/0470845767.ch24.

López, M.F. et al. 2022. Photoinduced reaction mechanisms in prototypical and bathy phytochromes. *Physical Chemistry Chemical Physics* 24(19), pp. 11967–11978. doi: 10.1039/D2CP00020B.

Luo, Z., Xu, X., Lin, D., Qu, J., Lin, F. and Li, J. 2024. Removing non-resonant background of CARS signal with generative adversarial network. *Applied Physics Letters* 124(26). doi: 10.1063/5.0201616.

Lupo Pasini, M., Mehta, K., Yoo, P. and Irle, S. 2023. Two excited-state datasets for quantum chemical UV-vis spectra of organic molecules. *Scientific Data* 10(1), p. 546. doi: 10.1038/s41597-023-02408-4.

Lv, X., Gong, L., Lin, S., Jin, P. and Huang, Z. 2022. Super-resolution stimulated Raman scattering microscopy with the phase-shifted spatial frequency modulation. *Optics Letters* 47(17), p. 4552. doi: 10.1364/OL.463087.

M. J. Frisch et al. 2016. Gaussian 09.

Macaluso, V., Salvadori, G., Cupellini, L. and Mennucci, B. 2021. The structural changes in the signaling mechanism of bacteriophytochromes in solution revealed by a multiscale computational investigation. *Chemical Science* 12(15), pp. 5555–5565. doi: 10.1039/D1SC00186H.

Mackerell, A.D., Feig, M. and Brooks, C.L. 2004. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal of Computational Chemistry* 25(11), pp. 1400–1415. doi: 10.1002/jcc.20065.

Maillard, J., Klehs, K., Rumble, C., Vauthey, E., Heilemann, M. and Fürstenberg, A. 2021. Universal quenching of common fluorescent probes by water and alcohols. *Chemical Science* 12(4), pp. 1352–1362. doi: 10.1039/D0SC05431C.

Marques, M.A.L. and Gross, E.K.U. 2004. TIME-DEPENDENT DENSITY FUNCTIONAL THEORY. *Annual Review of Physical Chemistry* 55(1), pp. 427–455. doi: 10.1146/annurev.physchem.55.091602.094449.

Masia, F., Glen, A., Stephens, P., Borri, P. and Langbein, W. 2013. Quantitative Chemical Imaging and Unsupervised Analysis Using Hyperspectral Coherent Anti-Stokes Raman Scattering Microscopy. *Analytical Chemistry* 85(22), pp. 10820–10828. doi: 10.1021/ac402303g.

Masters, T.A., Marsh, R.J., Armoogum, D.A., Nicolaou, N., Larijani, B. and Bain, A.J. 2013. Restricted State Selection in Fluorescent Protein Förster Resonance Energy Transfer. *Journal of the American Chemical Society* 135(21), pp. 7883–7890. doi: 10.1021/ja312230b.

Matlashov, M.E. et al. 2020a. A set of monomeric near-infrared fluorescent proteins for multicolor imaging across scales. *Nature Communications* 11(1), p. 239. doi: 10.1038/s41467-019-13897-6.

Matlashov, M.E. et al. 2020b. A set of monomeric near-infrared fluorescent proteins for multicolor imaging across scales. *Nature Communications* 11(1), p. 239. doi: 10.1038/s41467-019-13897-6.

Matz, M. V., Fradkov, A.F., Labas, Y.A., Savitsky, A.P., Zaraisky, A.G., Markelov, M.L. and Lukyanov, S.A. 1999. Fluorescent proteins from nonbioluminescent Anthozoa species. *Nature Biotechnology* 17(10), pp. 969–973. doi: 10.1038/13657.

Mennucci, B. 2012. Polarizable continuum model. *WIREs Computational Molecular Science* 2(3), pp. 386–404. doi: 10.1002/wcms.1086.

Miao, Y., Qian, N., Shi, L., Hu, F. and Min, W. 2021. 9-Cyanopyronin probe palette for super-multiplexed vibrational imaging. *Nature Communications* 12(1), p. 4518. doi: 10.1038/s41467-021-24855-6.

Miyake-Stoner, S.J., Miller, A.M., Hammill, J.T., Peeler, J.C., Hess, K.R., Mehl, R.A. and Brewer, S.H. 2009. Probing Protein Folding Using Site-Specifically Encoded Unnatural Amino Acids as FRET Donors with Tryptophan. *Biochemistry* 48(25), pp. 5953–5962. doi: 10.1021/bi900426d.

Modi, V., Donnini, S., Groenhof, G. and Morozov, D. 2019. Protonation of the Biliverdin IXα Chromophore in the Red and Far-Red Photoactive States of a Bacteriophytochrome. *The Journal of Physical Chemistry B* 123(10), pp. 2325–2334. doi: 10.1021/acs.jpcb.9b01117.

Mosselmans, J.F.W. et al. 2009. I18 – the microfocus spectroscopy beamline at the Diamond Light Source. *Journal of Synchrotron Radiation* 16(6), pp. 818–824. doi: 10.1107/S0909049509032282.

Mroginski, M.A., Murgida, D.H. and Hildebrandt, P. 2006. Calculation of Vibrational Spectra of Linear Tetrapyrroles. 4. Methine Bridge C−H Out-of-Plane Modes. *The Journal of Physical Chemistry A* 110(36), pp. 10564–10574. doi: 10.1021/jp063128x.

Mroginski, M.A., Murgida, D.H. and Hildebrandt, P. 2007. The Chromophore Structural Changes during the Photocycle of Phytochrome: A Combined Resonance Raman and Quantum Chemical Approach. *Accounts of Chemical Research* 40(4), pp. 258–266. doi: 10.1021/ar6000523.

Mukherjee, S., Manna, P., Douglas, N., Chapagain, P.P. and Jimenez, R. 2023. Conformational Dynamics of mCherry Variants: A Link between Side-Chain Motions and Fluorescence Brightness. *The Journal of Physical Chemistry B* 127(1), pp. 52–61. doi: 10.1021/acs.jpcb.2c05584.

Murdachaew, G., Varner, M.E., van der Veer, W.E., Gerber, R.B. and Phillips, L.F. 2014. Raman spectroscopy of solutions and interfaces containing nitrogen dioxide, water, and 1,4 dioxane: Evidence for repulsion of surface water by NO2 gas. *The Journal of Chemical Physics* 140(18). doi: 10.1063/1.4874640.

Murshudov, G.N. et al. 2011. *REFMAC* 5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D Biological Crystallography* 67(4), pp. 355–367. doi: 10.1107/S0907444911001314.

Myšková, J., Rybakova, O., Brynda, J., Khoroshyy, P., Bondar, A. and Lazar, J. 2020. Directionality of light absorption and emission in representative fluorescent proteins. *Proceedings of the National Academy of Sciences* 117(51), pp. 32395–32401. doi: 10.1073/pnas.2017379117.

Neugebauer, U., März, A., Henkel, T., Schmitt, M. and Popp, J. 2012. Spectroscopic detection and quantification of heme and heme degradation products. *Analytical and Bioanalytical Chemistry* 404(10), pp. 2819–2829. doi: 10.1007/s00216-012-6288-9.

Nishiwaki, N. 2020. A Walk through Recent Nitro Chemistry Advances. *Molecules* 25(16), p. 3680. doi: 10.3390/molecules25163680.

Oliinyk, O.S., Baloban, M., Clark, C.L., Carey, E., Pletnev, S., Nimmerjahn, A. and Verkhusha, V. V. 2022a. Single-domain near-infrared protein provides a scaffold for antigen-dependent fluorescent nanobodies. *Nature Methods* 19(6), pp. 740–750. doi: 10.1038/s41592-022-01467-6.

Oliinyk, O.S., Baloban, M., Clark, C.L., Carey, E., Pletnev, S., Nimmerjahn, A. and Verkhusha, V. V. 2022b. Single-domain near-infrared protein provides a scaffold for antigen-dependent fluorescent nanobodies. *Nature Methods* 19(6), pp. 740–750. doi: 10.1038/s41592-022-01467-6.

Olzmann, J.A. and Carvalho, P. 2019. Dynamics and functions of lipid droplets. *Nature Reviews Molecular Cell Biology* 20(3), pp. 137–155. doi: 10.1038/s41580-018-0085-z.

Omary, M.A. and Patterson, H.H. 2017. Luminescence, Theory. In: *Encyclopedia of Spectroscopy and Spectrometry*. Elsevier, pp. 636–653. doi: 10.1016/B978-0-12-803224-4.00193-X.

Oudar, J.-L., Smith, R.W. and Shen, Y.R. 1979. Polarization-sensitive coherent anti-Stokes Raman spectroscopy. *Applied Physics Letters* 34(11), pp. 758–760. doi: 10.1063/1.90663.

Ozeki, Y., Dake, F., Kajiyama, S., Fukui, K. and Itoh, K. 2009. Analysis and experimental assessment of the sensitivity of stimulated Raman scattering microscopy. *Optics Express* 17(5), p. 3651. doi: 10.1364/OE.17.003651.

Parchaňský, V., Kapitán, J. and Bouř, P. 2014. Inspecting chiral molecules by Raman optical activity spectroscopy. *RSC Adv.* 4(100), pp. 57125–57136. doi: 10.1039/C4RA10416A.

Pascual-ahuir, J.L., Silla, E. and Tuñon, I. 1994. GEPOL: An improved description of molecular surfaces. III. A new algorithm for the computation of a solvent-excluding surface. *Journal of Computational Chemistry* 15(10), pp. 1127–1138. doi: 10.1002/jcc.540151009.

Pauli, W. 1925. Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren. *Zeitschrift für Physik* 31(1), pp. 765–783. doi: 10.1007/BF02980631.

Pearson, D.A., Lister-James, J., McBride, W.J., Wilson, D.M., Martel, L.J., Civitello, E.R. and Dean, R.T. 1996. Thrombus Imaging Using Technetium-99m-Labeled High-Potency GPIIb/IIIa Receptor Antagonists. Chemistry and Initial Biological Studies. *Journal of Medicinal Chemistry* 39(7), pp. 1372–1382. doi: 10.1021/jm950112e.

Pérez-Jiménez, A.I., Lyu, D., Lu, Z., Liu, G. and Ren, B. 2020. Surface-enhanced Raman spectroscopy: benefits, trade-offs and future developments. *Chemical Science* 11(18), pp. 4563–4577. doi: 10.1039/D0SC00809E.

Planck, M. 1901. Ueber das Gesetz der Energieverteilung im Normalspectrum. *Annalen der Physik* 309(3), pp. 553–563. doi: 10.1002/andp.19013090310.

Pletnev, S. 2022. Crystal structure of near-infrared fluorescent protein miRFP670nano3. doi: 10.2210/pdb7lsc/pdb.

Pletnev, S. [no date]. Near-infrared fluorescent protein with enhanced brightness. *To be published.*

Pletnev, S., Subach, F. V., Dauter, Z., Wlodawer, A. and Verkhusha, V. V. 2010. Understanding Blue-to-Red Conversion in Monomeric Fluorescent Timers and Hydrolytic Degradation of Their Chromophores. *Journal of the American Chemical Society* 132(7), pp. 2243–2253. doi: 10.1021/ja908418r.

Polyakov, I. V., Grigorenko, B.L., Mironov, V.A. and Nemukhin, A. V. 2018. Modeling structure and excitation of biliverdin-binding domains in infrared fluorescent proteins. *Chemical Physics Letters* 710, pp. 59–63. doi: 10.1016/j.cplett.2018.08.068.

Porezag, D. and Pederson, M.R. 1996. Infrared intensities and Raman-scattering activities within density-functional theory. *Physical Review B* 54(11), pp. 7830–7836. doi: 10.1103/PhysRevB.54.7830.

Pronk, S. et al. 2013. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29(7), pp. 845–854. Available at: <Go to ISI>://WOS:000316695700004.

Pruccoli, A., Kocademir, M., Winterhalder, M.J. and Zumbusch, A. 2023. Electronically Preresonant Stimulated Raman Scattering Microscopy of Weakly Fluorescing Chromophores. *The Journal of Physical Chemistry B* 127(27), pp. 6029–6037. doi: 10.1021/acs.jpcb.3c01407.

Qi, Y. et al. 2023. Applications of Raman Spectroscopy in Clinical Medicine.

Qian, N. and Min, W. 2022. Super-multiplexed vibrational probes: Being colorful makes a difference. *Current opinion in chemical biology* 67, p. 102115. doi: 10.1016/j.cbpa.2021.102115.

Raček, T., Schindler, O., Toušek, D., Horský, V., Berka, K., Koča, J. and Svobodová, R. 2020. Atomic Charge Calculator II: web-based tool for the calculation of partial atomic charges. *Nucleic Acids Research* 48(W1), pp. W591–W596. doi: 10.1093/nar/gkaa367.

Ramakrishnan, V. 2015. The Diamond Light Source and the challenges ahead for structural biology: some informal remarks. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 373(2036), p. 20130156. doi: 10.1098/rsta.2013.0156.

Raman, C. V. and Krishnan, K.S. 1928. A New Type of Secondary Radiation. *Nature* 121(3048), pp. 501–502. doi: 10.1038/121501c0.

Reddington, S.C., Driezis, S., Hartley, A.M., Watson, P.D., Rizkallah, P.J. and Jones, D.D. 2015. Genetically encoded phenyl azide photochemistry drives positive and negative functional modulation of a red fluorescent protein. *RSC Advances* 5(95), pp. 77734–77738. doi: 10.1039/C5RA13552D.

Reddington, S.C., Rizkallah, P.J., Watson, P.D., Pearson, R., Tippmann, E.M. and Jones, D.D. 2013. Different Photochemical Events of a Genetically Encoded Phenyl Azide Define and Modulate GFP Fluorescence. *Angewandte Chemie International Edition* 52(23), pp. 5974–5977. doi: 10.1002/anie.201301490.

Reddy, A., Caler, E. V. and Andrews, N.W. 2001. Plasma Membrane Repair Is Mediated by Ca2+-Regulated Exocytosis of Lysosomes. *Cell* 106(2), pp. 157–169. doi: 10.1016/S0092-8674(01)00421-4.

Regan, D., Williams, J., Borri, P. and Langbein, W. 2019. Lipid Bilayer Thickness Measured by Quantitative DIC Reveals Phase Transitions and Effects of Substrate Hydrophilicity. *Langmuir* 35(43), pp. 13805–13814. doi: 10.1021/acs.langmuir.9b02538.

Riek, C. et al. 2016. Stimulated Raman scattering microscopy by Nyquist modulation of a two-branch ultrafast fiber source. *Optics Letters* 41(16), p. 3731. doi: 10.1364/OL.41.003731.

Rigneault, H. and Gachet, D. 2012. Background-free Coherent Raman Imaging: The CARS and SRS Contrast Mechanisms. pp. 347–372. doi: 10.1007/978-3-642-28252-2_12.

Robson, P.R.H., McCormac, A.C., Irvine, A.S. and Smith, H. 1996. Genetic engineering of harvest index in tobacco through overexpression of a phytochrome gene. *Nature Biotechnology* 14(8), pp. 995–998. doi: 10.1038/nbt0896-995.

Rogers, O.C., Johnson, D.M. and Firnberg, E. 2019a. mRhubarb: Engineering of monomeric, red-shifted, and brighter variants of iRFP using structure-guided multi-site mutagenesis. *Scientific Reports* 9(1), p. 15653. doi: 10.1038/s41598-019-52123-7.

Rogers, O.C., Johnson, D.M. and Firnberg, E. 2019b. mRhubarb: Engineering of monomeric, red-shifted, and brighter variants of iRFP using structure-guided multi-site mutagenesis. *Scientific Reports* 9, p. 8. Available at: <Go to ISI>://WOS:000493276600083.

Saftig, P. and Klumperman, J. 2009. Lysosome biogenesis and lysosomal membrane proteins: trafficking meets function. *Nature Reviews Molecular Cell Biology* 10(9), pp. 623–635. doi: 10.1038/nrm2745.

Sahni, V. 2016. Hohenberg–Kohn, Kohn–Sham, and Runge-Gross Density Functional Theories. In: *Quantal Density Functional Theory*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 135–183. doi: 10.1007/978-3-662-49842-2_4.

Salomon-Ferrer, R., Case, D.A. and Walker, R.C. 2013. An overview of the Amber biomolecular simulation package. *Wiley Interdisciplinary Reviews-Computational Molecular Science* 3(2), pp. 198–210. Available at: <Go to ISI>://WOS:000318044900007.

Sands, B.L., Welsh, M.J., Kin, S., Marhatta, R., Hinkle, J.D. and Bayram, S.B. 2007. Raman scattering spectroscopy of liquid nitrogen molecules: An advanced undergraduate physics laboratory experiment. *American Journal of Physics* 75(6), pp. 488–495. doi: 10.1119/1.2721584.

Sasidharan, S., Gosu, V., Tripathi, T. and Saudagar, P. 2023. Molecular Dynamics Simulation to Study Protein Conformation and Ligand Interaction. In: *Protein Folding Dynamics and Stability*. Singapore: Springer Nature Singapore, pp. 107–127. doi: 10.1007/978-981-99-2079-2_6.

Sato, E.T. and Martinho, H. 2018. First-principles calculations of Raman vibrational modes in the fingerprint region for connective tissue. *Biomedical Optics Express* 9(4), p. 1728. doi: 10.1364/BOE.9.001728.

Schmidt, E.L., Ou, Z., Ximendes, E., Cui, H., Keck, C.H.C., Jaque, D. and Hong, G. 2024. Near-infrared II fluorescence imaging. *Nature Reviews Methods Primers* 4(1), p. 23. doi: 10.1038/s43586-024-00301-x.

Schmidt, M.W., Ivanic, J. and Ruedenberg, K. 2014. Covalent bonds are created by the drive of electron waves to lower their kinetic energy through expansion. *The Journal of Chemical Physics* 140(20). doi: 10.1063/1.4875735.

Schneider, C.A., Rasband, W.S. and Eliceiri, K.W. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* 9(7), pp. 671–675. doi: 10.1038/nmeth.2089.

Schrödinger, E. 1926. Quantisierung als Eigenwertproblem. *Annalen der Physik* 386(18), pp. 109–139. doi: 10.1002/andp.19263861802.

Schrödinger, LLC. 2015. The PyMOL Molecular Graphics System.

Shaner, N.C., Campbell, R.E., Steinbach, P.A., Giepmans, B.N.G., Palmer, A.E. and Tsien, R.Y. 2004a. Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nature Biotechnology* 22(12), pp. 1567–1572. doi: 10.1038/nbt1037.

Shaner, N.C., Campbell, R.E., Steinbach, P.A., Giepmans, B.N.G., Palmer, A.E. and Tsien, R.Y. 2004b. Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp red fluorescent protein. *Nature Biotechnology* 22(12), pp. 1567–1572. Available at: <Go to ISI>://WOS:000225638600028.

Shcherbo, D. et al. 2009. Practical and reliable FRET/FLIM pair of fluorescent proteins. *BMC Biotechnology* 9(1), p. 24. doi: 10.1186/1472-6750-9-24.

Shi, L., Xiong, H., Shen, Y., Long, R., Wei, L. and Min, W. 2018. Electronic Resonant Stimulated Raman Scattering Micro-Spectroscopy. *The Journal of Physical Chemistry B* 122(39), pp. 9218–9224. doi: 10.1021/acs.jpcb.8b07037.

Shimomura, O. 1979. STRUCTURE OF THE CHROMOPHORE OF AEQUOREA GREEN FLUORESCENT PROTEIN. *Febs Letters* 104(2), pp. 220–222. Available at: <Go to ISI>://WOS:A1979HG73400003.

Shimomura, O., Johnson, F.H. and Saiga, Y. 1962. EXTRACTION, PURIFICATION AND PROPERTIES OF AEQUORIN, A BIOLUMINESCENT PROTEIN FROM LUMINOUS HYDROMEDUSAN, AEQUOREA. *Journal of Cellular and Comparative Physiology* 59(3), pp. 223-. Available at: <Go to ISI>://WOS:A19624094B00012.

Shipp, D.W., Sinjab, F. and Notingher, I. 2017. Raman spectroscopy: techniques and applications in the life sciences. *Advances in Optics and Photonics* 9(2), p. 315. doi: 10.1364/AOP.9.000315.

Shu, X., Shaner, N.C., Yarbrough, C.A., Tsien, R.Y. and Remington, S.J. 2006a. Novel Chromophores and Buried Charges Control Color in mFruits ′. *Biochemistry* 45(32), pp. 9639–9647. doi: 10.1021/bi060773l.

Shu, X., Shaner, N.C., Yarbrough, C.A., Tsien, R.Y. and Remington, S.J. 2006b. Novel Chromophores and Buried Charges Control Color in mFruits ′. *Biochemistry* 45(32), pp. 9639–9647. doi: 10.1021/bi060773l.

Shu, X., Shaner, N.C., Yarbrough, C.A., Tsien, R.Y. and Remington, S.J. 2006c. Novel Chromophores and Buried Charges Control Color in mFruits ′. *Biochemistry* 45(32), pp. 9639–9647. doi: 10.1021/bi060773l.

Shu, X.K., Shaner, N.C., Yarbrough, C.A., Tsien, R.Y. and Remington, S.J. 2006d. Novel chromophores and buried charges control color in mFruits. *Biochemistry* 45(32), pp. 9639–9647. Available at: <Go to ISI>://WOS:000239596600001.

Shukla, R. and Tripathi, T. 2020. Molecular Dynamics Simulation of Protein and Protein–Ligand Complexes. In: *Computer-Aided Drug Design*. Singapore: Springer Singapore, pp. 133–161. doi: 10.1007/978-981-15-6815-2_7.

Sinha, P., Boesch, S.E., Gu, C., Wheeler, R.A. and Wilson, A.K. 2004. Harmonic Vibrational Frequencies: Scaling Factors for HF, B3LYP, and MP2 Methods in Combination with Correlation Consistent Basis Sets. *The Journal of Physical Chemistry A* 108(42), pp. 9213–9217. doi: 10.1021/jp048233q.

Smith, R., Wright, K.L. and Ashton, L. 2016. Raman spectroscopy: an evolving technique for live cell studies. *The Analyst* 141(12), pp. 3590–3600. doi: 10.1039/C6AN00152A.

Sousa da Silva, A.W. and Vranken, W.F. 2012. ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Research Notes* 5(1), p. 367. doi: 10.1186/1756-0500-5-367.

Spezia, R., Aschi, M., Di Nola, A., Di Valentin, M., Carbonera, D. and Amadei, A. 2003. The Effect of Protein Conformational Flexibility on the Electronic Properties of a Chromophore. *Biophysical Journal* 84(5), pp. 2805–2813. doi: 10.1016/S0006-3495(03)70010-1.

Van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E. and Berendsen, H.J.C. 2005. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* 26(16), pp. 1701–1718. Available at: <Go to ISI>://WOS:000233021400004.

Steindal, A.H., Ruud, K., Frediani, L., Aidas, K. and Kongsted, J. 2011. Excitation Energies in Solution: The Fully Polarizable QM/MM/PCM Method. *The Journal of Physical Chemistry B* 115(12), pp. 3027–3037. doi: 10.1021/jp1101913.

Stepanenko, O., Stepanenko, O., Kuznetsova, I., Shcherbakova, D., Verkhusha, V. and Turoverov, K. 2017. Interaction of Biliverdin Chromophore with Near-Infrared Fluorescent Protein BphP1-FP Engineered from Bacterial Phytochrome. *International Journal of Molecular Sciences* 18(5), p. 1009. doi: 10.3390/ijms18051009.

Stepanenko, O. V., Stepanenko, O. V., Kuznetsova, I.M., Verkhusha, V. V. and Turoverov, K.K. 2013. Beta-Barrel Scaffold of Fluorescent Proteins. pp. 221–278. doi: 10.1016/B978-0-12-407699-0.00004-2.

Stepanenko, O. V., Stepanenko, O. V., Shcherbakova, D.M., Kuznetsova, I.M., Turoverov, K.K. and Verkhusha, V. V. 2011. Modern Fluorescent Proteins: From Chromophore Formation to Novel Intracellular Applications. *BioTechniques* 51(5), pp. 313–327. doi: 10.2144/000113765.

Stephens, P.J., Devlin, F.J., Chabalowski, C.F. and Frisch, M.J. 1994. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *The Journal of Physical Chemistry* 98(45), pp. 11623–11627. doi: 10.1021/j100096a001.

Stillinger, F.H. et al. 1995. *Mathematical Challenges from Theoretical/Computational Chemistry*. Washington, D.C.: National Academies Press. doi: 10.17226/4886.

Stokes, A.H. et al. 2000a. *p* -Ethynylphenylalanine. *Journal of Neurochemistry* 74(5), pp. 2067–2073. doi: 10.1046/j.1471-4159.2000.0742067.x.

Stokes, A.H. et al. 2000b. *p* -Ethynylphenylalanine. *Journal of Neurochemistry* 74(5), pp. 2067–2073. doi: 10.1046/j.1471-4159.2000.0742067.x.

Strongin, D.E. et al. 2007. Structural rearrangements near the chromophore influence the maturation speed and brightness of DsRed variants. *Protein Engineering, Design and Selection* 20(11), pp. 525–534. doi: 10.1093/protein/gzm046.

Studier, F.W. 2005. Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification* 41(1), pp. 207–234. Available at: <Go to ISI>://WOS:000228564900025.

Subach, F. V, Piatkevich, K.D. and Verkhusha, V. V. 2011. Directed molecular evolution to design advanced red fluorescent proteins. *Nature Methods* 8(12), pp. 1019–1026. Available at: <Go to ISI>://WOS:000297931700010.

Subach, F. V. and Verkhusha, V. V. 2012. Chromophore Transformations in Red Fluorescent Proteins. *Chemical Reviews* 112(7), pp. 4308–4327. doi: 10.1021/cr2001965.

Sun, M., Mu, X. and Li, R. 2024. The Principle, Application and Imaging of CARS. pp. 41–57. doi: 10.1007/978-981-99-3637-3_3.

Tachibana, S.R. et al. 2021. An Engineered Biliverdin-Compatible Cyanobacteriochrome Enables a Unique Ultrafast Reversible Photoswitching Pathway. *International Journal of Molecular Sciences* 22(10), p. 5252. doi: 10.3390/ijms22105252.

Tallents, G.J. 2018. Radiative Transitions between Discrete Quantum States. In: *An Introduction to the Atomic and Radiation Physics of Plasmas*. Cambridge University Press, pp. 177–207. doi: 10.1017/9781108303538.011.

Topol, I., Collins, J., Savitsky, A. and Nemukhin, A. 2011a. Computational strategy for tuning spectral properties of red fluorescent proteins. *Biophysical Chemistry* 158(2–3), pp. 91–95. doi: 10.1016/j.bpc.2011.05.016.

Topol, I., Collins, J., Savitsky, A. and Nemukhin, A. 2011b. Computational strategy for tuning spectral properties of red fluorescent proteins. *Biophysical Chemistry* 158(2–3), pp. 91–95. doi: 10.1016/j.bpc.2011.05.016.

Tsien, R.Y. 1998. The green fluorescent protein. *Annual Review of Biochemistry* 67, pp. 509–544. Available at: <Go to ISI>://WOS:000075721700018.

Tsikritsis, D., Legge, E.J. and Belsey, N.A. 2022. Practical considerations for quantitative and reproducible measurements with stimulated Raman scattering microscopy. *The Analyst* 147(21), pp. 4642–4656. doi: 10.1039/D2AN00817C.

Tsuboi, M., Benevides, J.M. and Thomas, G.J. 2009. Raman tensors and their application in structural studies of biological systems. *Proceedings of the Japan Academy. Series B, Physical and biological sciences* 85(3), pp. 83–97. doi: 10.2183/pjab.85.83.

Tsuneda, T. 2014. Exchange-Correlation Functionals. In: *Density Functional Theory in Quantum Chemistry*. Tokyo: Springer Japan, pp. 101–124. doi: 10.1007/978-4-431-54825-6_5.

Tucker, M.J., Oyola, R. and Gai, F. 2006. A novel fluorescent probe for protein binding and folding studies: *p* -cyano-phenylalanine. *Biopolymers* 83(6), pp. 571–576. doi: 10.1002/bip.20587.

Tunyasuvunakool, K. et al. 2021. Highly accurate protein structure prediction for the human proteome. *Nature* 596(7873), pp. 590–596. doi: 10.1038/s41586-021-03828-1.

Tzeliou, C.E., Mermigki, M.A. and Tzeli, D. 2022. Review on the QM/MM Methodologies and Their Application to Metalloproteins. *Molecules* 27(9), p. 2660. doi: 10.3390/molecules27092660.

Ullrich, C.A. 2011. *Time-Dependent Density-Functional Theory*. Oxford University Press. doi: 10.1093/acprof:oso/9780199563029.001.0001.

Váczi, T., Himics, L., Bruzzone, M., Veres, M. and dal Maschio, M. 2023. Spectrally Focused Stimulated Raman Scattering (sf-SRS) Microscopy for Label-Free Investigations of Molecular Mechanisms in Living Organisms. pp. 393–416. doi: 10.1007/978-1-0716-2764-8_13.

Varadi, M. et al. 2024. AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Research* 52(D1), pp. D368–D375. doi: 10.1093/nar/gkad1011.

Vardaki, M.Z., Gregoriou, V.G. and Chochos, C.L. 2024a. Biomedical applications, perspectives and tag design concepts in the cell – silent Raman window. *RSC Chemical Biology* 5(4), pp. 273–292. doi: 10.1039/D3CB00217A.

Vardaki, M.Z., Gregoriou, V.G. and Chochos, C.L. 2024b. Biomedical applications, perspectives and tag design concepts in the cell – silent Raman window. *RSC Chemical Biology* 5(4), pp. 273–292. doi: 10.1039/D3CB00217A.

Waddar, B., Parne, S.R., Gandi, S., Prasanth, G.R., Yaseen, M. and Kariduraganavar, M.Y. 2024. The second-order nonlinear optical properties of novel triazolo[3,4-b] [1, 3, 4] thiadiazole derivative chromophores using DFT calculations. *Structural Chemistry* 35(1), pp. 253–264. doi: 10.1007/s11224-023-02178-0.

Walther, K.A., Papke, B., Sinn, M.B., Michel, K. and Kinkhabwala, A. 2011. Precise measurement of protein interacting fractions with fluorescence lifetime imaging microscopy. *Molecular BioSystems* 7(2), p. 322. doi: 10.1039/c0mb00132e.

Wang, L., Xie, J., Deniz, A.A. and Schultz, P.G. 2003. Unnatural Amino Acid Mutagenesis of Green Fluorescent Protein. *The Journal of Organic Chemistry* 68(1), pp. 174–176. doi: 10.1021/jo026570u.

Weeks, C.L., Polishchuk, A., Getahun, Z., DeGrado, W.F. and Spiro, T.G. 2008. Investigation of an unnatural amino acid for use as a resonance Raman probe: detection limits and solvent and temperature dependence of the vC□N band of 4-cyanophenylalanine. *Journal of Raman Spectroscopy* 39(11), pp. 1606–1613. doi: 10.1002/jrs.2067.

Wegele, R., Tasler, R., Zeng, Y., Rivera, M. and Frankenberg-Dinkel, N. 2004. The Heme Oxygenase(s)-Phytochrome System of Pseudomonas aeruginosa. *Journal of Biological Chemistry* 279(44), pp. 45791–45802. doi: 10.1074/jbc.M408303200.

Wei, L. et al. 2017. Super-multiplex vibrational imaging. *Nature* 544(7651), pp. 465-+. Available at: <Go to ISI>://WOS:000400051900040.

Wei, L. and Min, W. 2018. Electronic Preresonance Stimulated Raman Scattering Microscopy. *The Journal of Physical Chemistry Letters* 9(15), pp. 4294–4301. doi: 10.1021/acs.jpclett.8b00204.

Welker, E.A. et al. 2015. Conformational Change with Steric Interactions Affects the Inner Sphere Component of Concerted Proton–Electron Transfer in a Pyridyl-Appended Radical Cation System. *The Journal of Organic Chemistry* 80(17), pp. 8705–8712. doi: 10.1021/acs.joc.5b01427.

Whitby, F.G., Phillips, J.D., Hill, C.P., McCoubrey, W. and Maines, M.D. 2002. Crystal Structure of a Biliverdin IXα Reductase Enzyme–Cofactor Complex. *Journal of Molecular Biology* 319(5), pp. 1199–1210. doi: 10.1016/S0022-2836(02)00383-2.

Williams, E.G. 1999. Spherical Waves. In: *Fourier Acoustics*. Elsevier, pp. 183–234. doi: 10.1016/B978-012753960-7/50006-1.

Wu, P.G. and Brand, L. 1994. Resonance Energy Transfer: Methods and Applications. *Analytical Biochemistry* 218(1), pp. 1–13. doi: 10.1006/abio.1994.1134.

Xu, T., Xu, Q. and Li, J. 2023. Toward the appropriate interpretation of Alphafold2. *Frontiers in Artificial Intelligence* 6. doi: 10.3389/frai.2023.1149748.

Yamada, T. and Mizuno, M. 2018. Characteristic Spectroscopic Features because of Cation–Anion Interactions Observed in the 700–950 cm $^{-1}$ Range of Infrared Spectroscopy for Various Imidazolium-Based Ionic Liquids. *ACS Omega* 3(7), pp. 8027–8035. doi: 10.1021/acsomega.8b00938.

Yang, X., Kuk, J. and Moffat, K. 2009a. Conformational differences between the Pfr and Pr states in *Pseudomonas aeruginosa* bacteriophytochrome. *Proceedings of the National Academy of Sciences* 106(37), pp. 15639–15644. doi: 10.1073/pnas.0902178106.

Yang, X., Kuk, J. and Moffat, K. 2009b. Conformational differences between the Pfr and Pr states in *Pseudomonas aeruginosa* bacteriophytochrome. *Proceedings of the National Academy of Sciences* 106(37), pp. 15639–15644. doi: 10.1073/pnas.0902178106.

Yang, X., Kuk, J. and Moffat, K. 2009c. Conformational differences between the Pfr and Pr states in *Pseudomonas aeruginosa* bacteriophytochrome. *Proceedings of the National Academy of Sciences* 106(37), pp. 15639–15644. doi: 10.1073/pnas.0902178106.

Yang, X., Kuk, J. and Moffat, K. 2009d. Conformational differences between the Pfr and Pr states in *Pseudomonas aeruginosa* bacteriophytochrome. *Proceedings of the National Academy of Sciences* 106(37), pp. 15639–15644. doi: 10.1073/pnas.0902178106.

Yu, H.S., Li, S.L. and Truhlar, D.G. 2016. Perspective: Kohn-Sham density functional theory descending a staircase. *The Journal of Chemical Physics* 145(13). doi: 10.1063/1.4963168.

Yuan, Y., Wang, D., Zhang, J., Liu, J., Chen, J. and Zhang, X.-E. 2018. Raman spectra of the GFP-like fluorescent proteins. *Biophysics Reports* 4(5), pp. 265–272. doi: 10.1007/s41048-018-0072-0.

Zanetti-Domingues, L.C., Tynan, C.J., Rolfe, D.J., Clarke, D.T. and Martin-Fernandez, M. 2013. Hydrophobic Fluorescent Probes Introduce Artifacts into Single Molecule Tracking

Experiments Due to Non-Specific Binding. *PLoS ONE* 8(9), p. e74200. doi: 10.1371/journal.pone.0074200.

Zhang, H. et al. 2023a. Quantitative assessment of near-infrared fluorescent proteins. *Nature Methods* 20(10), pp. 1605–1616. doi: 10.1038/s41592-023-01975-z.

Zhang, S., Qi, Y., Tan, S.P.H., Bi, R. and Olivo, M. 2023b. Molecular Fingerprint Detection Using Raman and Infrared Spectroscopy Technologies for Cancer Detection: A Progress Review. *Biosensors* 13(5), p. 557. doi: 10.3390/bios13050557.

Zwiebach, B. 2022. *Mastering Quantum Mechanics: Essentials, Theory, and Applications*. Cambridge, MA: MIT Press.

9 Supplementary Material

9.1 Supplementary Tables

Table S1: DNA sequences of WT proteins

WT protein	DNA sequence
mRhubarb720	ATGGGCAGCAGCCATCACCATCATCACCACAGCCAGGATCCGATGGCTGAAGGATCCGTCGCCA GGCAGCCTGACCTCTTGACCTGCGACGATGAGCCGATCCATATCCCCGGTGCCATCCAACCGCA TGGACTGCTGCTCGCCCTCGCCGCCGACATGACGATCGTTGCCGGCAGCGACAACCTTCCCGAA CTCACCGGACTGGCGATCGGCGCCCTGATCGGCCGCTCTGCGGCCGATGTCTTCGACTCGGAGA CGCACAACCGTCTGACGATCGCCTTGGCCGAGCCCGGGGCGGCCGTCGGAGCACCGATCACTG TCGGCTTCACGATGCGAAAGGACGCAGGCTTCATCGGCTCCTGGCATCGCCATGATCAGCTCAT CTTCCTCGAGCTCGAGCCTCCCCAGCGGGACGTCGCCGAGCCGCAGGCGTTCTTCCGCCACACC AACAGCGCCATCCGCCGCCTGCAGGCCGCCGAAACCTTGGAGAGCGCCTGCGCCGCCGCGGCG CAAGAGGTGCGGAAGATTACCGGCTTCGATCGGGTGATGATCTATCGCTTCGCCTCCGACTTCA GCGGCGAAGTGATCGCAGAGGATCGGTGCGCCGAGGTCGAGTCAAAACTAGGCCAGCACTATC CTGCCTCAGATATTCCGGCGCAGGCCCGTCGGCTCTATACCATCAACCCGGTACGGATCATTCCC GATATCAATTATCGGCCGGTGCCGGTCACCCCAGACCTCAATCCGGTCACCGGGCGGCCGATTG ATCTTAGCTTCGCCATCCTGCGCAGCGTCTCGCCCGTCCATCTGGAATTCATGCGCAACATAGGC ATGCACGGCACGATGTCGATCTCGATTTTGCGCGGCGAGCGACTGTGGGGATTGATCGTTTGCC ATCACCGAACGCCGTACTACGTCGATCTCGATGGCCGCCAAGCCTGCGAGCTAGTCGCCCAGGT TCTGGCCCGGGCGATCGGCGTGATGGAAGAGTAA
emiRFP670	ATGGGGTAGTCACCATCACCACCACCATGGCAGATCTATGGCGGAAGGTTCCGTCGCCCGACAG CCTGATTTGTTGACATGCGAACATGAAGAGATTCACCTCGCCGGCAGTATCCAACCGCATGGAG CGCTTTTAGTGGTAAGCGAACATGATCATCGCGTCATCCAGGCAAGCGCGAACGCAGCAGAATT TTTGAATCTTGGAAGCGTACTCGGTGTTCCGTTAGCCGAGATTGATGGCGATCTGTTGATTAAAA TATTGCCGCATCTCGATCCCACCGCCGAAGGAATGCCGGTAGCTGTTCGCTGTCGGATTGGGAA TCCCTCTACGGAGTACTGCGGTCTGATGCATCGTCCACCAGAAGGTGGGCTTATCATCGAATTA GAACGTGCCGGACCGTCGATAGATCTGTCAGGTACGCTCGCTCCTGCGCTGGAGCGCATTCGTA CTGCTGGTAGTCTGCGCGCTCTGTGTGATGACACTGTGCTGTTATTTCAACAGTGTACCGGCTAC GACCGGGTTATGGTTTATCGTTTTGATGAGCAAGGTCACGGCCTGGTATTTTCCGAGTGCCATG TGCCTGGTTTAGAATCTTATTTCGGCAACCGTTATCCTTCATCTACTGTCCCACAGATGGCTAGAC AGCTGTACGTGCGTCAGCGAGTCCGCGTTTTAGTGGACGTTACCTATCAACCAGTTCCACTGGA ACCTCGTCTTTCACCCTTAACCGGGAGAGATCTTGACATGTCGGGCTGCTTTCTGCGCTCCATGA

	GTCCGTGTCATTTGCAGTTCCTGAAGGACATGGGCGTGCGCGCAACACTGGCGGTGTCACTGGT GGTCGGGGGAAAACTGTGGGGGTTAGTTGTATGTCACCATTATCTGCCGAGATTCATTCGTTTT GAGCTTCGAGCGATATGCAAACGGCTCGCCGAACGGATTGCAACGCGTATCACAGCACTTGAA AGCTAA
miRFP670nano3	ATGGGGAGCCACCATCACCATCACCATGGCAGATCTGCAGCTGGTACCATGGCAAACCTGGACA AGATGCTGAACACCACCGTGACCGAGGTGCGCAAGTTCCTGCAAGCGGACAGAGTGTGCGTGT TCAAGTTCGAGGAAGATTACTCCGGCACCGTCTCGCACGAAGCCGTGGACGACAGATGGATTA GCATCCTGAAGACCCAGGTGCAGGACAGATACTTCATGGAAACCAGAGGCGAGGAATACGTCC ACGGCAGATACCAGGCCATCGCCGACATCTACACAGCCAATCTGGTCGAGTGCTACAGAGACCT GCTGATCGAGTTTCAGGTGCGGGCCATTCTGGCTGTCCCCATCCTGCAAGGCAAGAAGCTGTGG GGCCTGCTGGTGGCCCACCAACTGGCCGGCCCTCGGGAGTGGCAGACCTGGGAAATCGACTTC CTGAAACAGCAAGCCGTGGTGATGGGCATCGCCATCCAGCAGAGCTGA
mCherry	ATGGGGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGG GTCGGGATCTGTACGACGATGACGATAAGGATCCGAGCTCGAGCATGGTGAGCAAGGGCGAG GAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTG AACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGAC CGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAG TTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTC CTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGT GACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAA CTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCG GATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACG GCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCG GCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGA ACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTAA

Table S2: Protein sequences of WT proteins

WT protein	**Protein sequence**
mRhubarb720	MGSSHHHHHHSQDPMAEGSVARQPDLLTCDDEPIHIPGAIQPHGLLLALAADMTIVAGSDNLPE LTGLAIGALIGRSAADVFDSETHNRLTIALAEPGAAVGAPITVGFTMRKDAGFIGSWHRHDQLIFLE LEPPQRDVAEPQAFFRHTNSAIRRLQAAETLESACAAAAQEVRKITGFDRVMIYRFASDFSGEVIAE DRCAEVESKLGQHYPASDIPAQARRLYTINPVRIIPDINYRPVPVTPDLNPVTGRPIDLSFAILRSVSP VHLEFMRNIGMHGTMSISILRGERLWGLIVCHHRTPYYVDLDGRQACELVAQVLARAIGVMEE*

emiRFP670	MGSHHHHHHGRSMAEGSVARQPDLLTCEHEEIHLAGSIQPHGALLVVSEHDHRVIQASANAAEF LNLGSVLGVPLAEIDGDLLIKILPHLDPTAEGMPVAVRCRIGNPSTEYCGLMHRPPEGGLIIELERAG PSIDLSGTLAPALERIRTAGSLRALCDDTVLLFQQCTGYDRVMVYRFDEQGHGLVFSECHVPGLESY FGNRYPSSTVPQMARQLYVRQRVRVLVDVTYQPVPLEPRLSPLTGRDLDMSGCFLRSMSPCHLQ FLKDMGVRATLAVSLVVGGKLWGLVVCHHYLPRFIRFELRAICKRLAERIATRITALES*
miRFP670nano3	MGSHHHHHHGRSAAGTMANLDKMLNTTVTEVRKFLQADRVCVFKFEEDYSGTVSHEAVDDRW ISILKTQVQDRYFMETRGEEYVHGRYQAIADIYTANLVECYRDLLIEFQVRAILAVPILQGKKLWGLL VAHQLAGPREWQTWEIDFLKQQAVVMGIAIQQS*
mCherry	MGGSHHHHHHGMASMTGGQQMGRDLYDDDDKDPSSSMVSKGEEDNMAIIKEFMRFKVHME GSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKL SFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSE RMYPEDGALKGEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQY ERAEGRHSTGGMDELYK*

Table S3: DNA sequences of mammalian cell constructs

Construct	DNA sequence
H2B-mRhubarb720	ATGCCAGAGCCAGCGAAGTCTGCTCCCGCCCCGAAAAAGGGCTCCAAGAAGGCGGTGACT AAGGCGCAGAAGAAAGGCGGCAAGAAGCGCAAGCGCAGCCGCAAGGAGAGCTATTCCAT CTATGTGTACAAGGTTCTGAAGCAGGTCCACCCTGACACCGGCATTTCGTCCAAGGCCATG GGCATCATGAATTCGTTTGTGAACGACATTTTCGAGCGCATCGCAGGTGAGGCTTCCCGCCT GGCGCATTACAACAAGCGCTCGACCATCACCTCCAGGGAGATCCAGACGGCCGTGCGCCTG CTGCTGCCTGGGGAGTTGGCCAAGCACGCCGTGTCCGAGGGTACTAAGGCCATCACCAAG TACACCAGCAGCTCTAAGGATCCACCGGTCGCCACCATGGGCAGCAGCAGCCAGGATCCGA TGGCTGAAGGATCCGTCGCCAGGCAGCCTGACCTCTTGACCTGCGACGATGAGCCGATCCA TATCCCCGGTGCCATCCAACCGCATGGACTGCTGCTCGCCCTCGCCGCCGACATGACGATCG TTGCCGGCAGCGACAACCTTCCCGAACTCACCGGACTGGCGATCGGCGCCCTGATCGGCCG CTCTGCGGCCGATGTCTTCGACTCGGAGACGCACAACCGTCTGACGATCGCCTTGGCCGAG CCCGGGGCGGCCGTCGGAGCACCGATCACTGTCGGCTTCACGATGCGAAAGGACGCAGGC TTCATCGGCTCCTGGCATCGCCATGATCAGCTCATCTTCCTCGAGCTCGAGCCTCCCCAGCG GGACGTCGCCGAGCCGCAGGCGTTCTTCCGCCACACCAACAGCGCCATCCGCCGCCTGCAG GCCGCCGAAACCTTGGAGAGCGCCTGCGCCGCCGCGGCGCAAGAGGTGCGGAAGATTACC GGCTTCGATCGGGTGATGATCTATCGCTTCGCCTCCGACTTCAGCGGCGAAGTGATCGCAG AGGATCGGTGCGCCGAGGTCGAGTCAAAACTAGGCCAGCACTATCCTGCCTCAGATATTCC GGCGCAGGCCCGTCGGCTCTATACCATCAACCCGGTACGGATCATTCCCGATATCAATTATC

	GGCCGGTGCCGGTCACCCCAGACCTCAATCCGGTCACCGGGCGGCCGATTGATCTTAGCTT CGCCATCCTGCGCAGCGTCTCGCCCGTCCATCTGGAATTCATGCGCAACATAGGCATGCAC GGCACGATGTCGATCTCGATTTTGCGCGGCGAGCGACTGTGGGGATTGATCGTTTGCCATC ACCGAACGCCGTACTACGTCGATCTCGATGGCCGCCAAGCCTGCGAGCTAGTCGCCCAGGT TCTGGCCCGGGCGATCGGCGTGATGGAAGAGTAA
LAMP1- mRhubarb720	ATGGCGGCCCCGGGCGCCCGGCGGCCGCTGCTCCTGTTGCTGCTGGCAGGCCTTGCACACA GCGCCCCAGCACTGTTCGAGGTGAAAGACAACAACGGCACAGCGTGTATAATGGCCAGCT TCTCTGCCTCCTTTCTGACCACCTATGAGGCTGGACATGTTTCTAAGGTCTCGAATATGACCC TGCCAGCCTCTGCAGAAGTCCTGAAGAATAGCAGCTCTTGTGGTGAAAAGAATGCTTCTGA GCCCACCCTCGCAATCACCTTTGGAGAAGGATATTTACTGAAACTCACCTTCACAAAAAACA CAACACGTTACAGTGTCCAGCACATGTATTTCACATATAACCTGTCAGACACACAATTCTTTC CCAATGCCAGCTCCAAAGGGCCCGACACTGTGGATTCCACAACTGACATCAAGGCAGACAT CAACAAAACATACCGATGTGTCAGCGACATCAGGGTCTACATGAAGAATGTGACCATTGTG CTCTGGGACGCTACTATCCAGGCCTACCTGCCGAGTAGCAACTTCAGCAAGGAAGAGACAC GCTGCCCACAGGATCAACCTTCCCCAACTACTGGGCCACCCAGCCCCTCACCACCACTCGTG CCCACAAACCCCAGTGTGTCCAAGTACAATGTGACTGGTGACAATGGAACCTGCCTGCTGG CCTCTATGGCACTGCAACTCAACATCACCTACATGAAGAAGGACAACACGACTGTGACCAG AGCATTCAACATCAACCCAAGTGACAAATATAGTGGGACTTGCGGTGCCCAGTTGGTGACC CTGAAGGTGGGGAACAAGAGCAGAGTCCTGGAGCTGCAGTTTGGGATGAATGCCACTTCT AGCCTGTTTTTCCTGCAAGGAGTTCAGTTGAACATGACTCTTCCTGATGCCATAGAGCCCAC GTTCAGCACCTCCAACTATTCCCTGAAAGCTCTTCAGGCCAGTGTCGGCAACTCATACAAGT GCAACAGTGAGGAGCACATCTTTGTCAGCAAGGCGCTCGCCCTCAATGTCTTCAGCGTGCA AGTCCAGGCTTTCAGGGTAGAAAGTGACAGGTTTGGGTCTGTGGAAGAGTGTGTACAGGA CGGTAACAACATGCTGATCCCCATTGCTGTGGGCGGGGCCCTGGCAGGGCTGGTCCTCATC GTCCTCATCGCCTACCTCATCGGCAGGAAGAGGAGTCACGCGGGCTATCAGACCATCCGGG ATCCACCGGTCGCCACCATGGGCAGCAGCAGCCAGGATCCGATGGCTGAAGGATCCGTCG CCAGGCAGCCTGACCTCTTGACCTGCGACGATGAGCCGATCCATATCCCCGGTGCCATCCA ACCGCATGGACTGCTGCTCGCCCTCGCCGCCGACATGACGATCGTTGCCGGCAGCGACAAC CTTCCCGAACTCACCGGACTGGCGATCGGCGCCCTGATCGGCCGCTCTGCGGCCGATGTCT TCGACTCGGAGACGCACAACCGTCTGACGATCGCCTTGGCCGAGCCCGGGGCGGCCGTCG GAGCACCGATCACTGTCGGCTTCACGATGCGAAAGGACGCAGGCTTCATCGGCTCCTGGCA TCGCCATGATCAGCTCATCTTCCTCGAGCTCGAGCCTCCCCAGCGGGACGTCGCCGAGCCG CAGGCGTTCTTCCGCCACACCAACAGCGCCATCCGCCGCCTGCAGGCCGCCGAAACCTTGG AGAGCGCCTGCGCCGCCGCGGCGCAAGAGGTGCGGAAGATTACCGGCTTCGATCGGGTGA TGATCTATCGCTTCGCCTCCGACTTCAGCGGCGAAGTGATCGCAGAGGATCGGTGCGCCGA GGTCGAGTCAAAACTAGGCCAGCACTATCCTGCCTCAGATATTCCGGCGCAGGCCCGTCGG

	CTCTATACCATCAACCCGGTACGGATCATTCCCGATATCAATTATCGGCCGGTGCCGGTCAC CCCAGACCTCAATCCGGTCACCGGGCGGCCGATTGATCTTAGCTTCGCCATCCTGCGCAGC GTCTCGCCCGTCCATCTGGAATTCATGCGCAACATAGGCATGCACGGCACGATGTCGATCTC GATTTTGCGCGGCGAGCGACTGTGGGGATTGATCGTTTGCCATCACCGAACGCCGTACTAC GTCGATCTCGATGGCCGCCAAGCCTGCGAGCTAGTCGCCCAGGTTCTGGCCCGGGCGATCG GCGTGATGGAAGAGTAA
pLifeAct- mRhubarb720	ATGGGCGTGGCCGACTTGATCAAGAAGTTCGAGTCCATCTCCAAGGAGGAGGGGGATCCA CCGGTCGCCACCATGGGCAGCAGCAGCCAGGATCCGATGGCTGAAGGATCCGTCGCCAGG CAGCCTGACCTCTTGACCTGCGACGATGAGCCGATCCATATCCCCGGTGCCATCCAACCGCA TGGACTGCTGCTCGCCCTCGCCGCCGACATGACGATCGTTGCCGGCAGCGACAACCTTCCC GAACTCACCGGACTGGCGATCGGCGCCCTGATCGGCCGCTCTGCGGCCGATGTCTTCGACT CGGAGACGCACAACCGTCTGACGATCGCCTTGGCCGAGCCCGGGGCGGCCGTCGGAGCAC CGATCACTGTCGGCTTCACGATGCGAAAGGACGCAGGCTTCATCGGCTCCTGGCATCGCCA TGATCAGCTCATCTTCCTCGAGCTCGAGCCTCCCCAGCGGGACGTCGCCGAGCCGCAGGCG TTCTTCCGCCACACCAACAGCGCCATCCGCCGCCTGCAGGCCGCCGAAACCTTGGAGAGCG CCTGCGCCGCCGCGGCGCAAGAGGTGCGGAAGATTACCGGCTTCGATCGGGTGATGATCT ATCGCTTCGCCTCCGACTTCAGCGGCGAAGTGATCGCAGAGGATCGGTGCGCCGAGGTCG AGTCAAAACTAGGCCAGCACTATCCTGCCTCAGATATTCCGGCGCAGGCCCGTCGGCTCTAT ACCATCAACCCGGTACGGATCATTCCCGATATCAATTATCGGCCGGTGCCGGTCACCCCAGA CCTCAATCCGGTCACCGGGCGGCCGATTGATCTTAGCTTCGCCATCCTGCGCAGCGTCTCGC CCGTCCATCTGGAATTCATGCGCAACATAGGCATGCACGGCACGATGTCGATCTCGATTTTG CGCGGCGAGCGACTGTGGGGATTGATCGTTTGCCATCACCGAACGCCGTACTACGTCGATC TCGATGGCCGCCAAGCCTGCGAGCTAGTCGCCCAGGTTCTGGCCCGGGCGATCGGCGTGA TGGAAGAGTAA

Table S4: DNA sequences of mammalian cell constructs

Construct	**Protein sequence**
H2B- mRhubarb720	MPEPAKSAPAPKKGSKKAVTKAQKKGGKKRKRSRKESYSIYVYKVLKQVHPDTGISSKAMGIM NSFVNDIFERIAGEASRLAHYNKRSTITSREIQTAVRLLLPGELAKHAVSEGTKAITKYTSSSKDPPV ATMGSSSQDPMAEGSVARQPDLLTCDDEPIHIPGAIQPHGLLLALAADMTIVAGSDNLPELTGL AIGALIGRSAADVFDSETHNRLTIALAEPGAAVGAPITVGFTMRKDAGFIGSWHRHDQLIFLELE PPQRDVAEPQAFFRHTNSAIRRLQAAETLESACAAAAQEVRKITGFDRVMIYRFASDFSGEVIAE DRCAEVESKLGQHYPASDIPAQARRLYTINPVRIIPDINYRPVPVTPDLNPVTGRPIDLSFAILRSV

	SPVHLEFMRNIGMHGTMSISILRGERLWGLIVCHHRTPYYVDLDGRQACELVAQVLARAIGVM EE*
LAMP1- mRhubarb720	MAAPGARRPLLLLLLAGLAHSAPALFEVKDNNGTACIMASFSASFLTTYEAGHVSKVSNMTLPA SAEVLKNSSSCGEKNASEPTLAITFGEGYLLKLTFTKNTTRYSVQHMYFTYNLSDTQFFPNASSKG PDTVDSTTDIKADINKTYRCVSDIRVYMKNVTIVLWDATIQAYLPSSNFSKEETRCPQDQPSPTT GPPSPSPPLVPTNPSVSKYNVTGDNGTCLLASMALQLNITYMKKDNTTVTRAFNINPSDKYSGT CGAQLVTLKVGNKSRVLELQFGMNATSSLFFLQGVQLNMTLPDAIEPTFSTSNYSLKALQASVG NSYKCNSEEHIFVSKALALNVFSVQVQAFRVESDRFGSVEECVQDGNNMLIPIAVGGALAGLVLI VLIAYLIGRKRSHAGYQTIRDPPVATMGSSSQDPMAEGSVARQPDLLTCDDEPIHIPGAIQPHGL LLALAADMTIVAGSDNLPELTGLAIGALIGRSAADVFDSETHNRLTIALAEPGAAVGAPITVGFT MRKDAGFIGSWHRHDQLIFLELEPPQRDVAEPQAFFRHTNSAIRRLQAAETLESACAAAAQEV RKITGFDRVMIYRFASDFSGEVIAEDRCAEVESKLGQHYPASDIPAQARRLYTINPVRIIPDINYRP VPVTPDLNPVTGRPIDLSFAILRSVSPVHLEFMRNIGMHGTMSISILRGERLWGLIVCHHRTPYY VDLDGRQACELVAQVLARAIGVMEE*
pLifeAct- mRhubarb720	MGVADLIKKFESISKEEGDPPVATMGSSSQDPMAEGSVARQPDLLTCDDEPIHIPGAIQPHGLL LALAADMTIVAGSDNLPELTGLAIGALIGRSAADVFDSETHNRLTIALAEPGAAVGAPITVGFTM RKDAGFIGSWHRHDQLIFLELEPPQRDVAEPQAFFRHTNSAIRRLQAAETLESACAAAAQEVRK ITGFDRVMIYRFASDFSGEVIAEDRCAEVESKLGQHYPASDIPAQARRLYTINPVRIIPDINYRPVP VTPDLNPVTGRPIDLSFAILRSVSPVHLEFMRNIGMHGTMSISILRGERLWGLIVCHHRTPYYVD LDGRQACELVAQVLARAIGVMEE*

9.2 Supplementary Figures

Non-fluorescent cells imaging

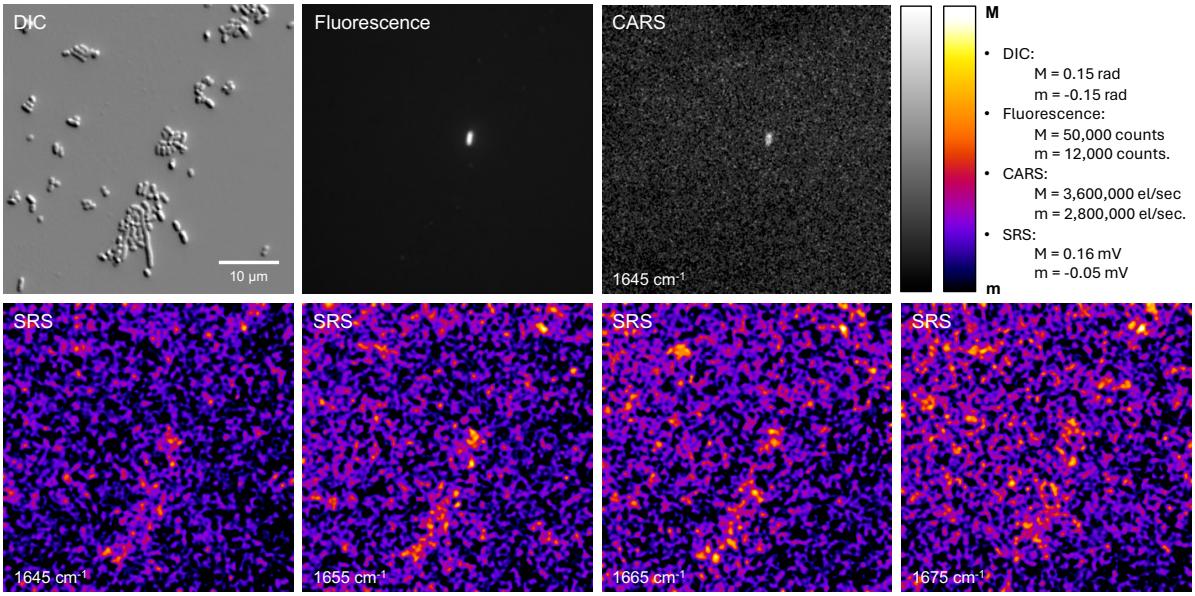


Figure S1: Mammalian cell imaging of non-fluorescent bacterial cells using false colour. (A) DIC, (B) fluorescence and (C) CARS at 1665 cm^{-1} present the same information highlighted in the previous figure. (D-G) SRS images were edited using ImageJ to provide false colouring for contrast of the Raman signal.

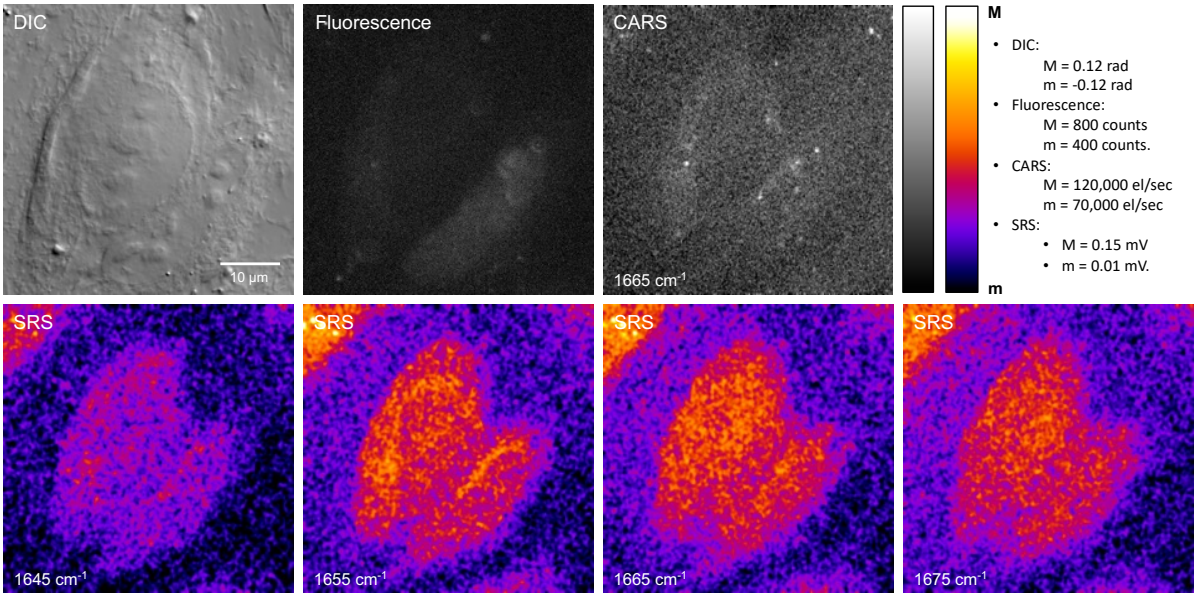


Figure S2: Mammalian cell imaging of non-fluorescent HeLa cells using false colour. (A) DIC, (B) fluorescence and (C) CARS at 1665 cm^{-1} present the same information highlighted in the previous figure. (D-G) SRS images were edited using ImageJ to provide false colouring for contrast of the Raman signal.

mRhubarb720 RMSD graphs

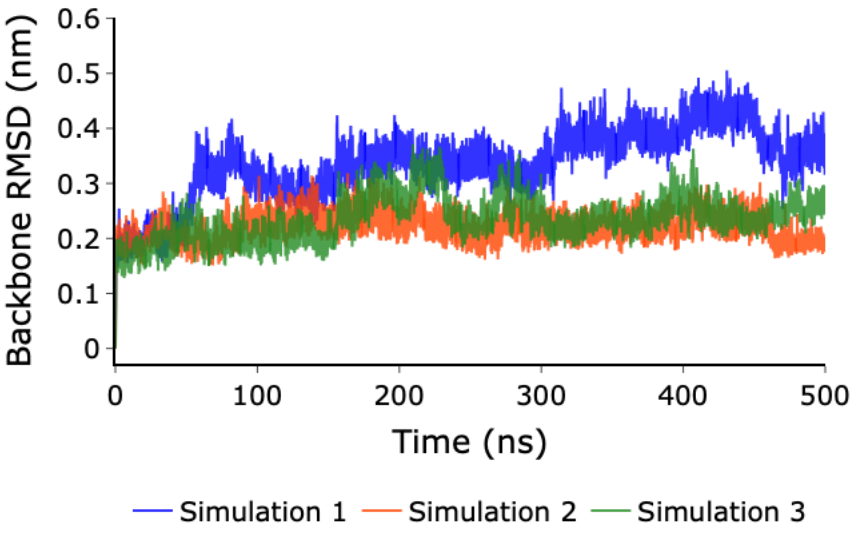


Figure S3: mRhubarb720 D177pCNPhe protein backbone RMSD analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

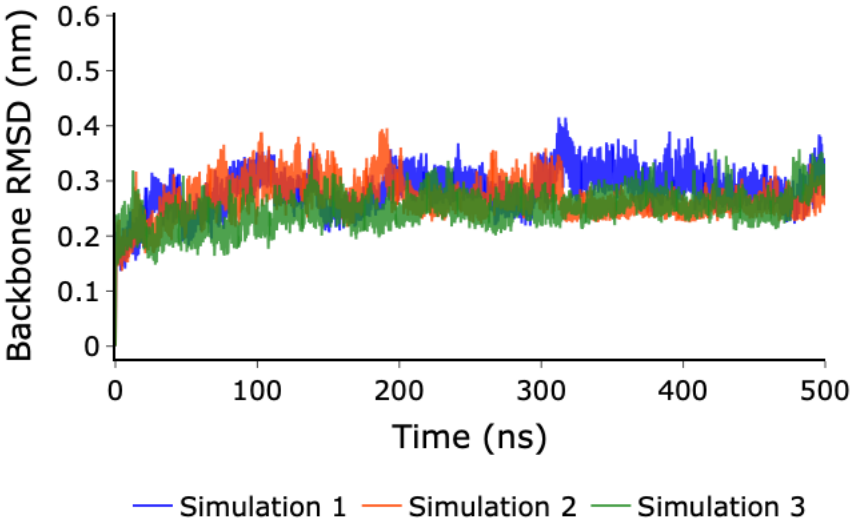


Figure S4: mRhubarb720 Y198pCNPhe protein backbone RMSD analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

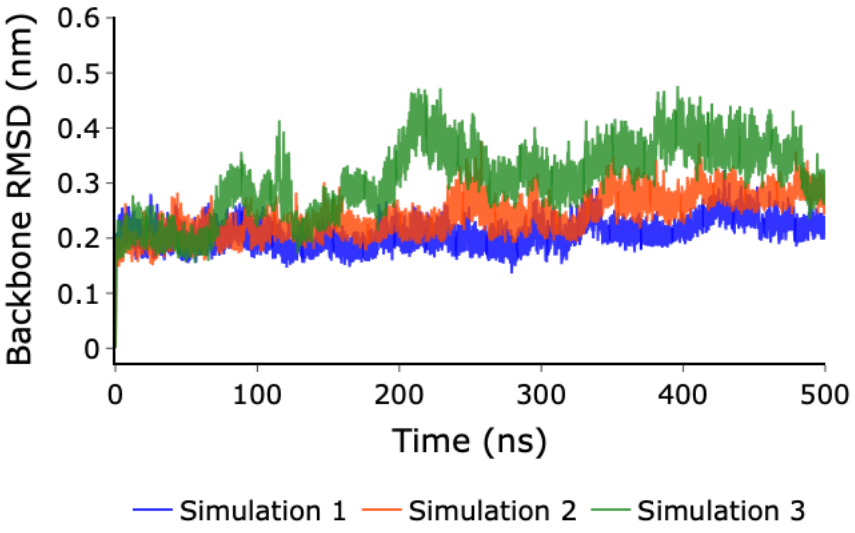


Figure S5: mRhubarb720 N224pCNPhe protein backbone RMSD analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

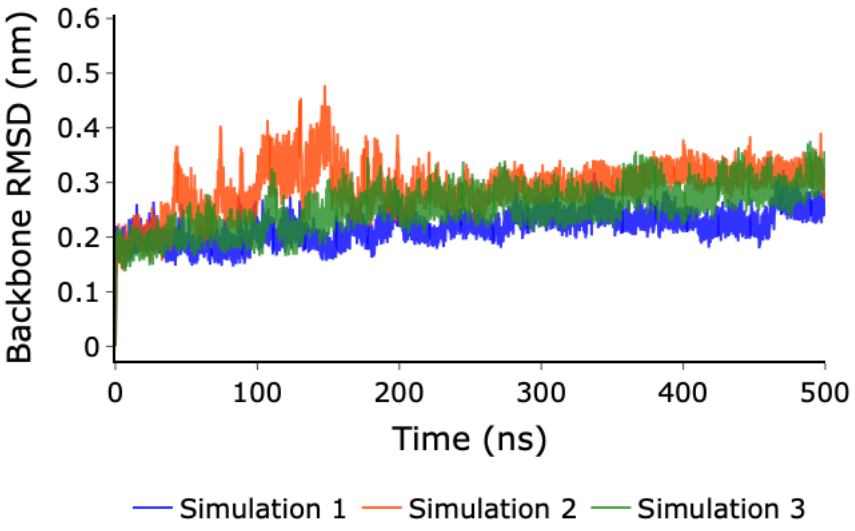


Figure S6: mRhubarb720 F258pCNPhe protein backbone RMSD analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

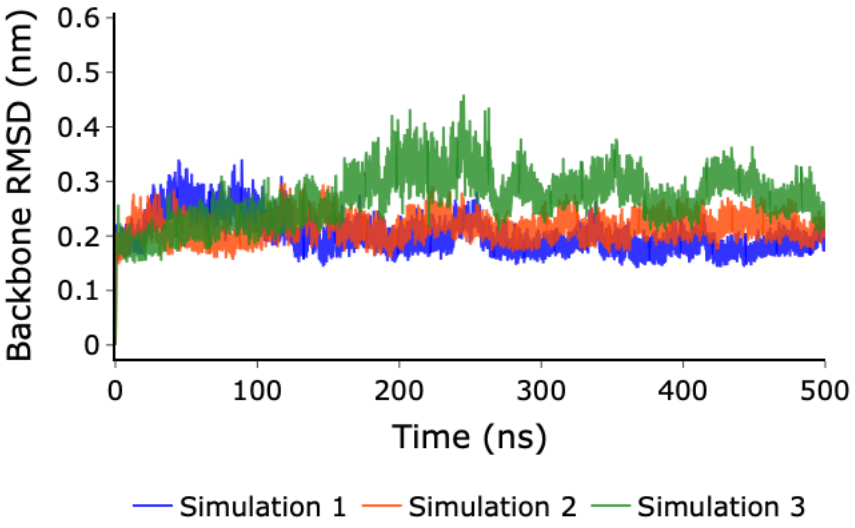


Figure S7: mRhubarb720 Y198/F258pCNPhe protein backbone RMSD analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

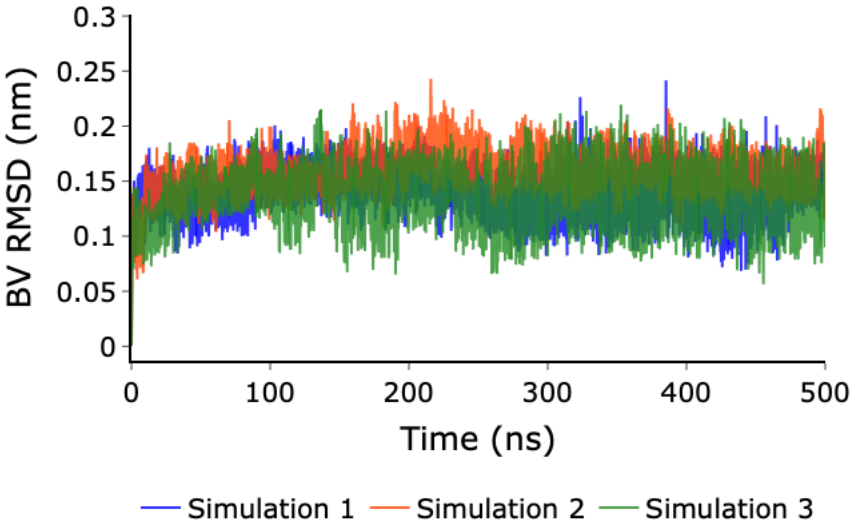


Figure S8: mRhubarb720 D177pCNPhe BV RMSD analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

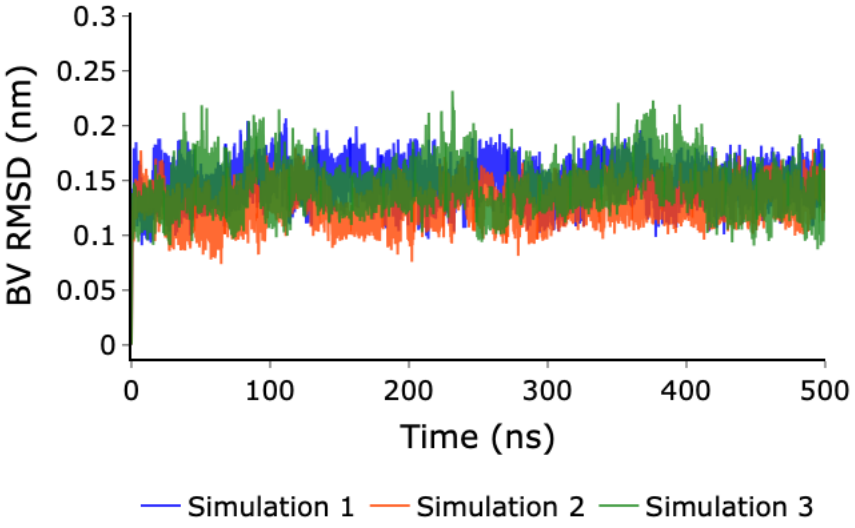


Figure S9: mRhubarb720 Y198pCNPhe BV RMSD analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

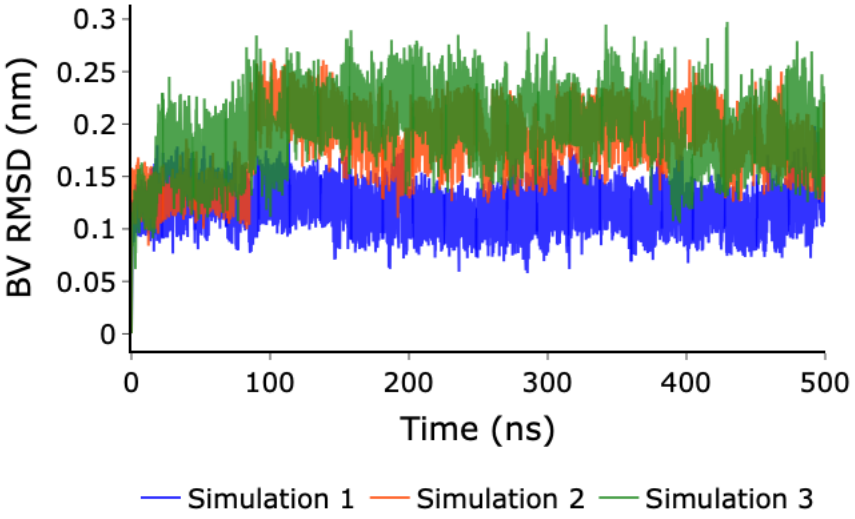


Figure S10: mRhubarb720 N224pCNPhe BV RMSD analysis over three simulations Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

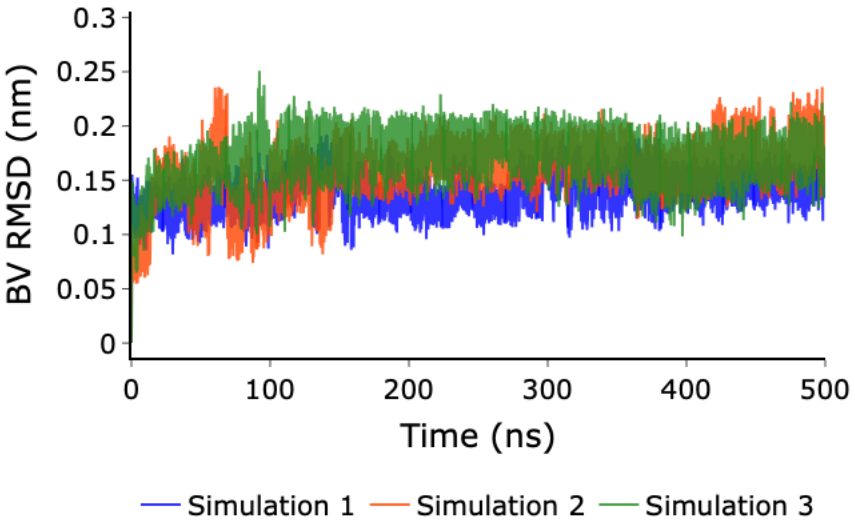


Figure S11: mRhubarb720 F258pCNPhe BV RMSD analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

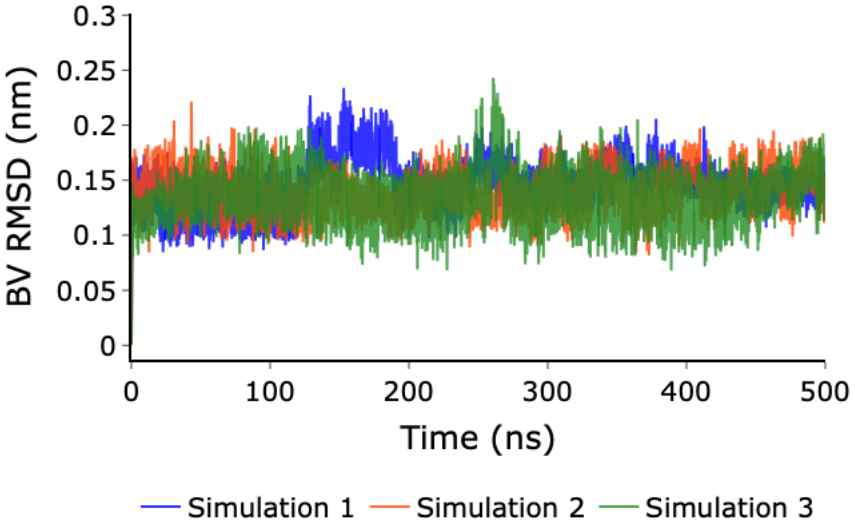


Figure S12: mRhubarb720 Y198/F258pCNPhe BV RMSD analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

mRhubarb720 RMSF graphs

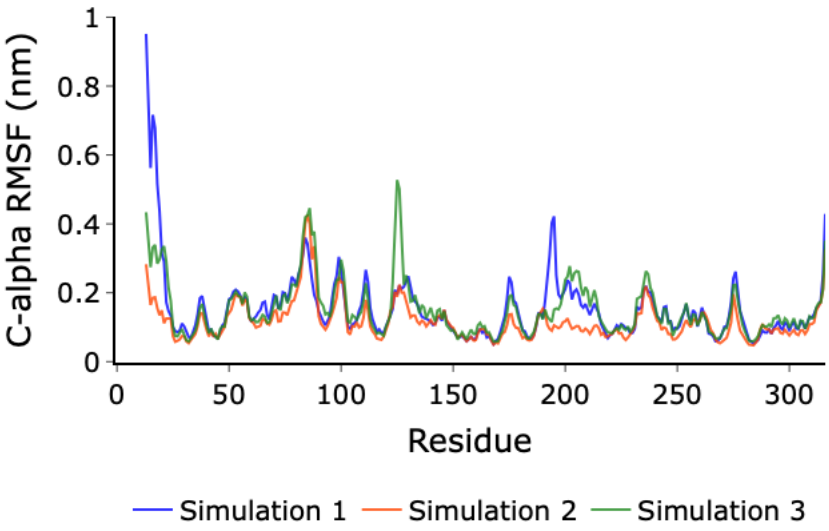


Figure S13: mRhubarb720 D177pCNPhe protein C-alpha RMSF analysis over three simulations

Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

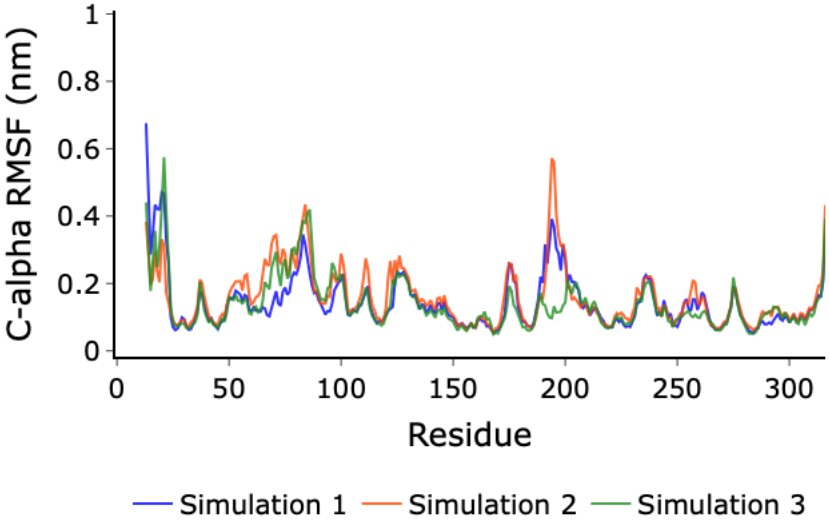


Figure S14: mRhubarb720 Y198pCNPhe protein C-alpha RMSF analysis over three simulations.

Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

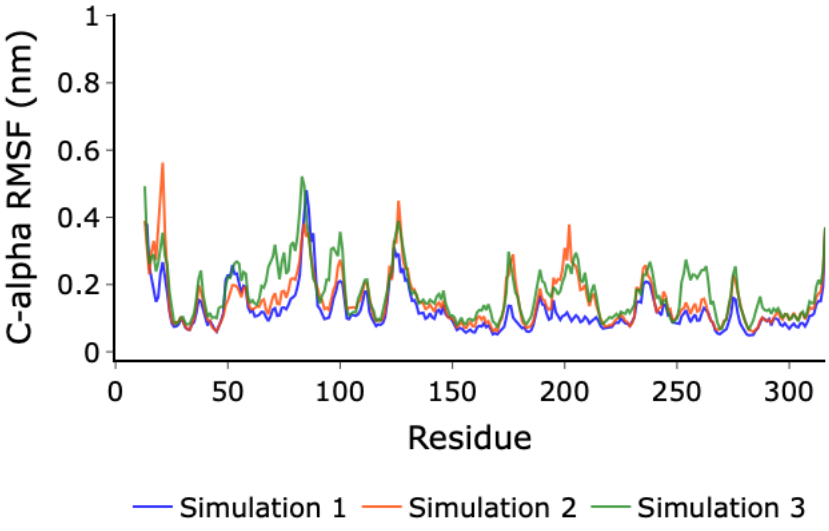


Figure S15: mRhubarb720 N224pCNPhe protein C-alpha RMSF analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

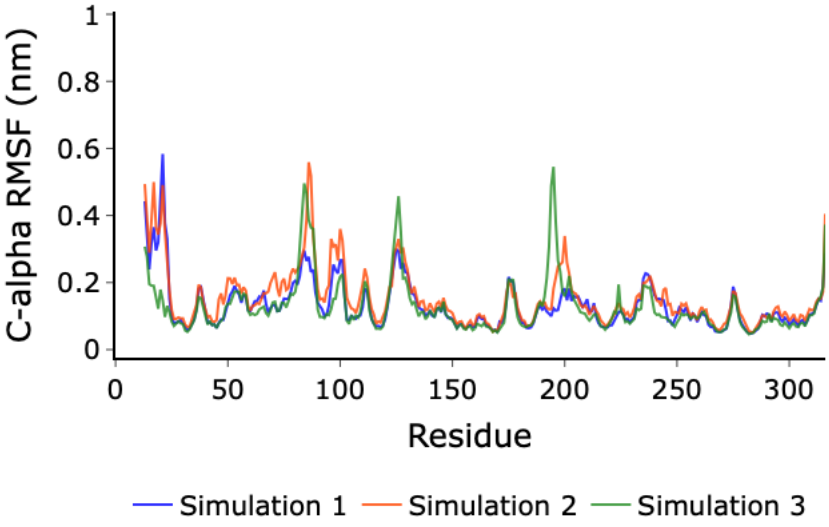


Figure S16: mRhubarb720 F258pCNPhe protein C-alpha RMSF analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

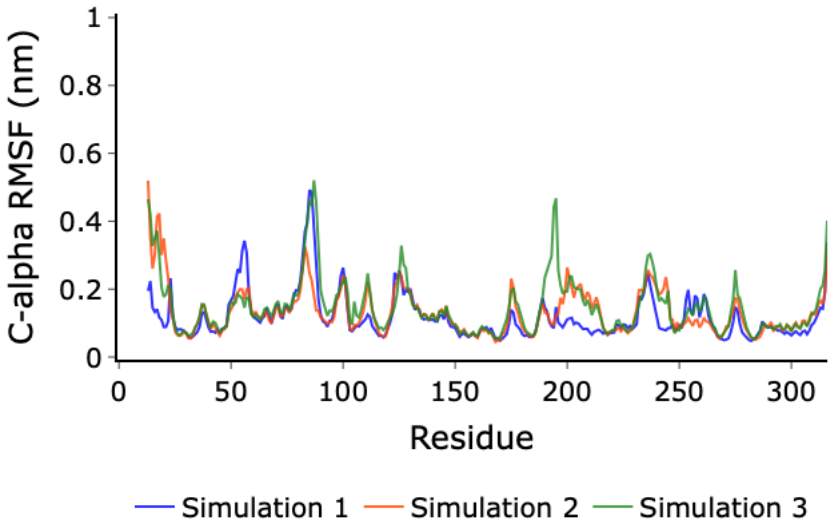


Figure S17: mRhubarb720 Y198/F258pCNPhe protein C-alpha RMSF analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

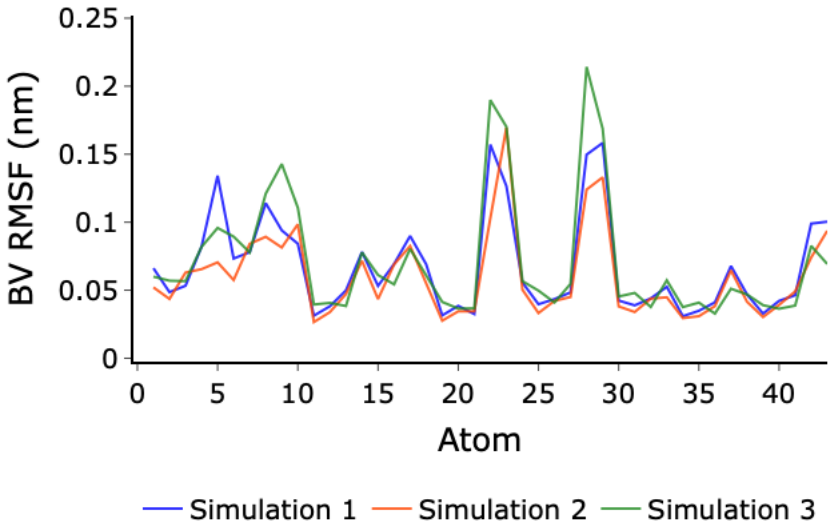


Figure S18: mRhubarb720 D177pCNPhe BV RMSF analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

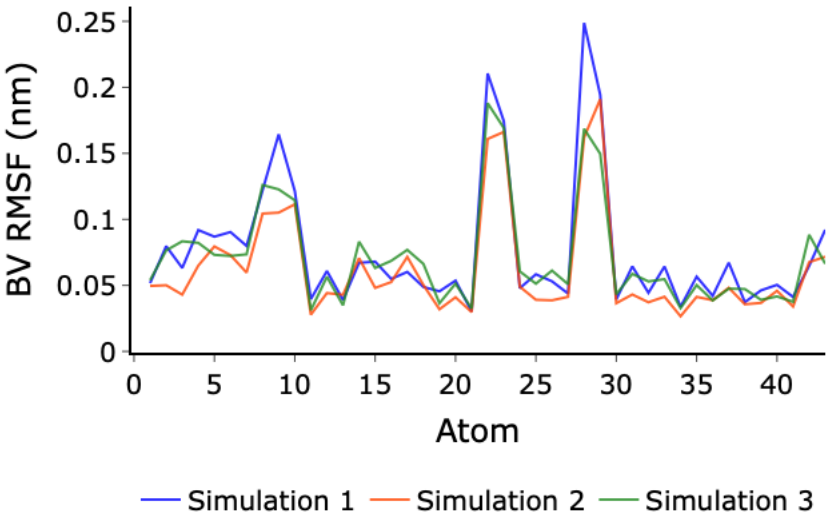


Figure S19: mRhubarb720 Y198pCNPhe BV RMSF analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

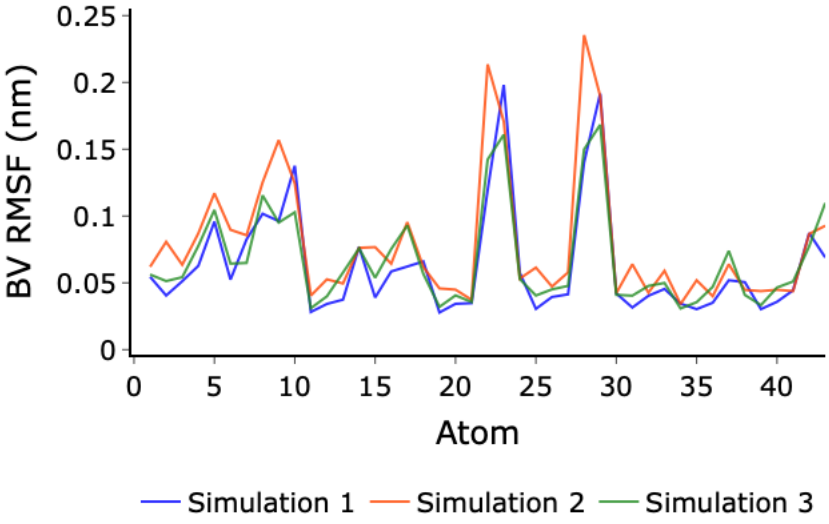


Figure S20: mRhubarb720 N224pCNPhe BV RMSF analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

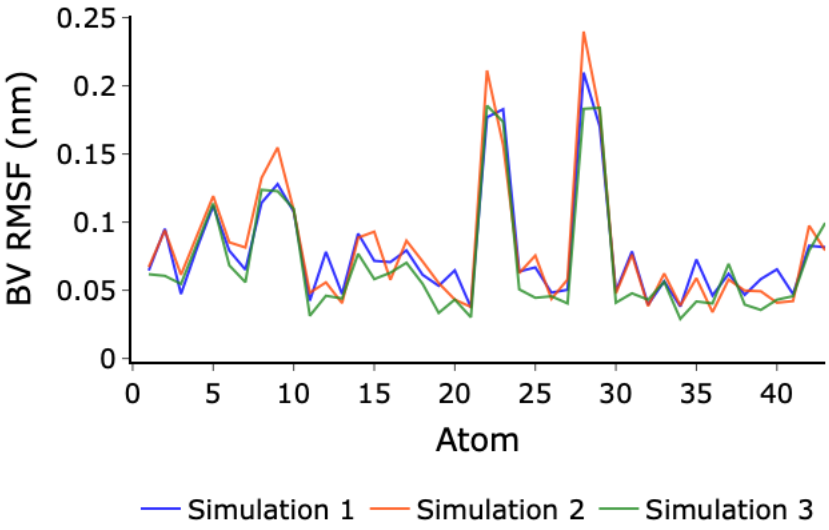


Figure S21: mRhubarb720 F258pCNPhe BV RMSF analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

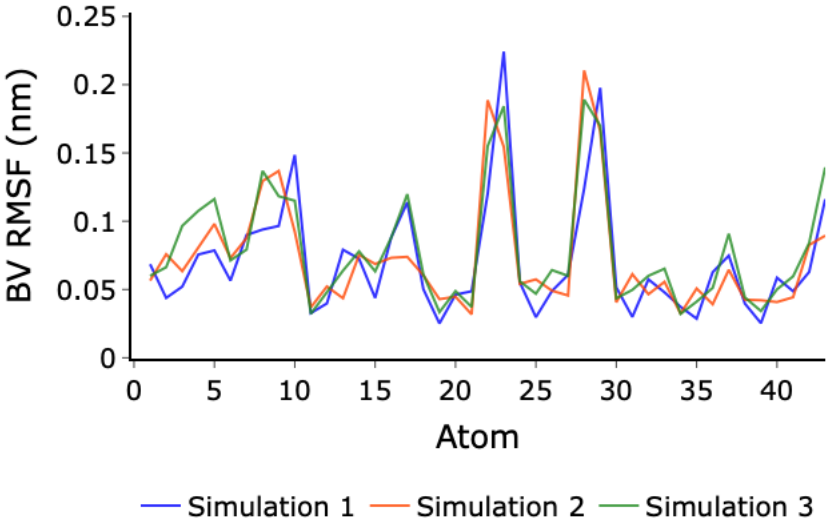


Figure S22: mRhubarb720 Y198/F258pCNPhe BV RMSF analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

mRhubarb720 H-bond analysis graphs

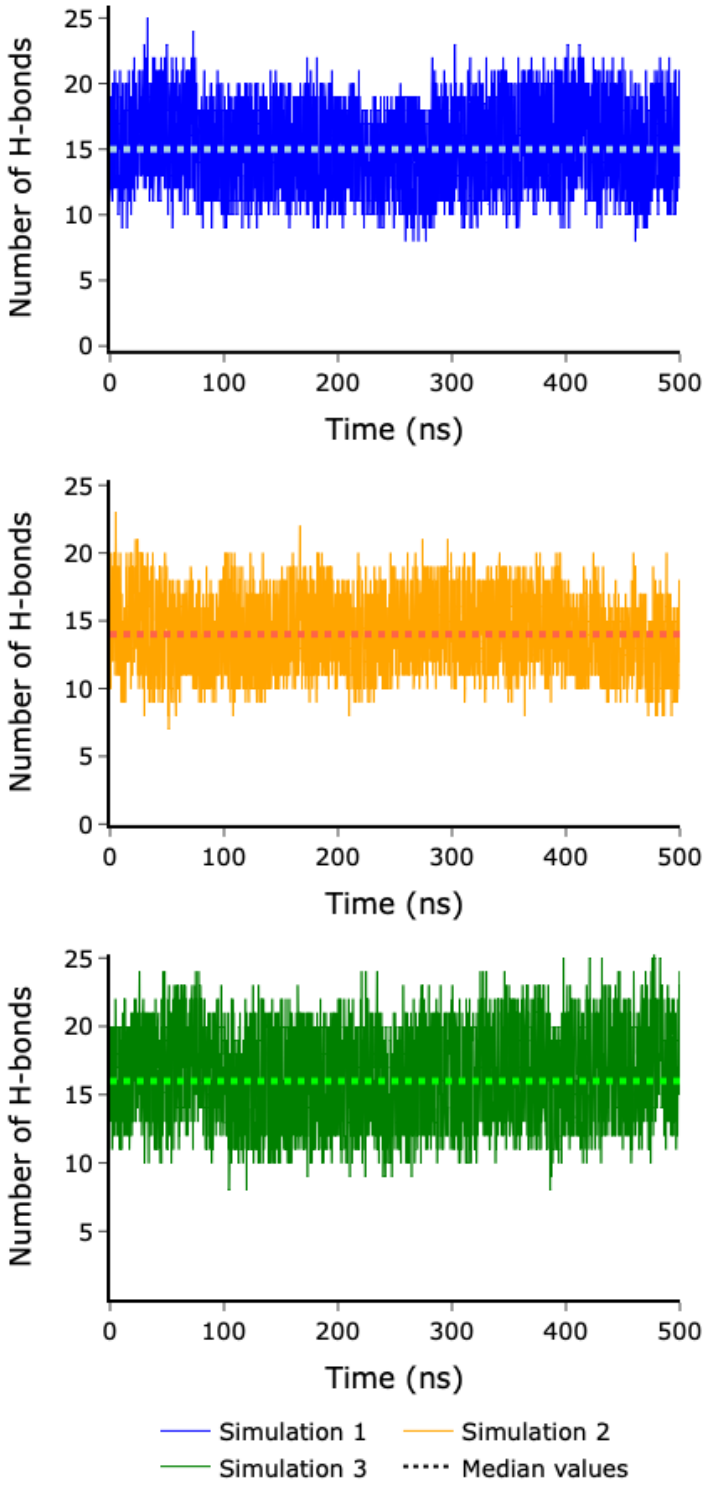


Figure S23: Total number of hydrogen bonds between BV and the surrounding environment in mRhubarb720 D177pCNPhe. These calculations considered the H-bonds between BV and the protein scaffold and solvent present in the system. Each simulation was plotted and was presented with each corresponding median values.

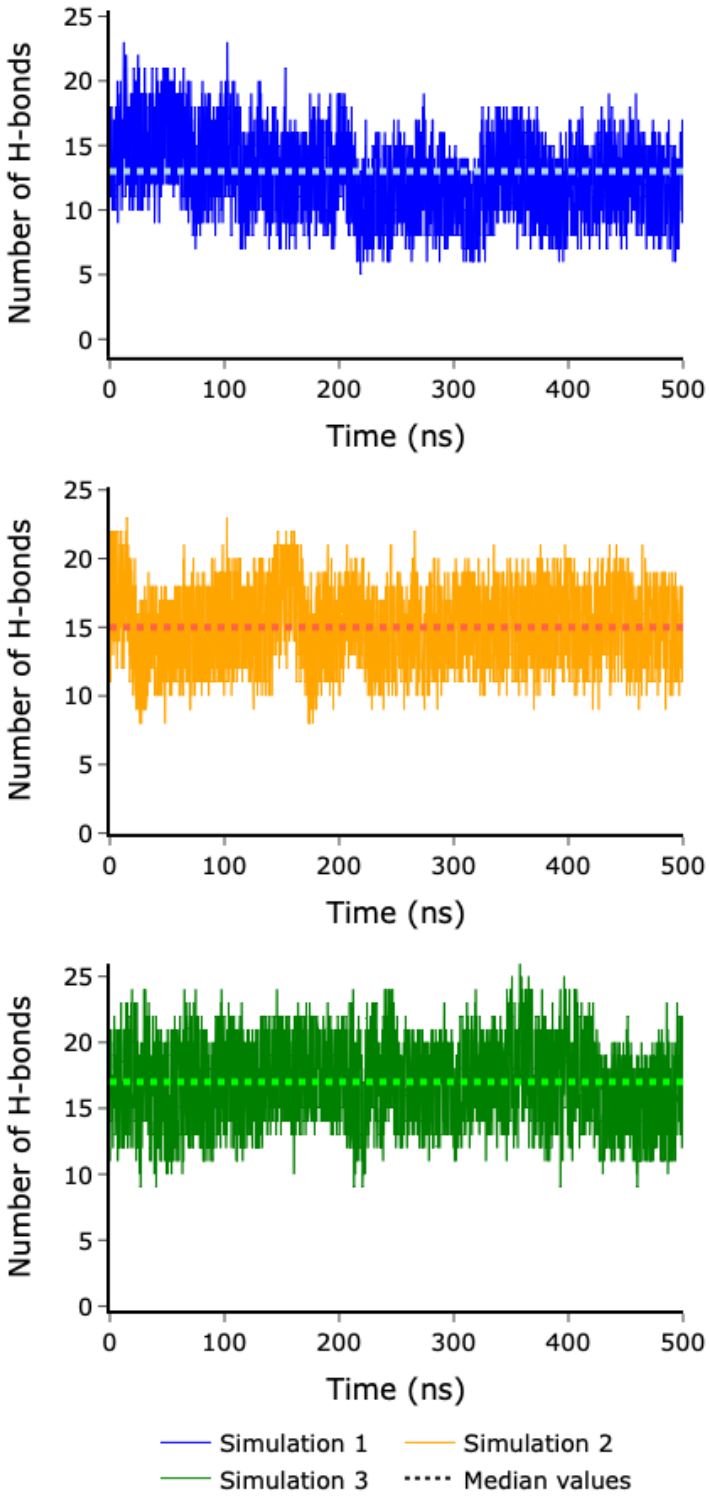


Figure S24: Total number of hydrogen bonds between BV and the surrounding environment in mRhubarb720 Y198pCNPhe. These calculations considered the H-bonds between BV and the protein scaffold and solvent present in the system. Each simulation was plotted and was presented with each corresponding median values.

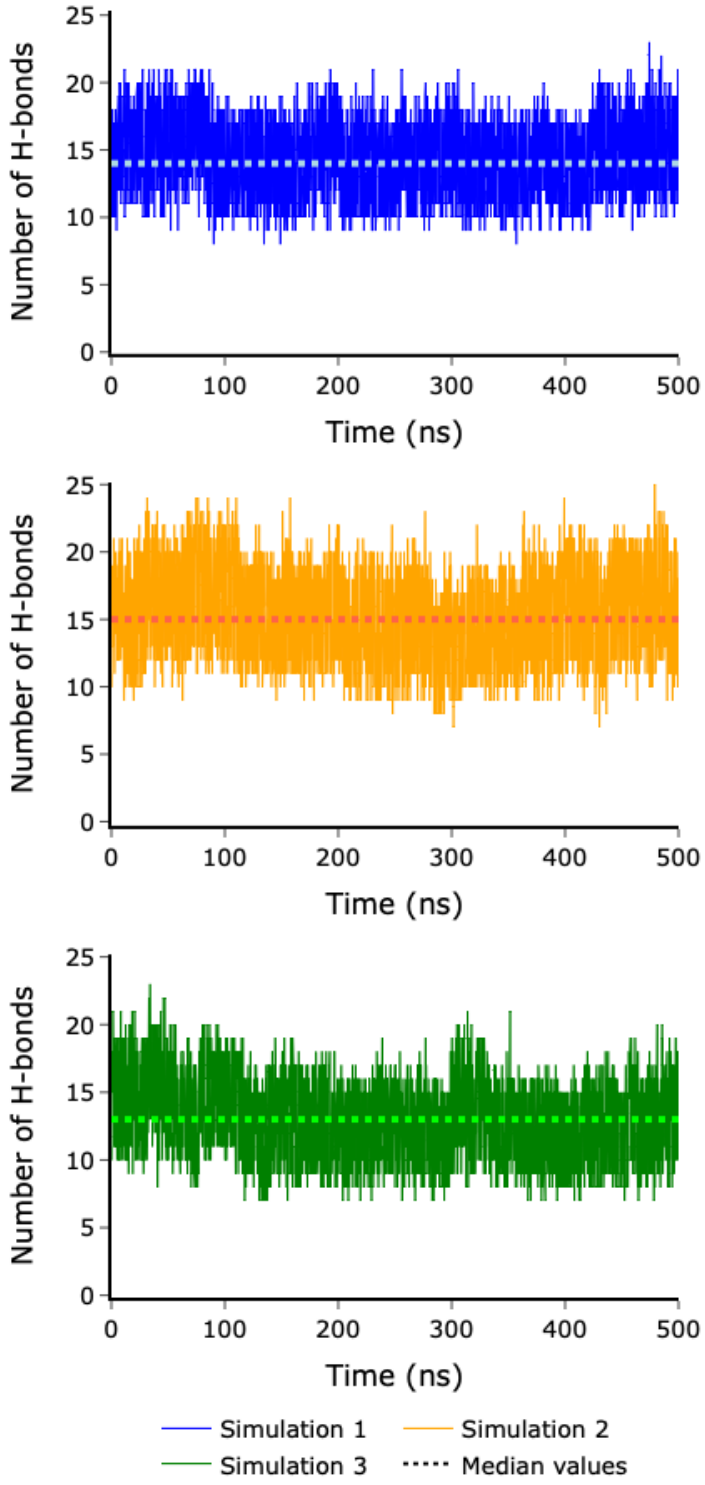


Figure S25: Total number of hydrogen bonds between BV and the surrounding environment in mRhubarb720 N224pCNPhe. These calculations considered the H-bonds between BV and the protein scaffold and solvent present in the system. Each simulation was plotted and was presented with each corresponding median values.

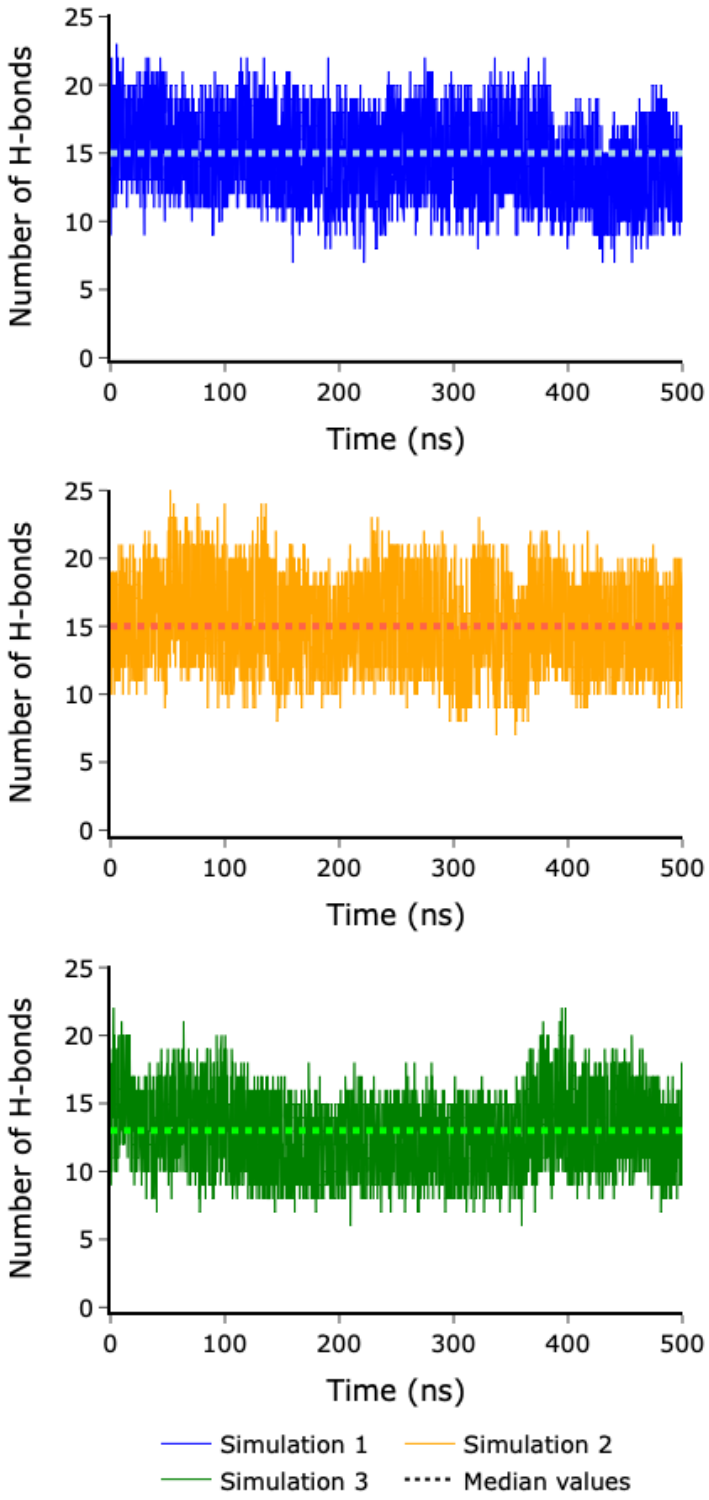


Figure S26: Total number of hydrogen bonds between BV and the surrounding environment in mRhubarb720 F258pCNPhe. These calculations considered the H-bonds between BV and the protein scaffold and solvent present in the system. Each simulation was plotted and was presented with each corresponding median values.

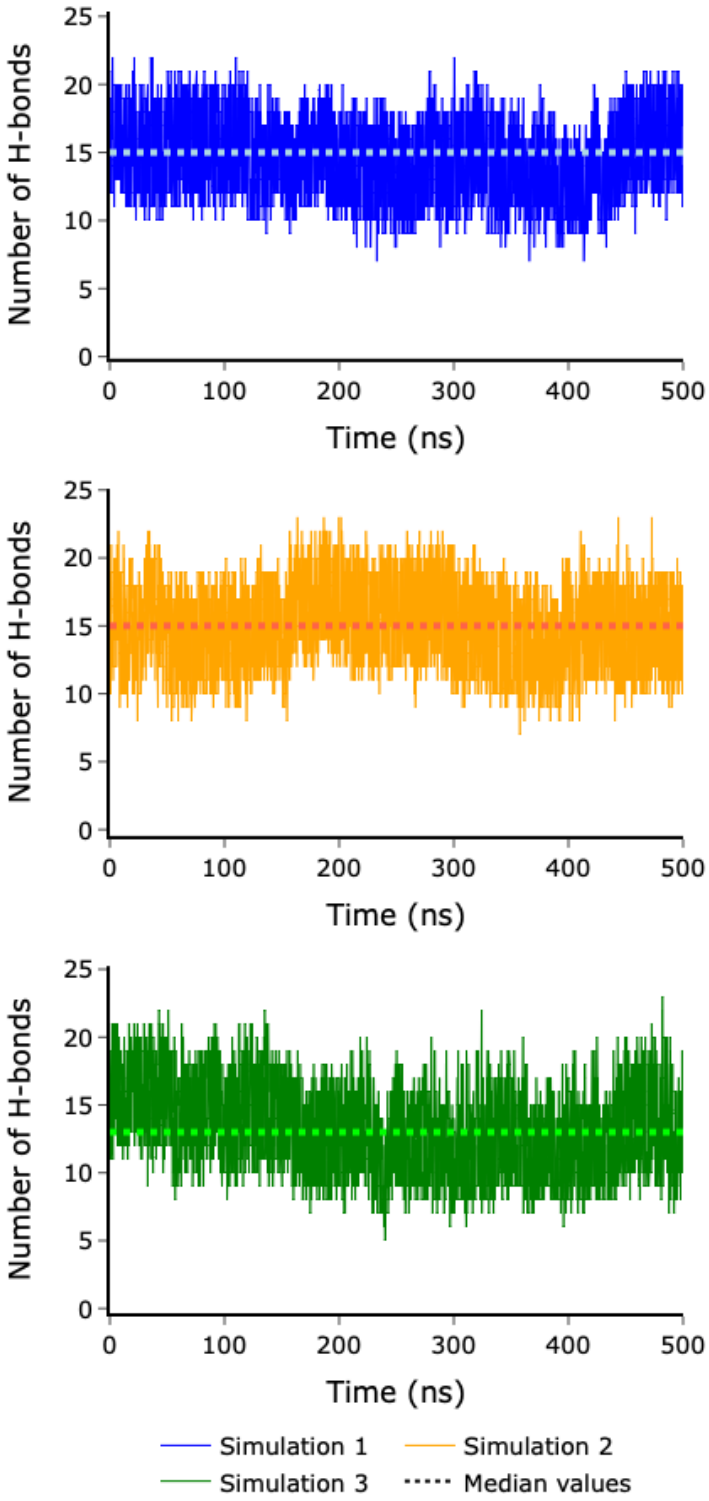


Figure S27: Total number of hydrogen bonds between BV and the surrounding environment in mRhubarb720 Y198/F258pCNPhe. These calculations considered the H-bonds between BV and the protein scaffold and solvent present in the system. Each simulation was plotted and was presented with each corresponding median values.

mRhubarb720 SASA graphs

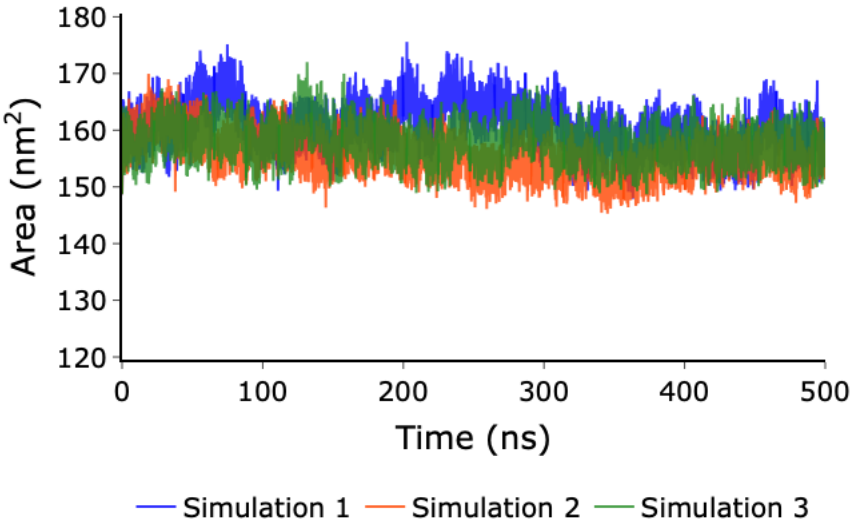


Figure S28: mRhubarb720 D177pCNPhe protein SASA analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

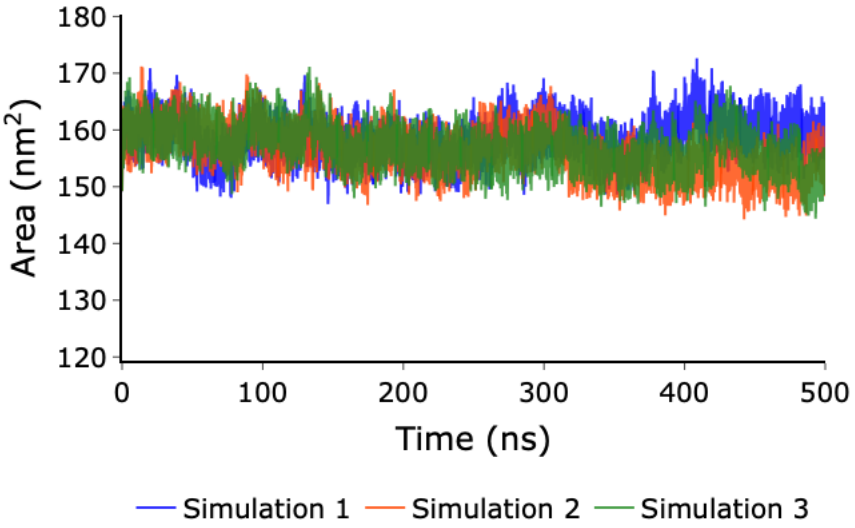


Figure S29: mRhubarb720 Y198pCNPhe protein SASA analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

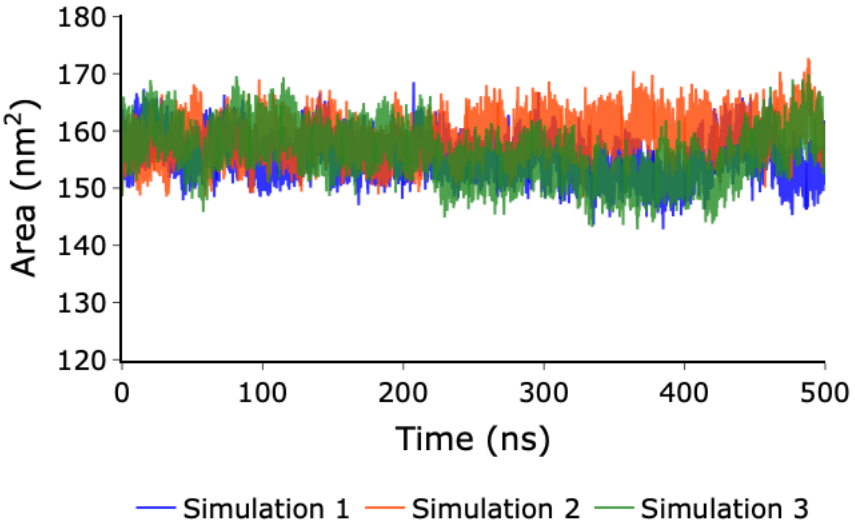


Figure S30: mRhubarb720 N224pCNPhe protein SASA analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

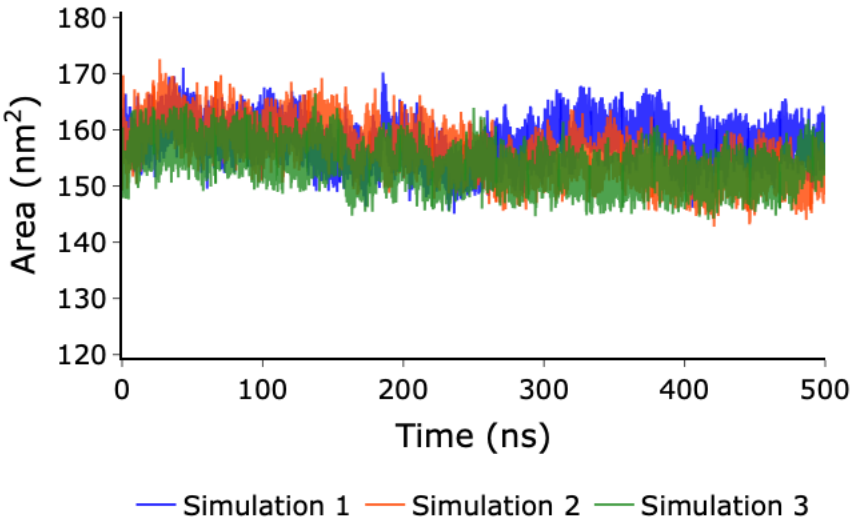


Figure S31: mRhubarb720 F258pCNPhe protein SASA analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

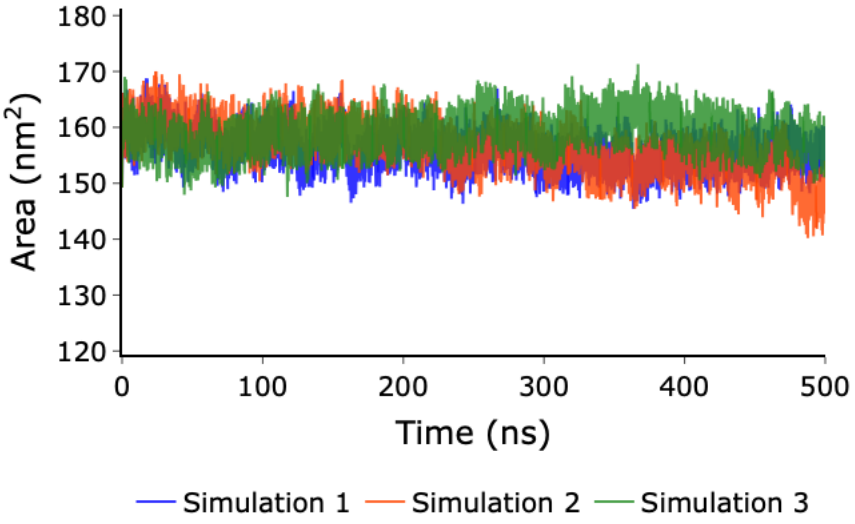


Figure S32: mRhubarb720 Y198/F258pCNPhe protein SASA analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

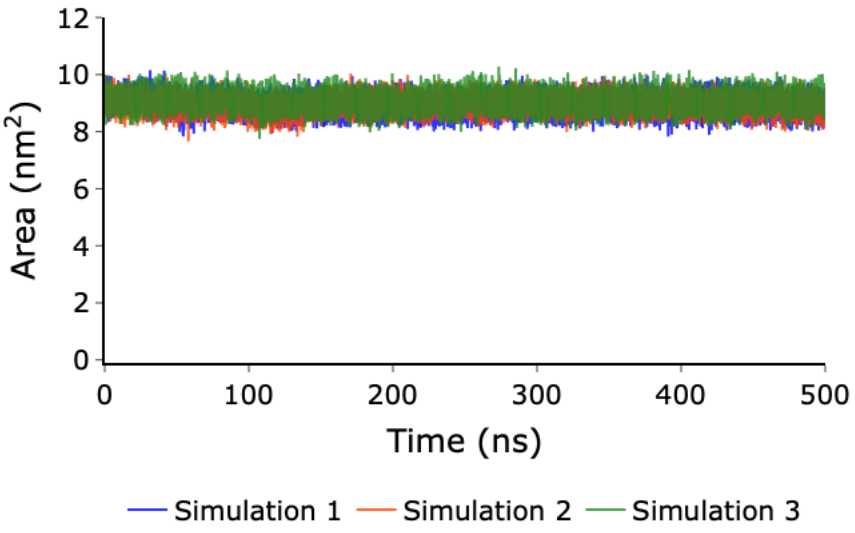


Figure S33: mRhubarb720 D177pCNPhe BV SASA analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

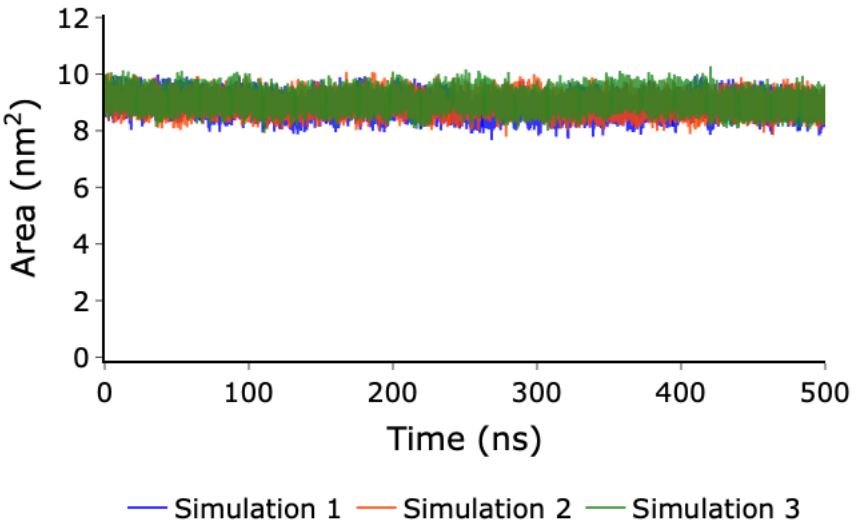


Figure S34: mRhubarb720 Y198pCNPhe BV SASA analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

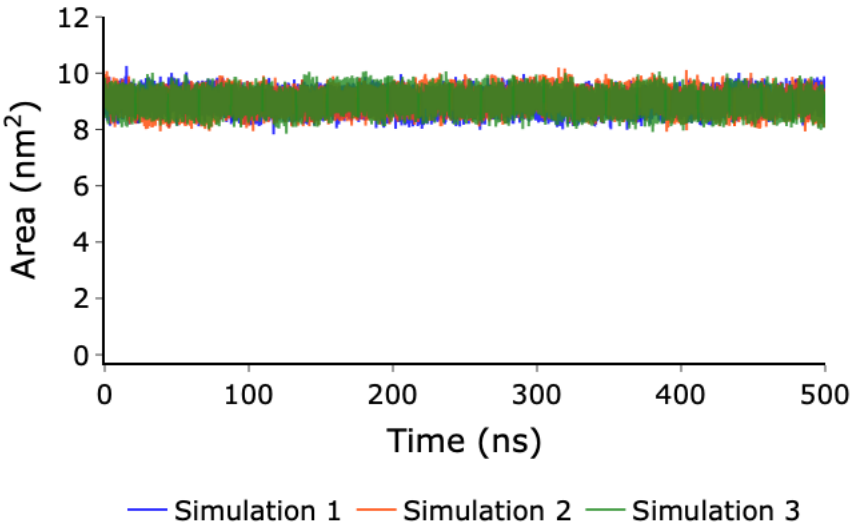


Figure S35: mRhubarb720 N224pCNPhe BV SASA analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

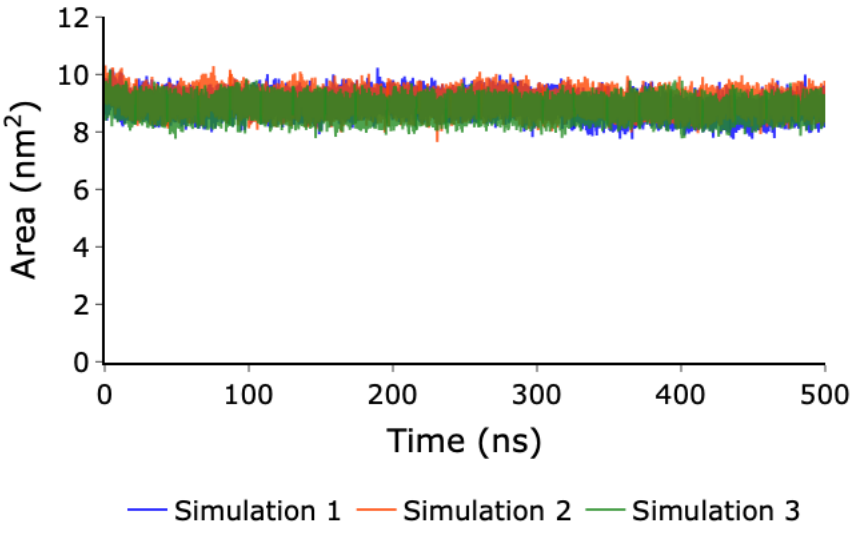


Figure S36: mRhubarb720 F258pCNPhe BV SASA analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

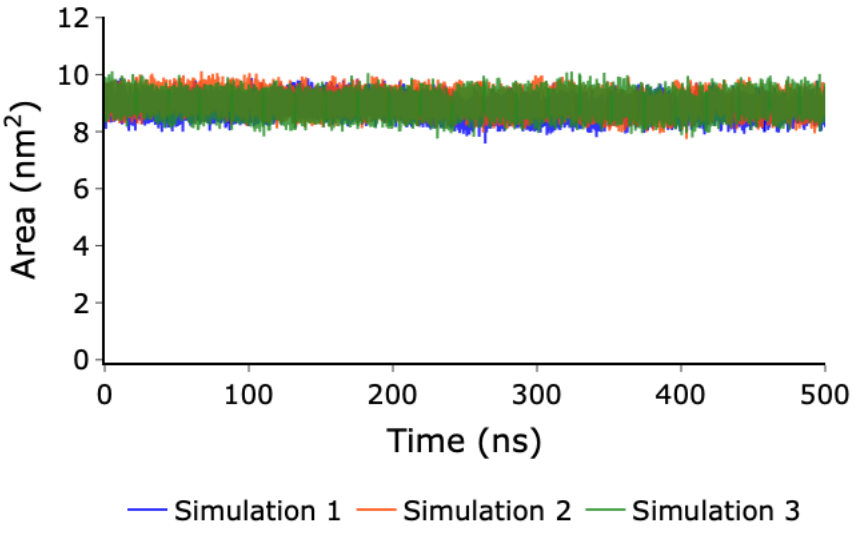


Figure S37: mRhubarb720 Y198/F258pCNPhe BV SASA analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

mRhubarb720 Rg graphs

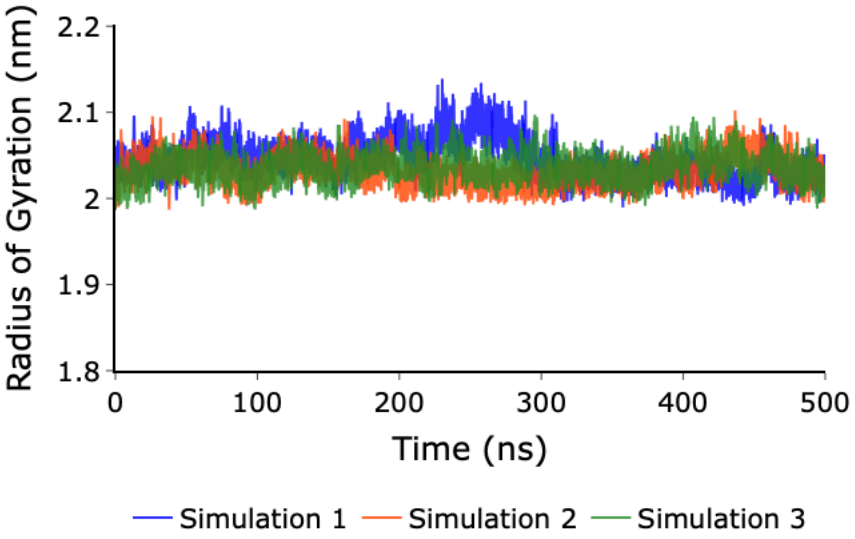


Figure S38: mRhubarb720 D177pCNPhe protein Rg analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

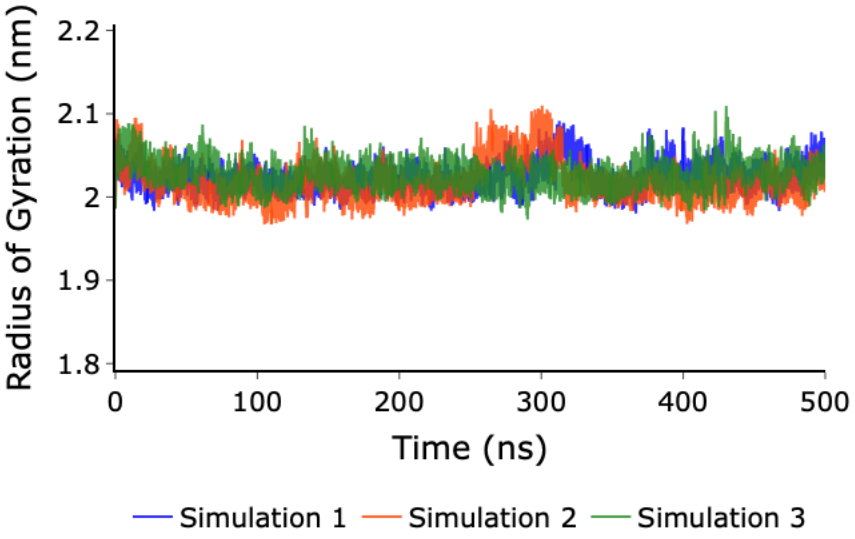


Figure S39: mRhubarb720 Y198pCNPhe protein Rg analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

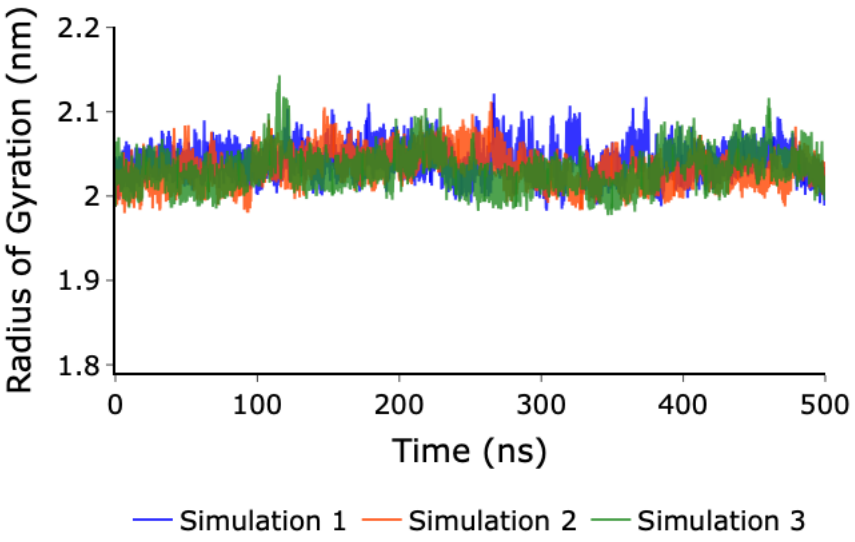


Figure S40: mRhubarb720 N224pCNPhe protein Rg analysis over three simulations Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

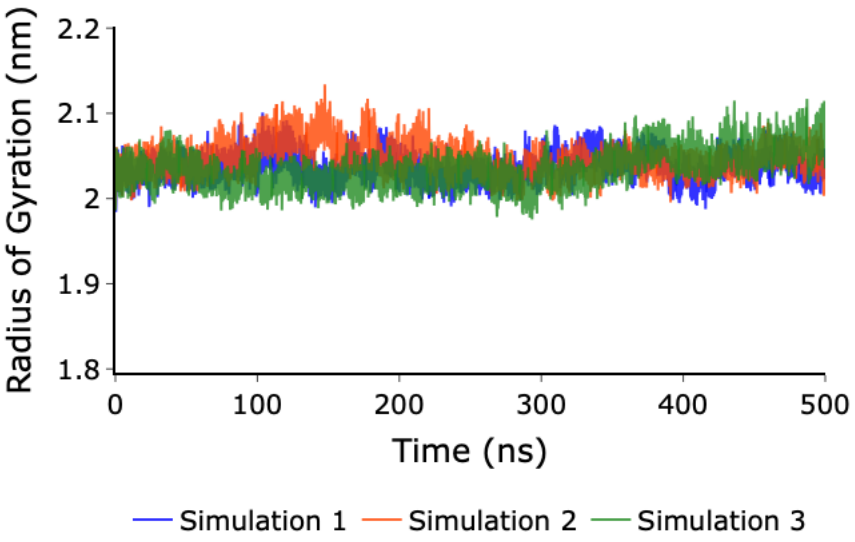


Figure S41: mRhubarb720 F258pCNPhe protein Rg analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

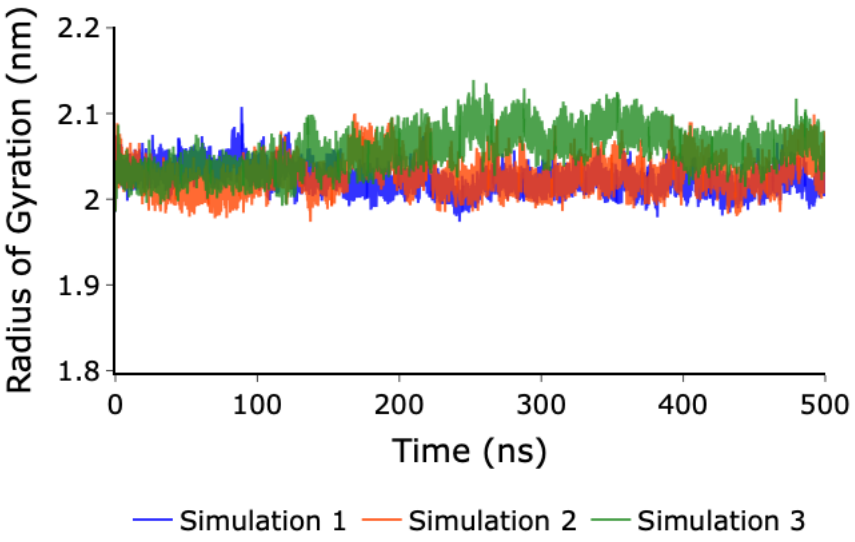


Figure S42: mRhubarb720 Y198/F258pCNPhe protein Rg analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

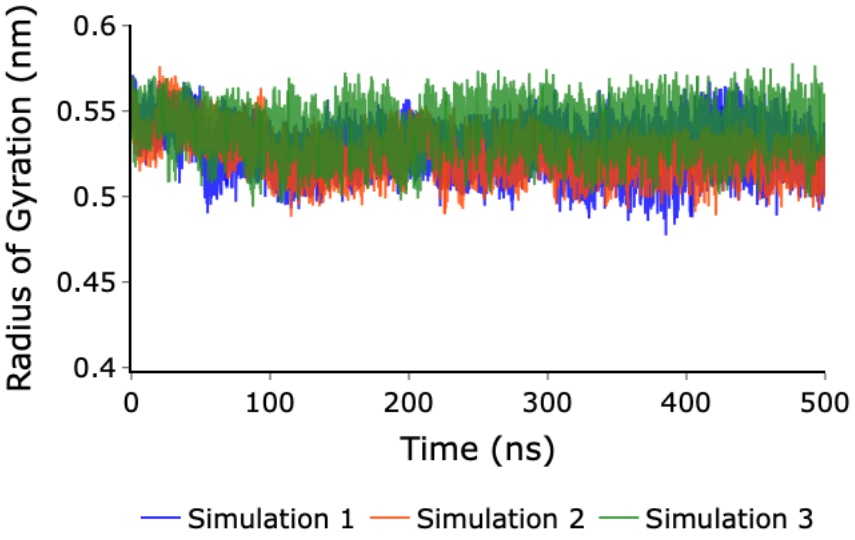


Figure S43: mRhubarb720 D177pCNPhe BV Rg analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

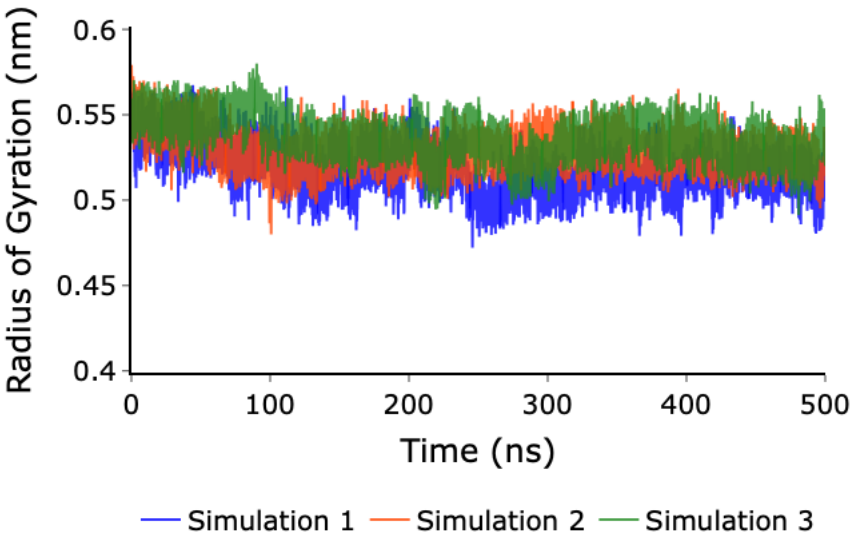


Figure S44: mRhubarb720 Y198pCNPhe BV Rg analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

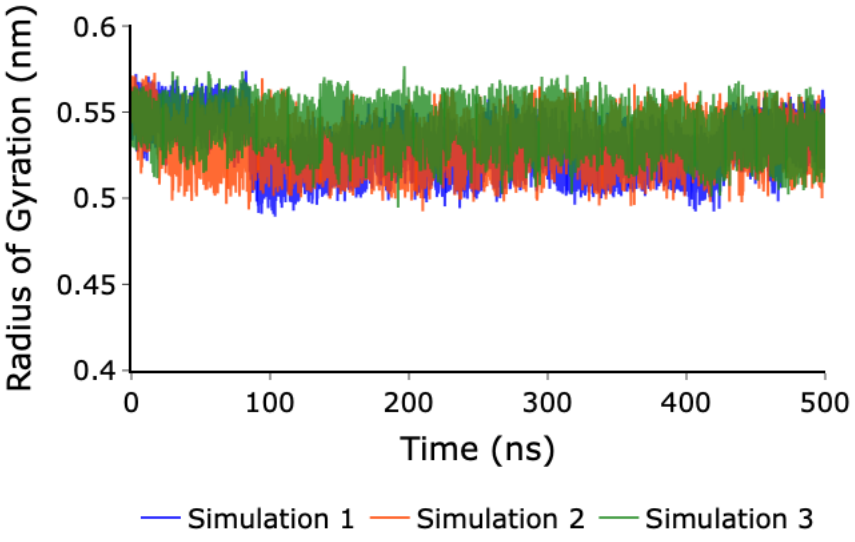


Figure S45: mRhubarb720 N224pCNPhe BV Rg analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

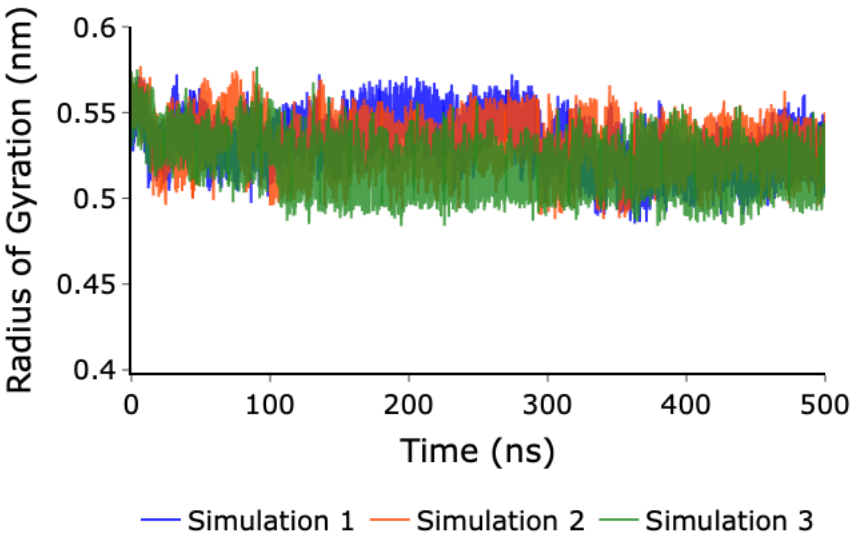


Figure S46: mRhubarb720 F258pCNPhe BV Rg analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

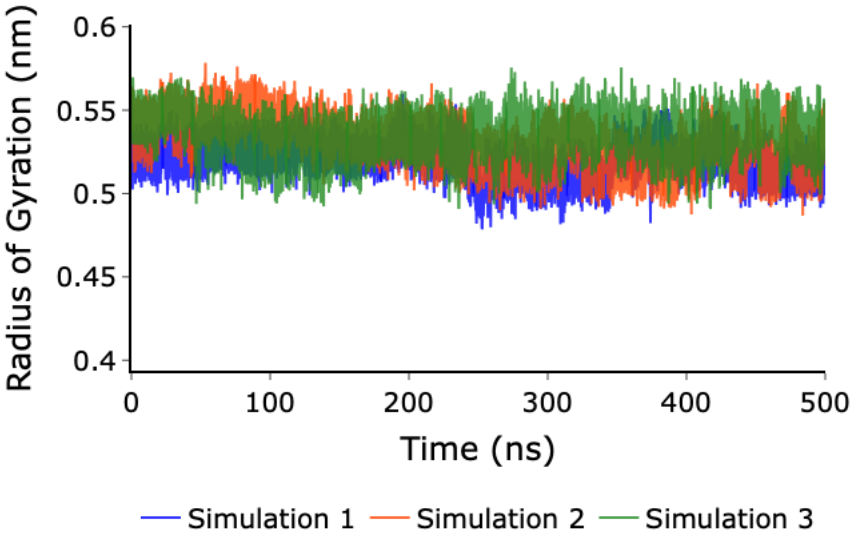


Figure S47: mRhubarb720 Y198/F258pCNPhe BV Rg analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

Free-BV MD graphs

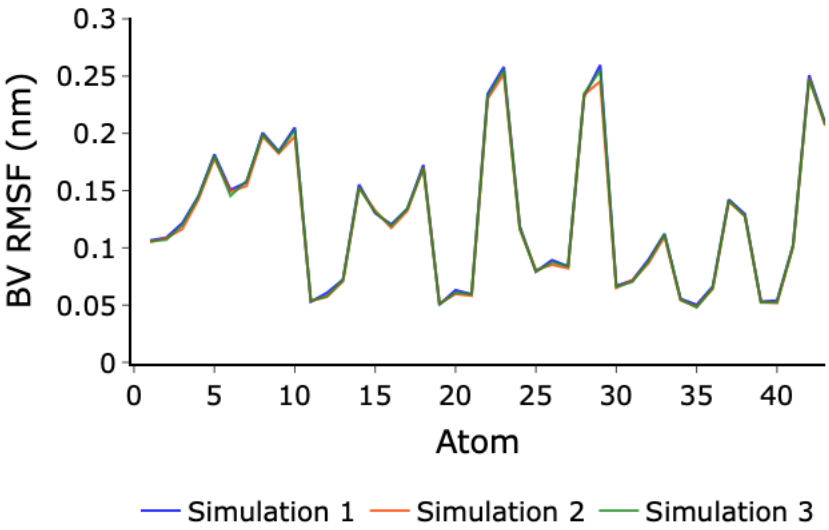


Figure S48: Free-BV RMSF analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

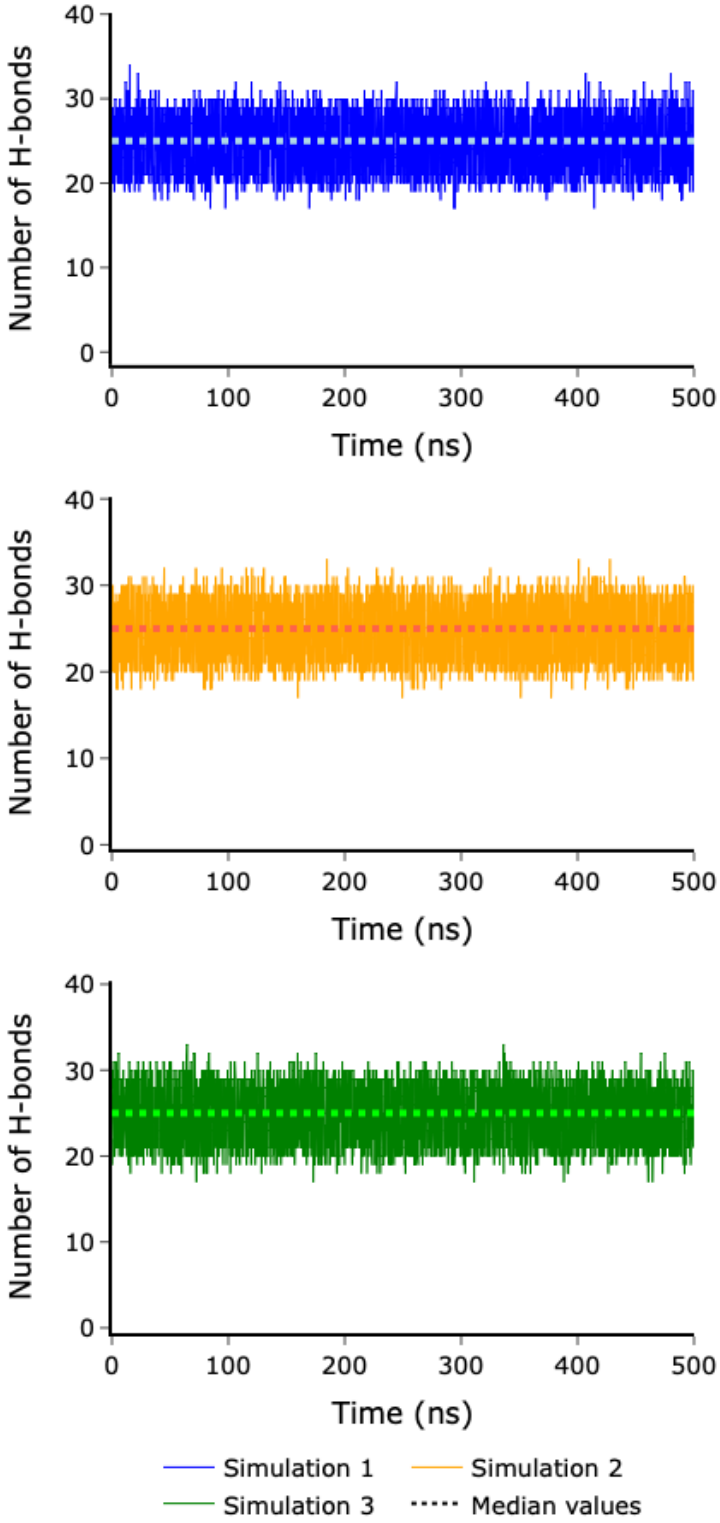


Figure S49: Total number of hydrogen bonds between Free-BV and the surrounding environment. These calculations considered the H-bonds between Free-BV and the solvent present in the system. Each simulation was plotted and was presented with each corresponding median values.

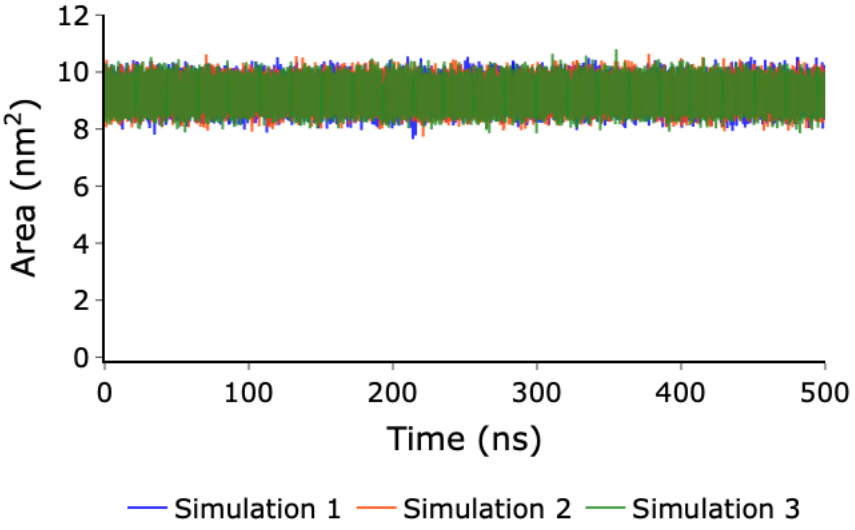


Figure S50: Free-BV SASA analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

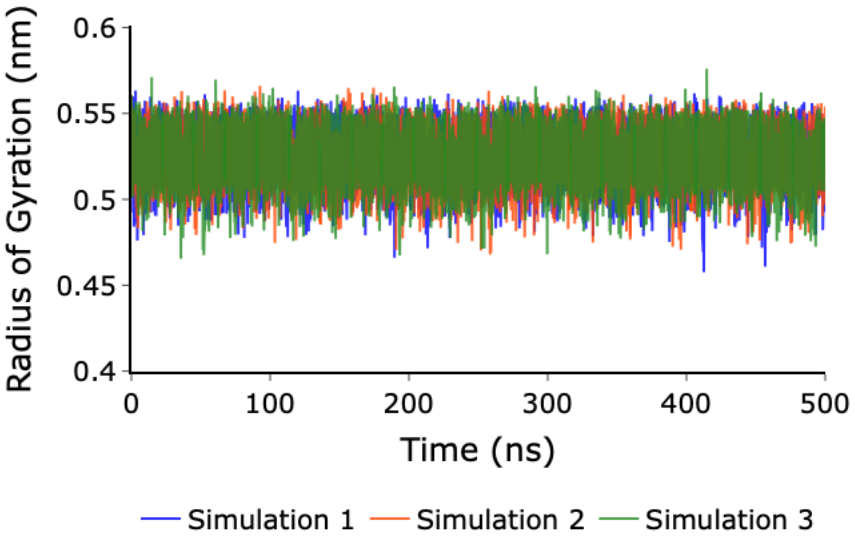


Figure S51: Free-BV Rg analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

mCherry RMSD graphs

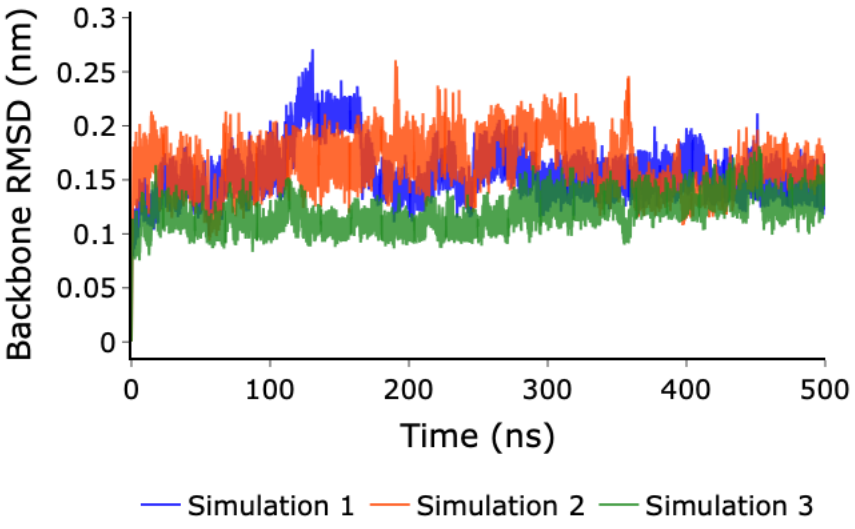


Figure S52: mCherry WT protein backbone RMSD analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

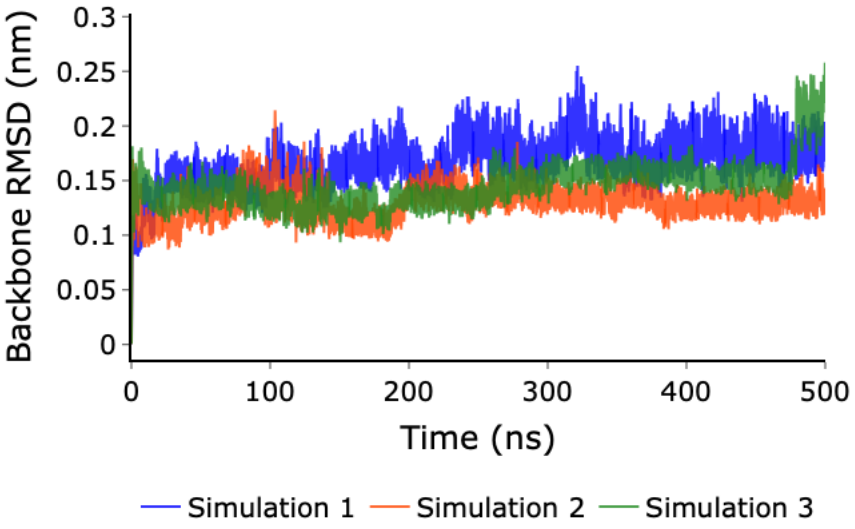


Figure S53: mCherry CRO-CN protein backbone RMSD analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

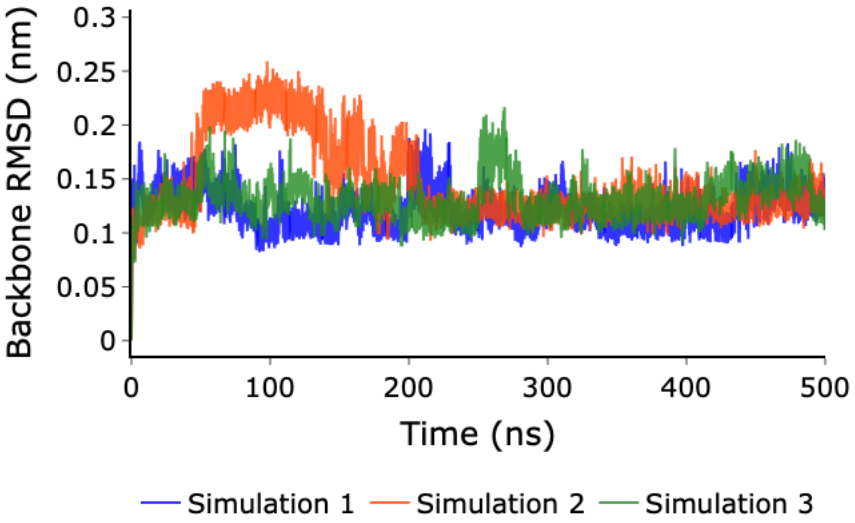


Figure S54: mCherry CRO-CC protein backbone RMSD analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

mCherry RMSF graphs

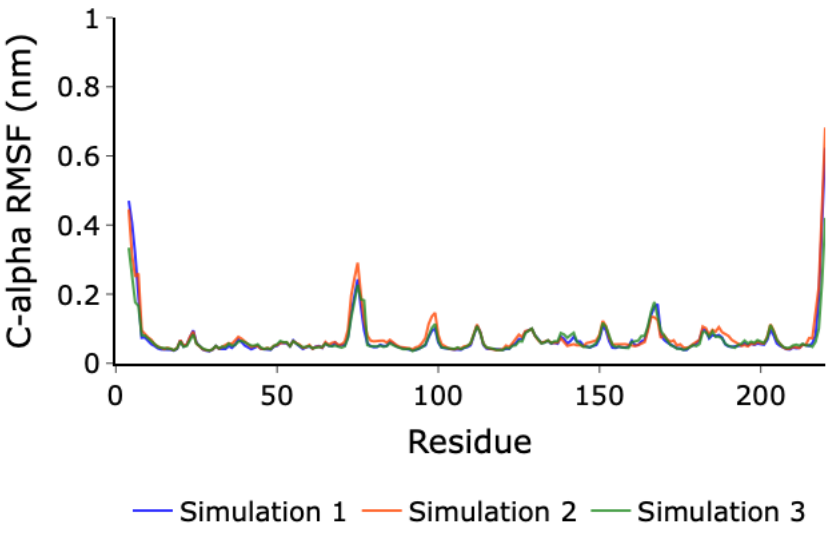


Figure S55: mCherry WT protein C-alpha RMSF analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

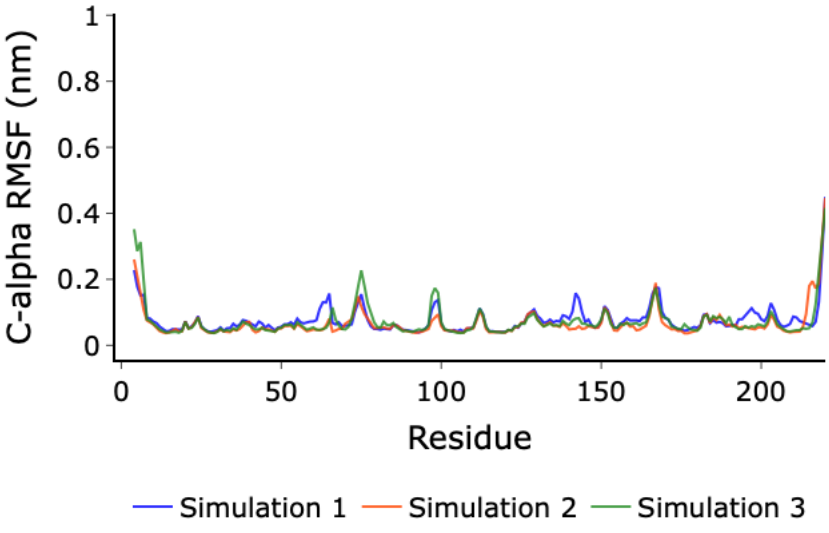


Figure S56: mCherry CRO-CN protein C-alpha RMSF analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

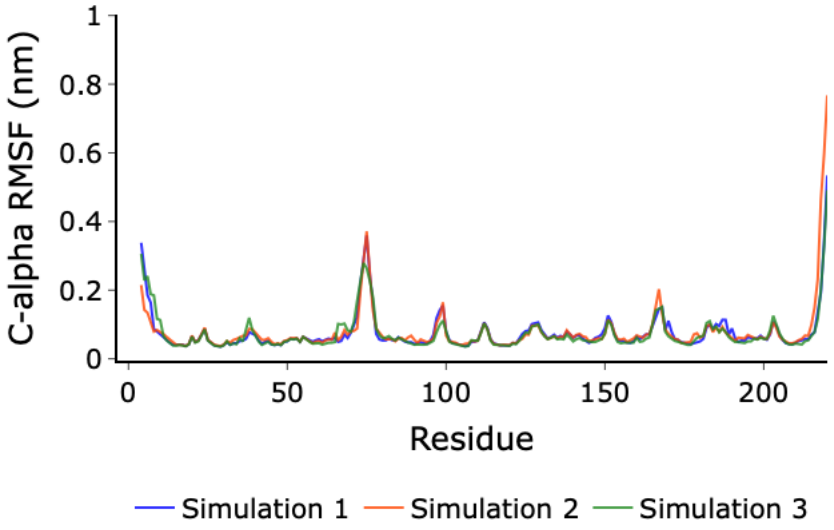


Figure S57: mCherry CRO-CC protein C-alpha RMSF analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

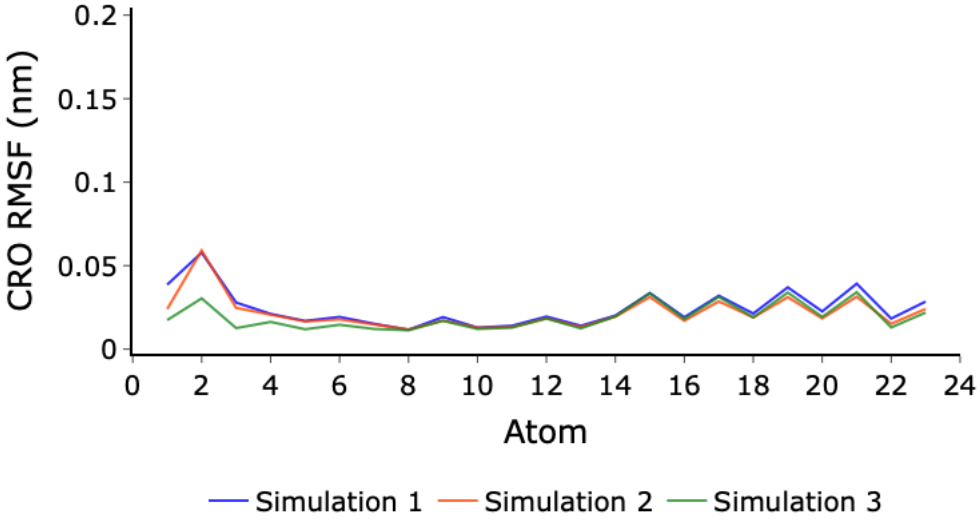


Figure S58: mCherry WT chromophore RMSF analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

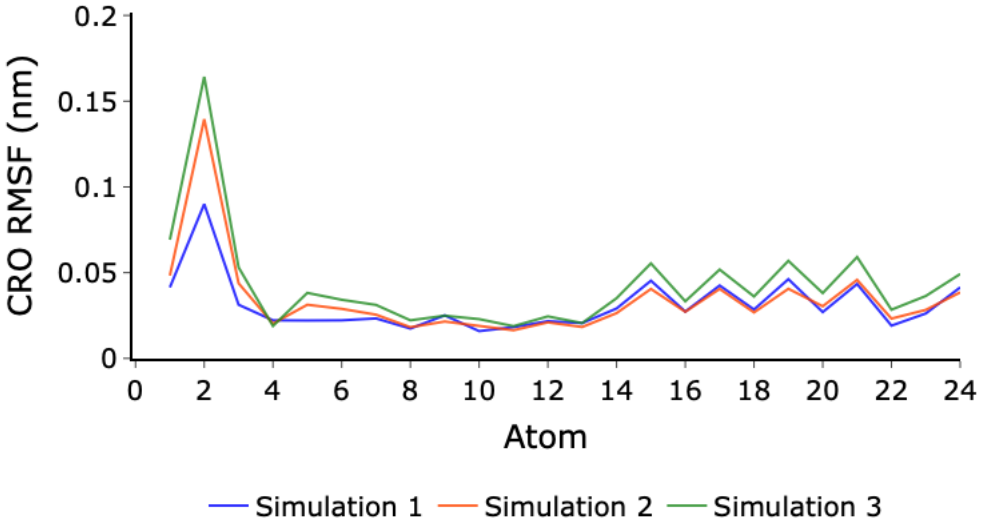


Figure S59: mCherry CRO-CN chromophore RMSF analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

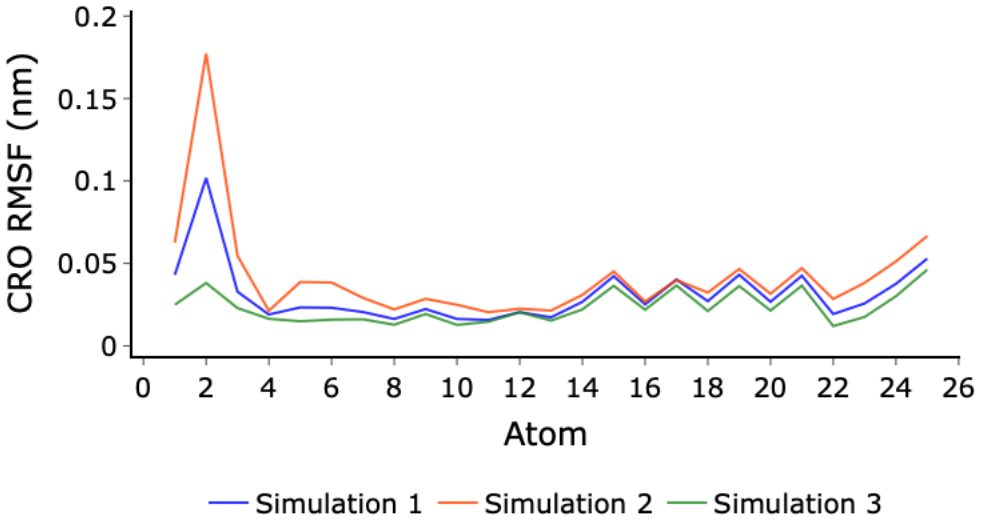


Figure S60: mCherry CRO-CC chromophore RMSF analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

mCherry SASA graphs

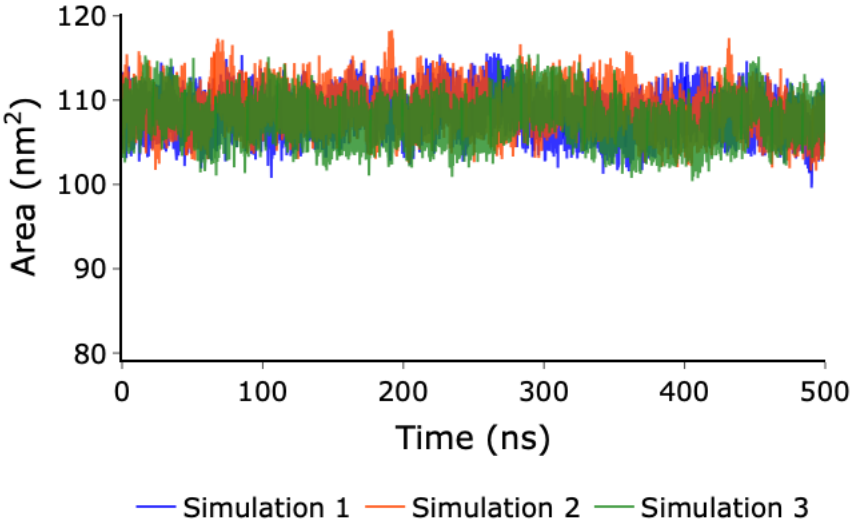


Figure S61: mCherry WT protein SASA analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

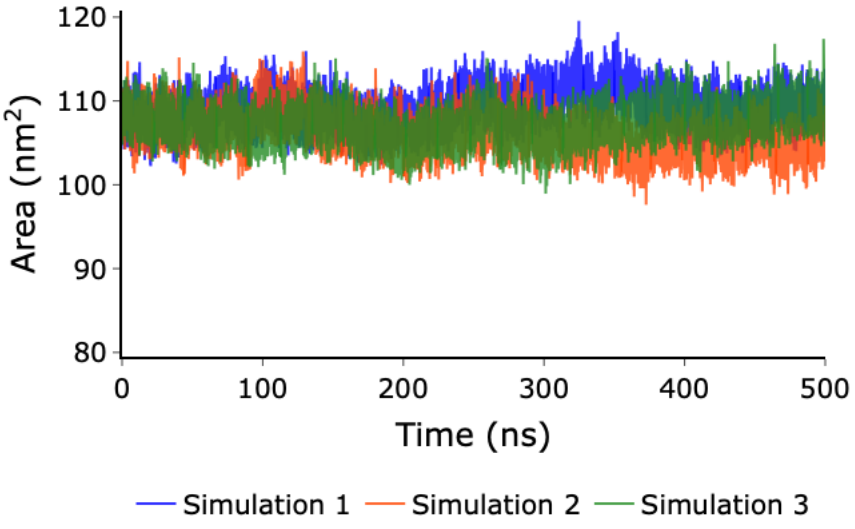


Figure S62: mCherry CRO-CN protein SASA analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

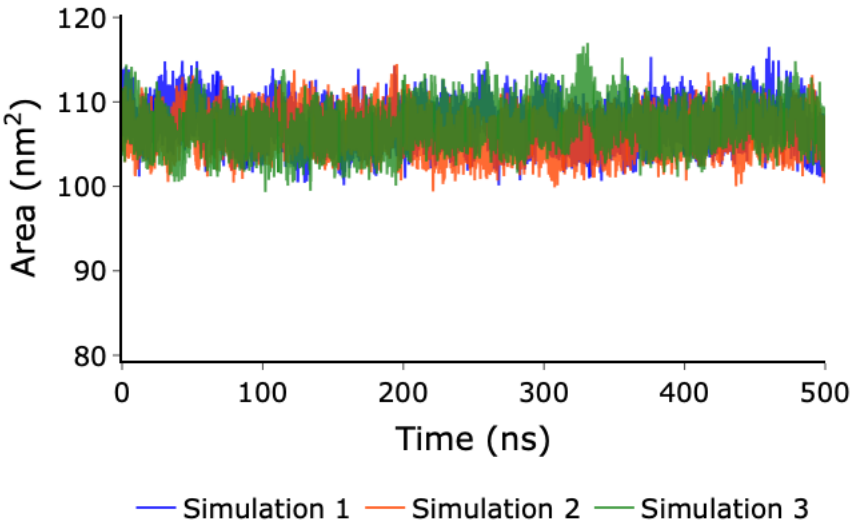


Figure S63: mCherry CRO-CC protein SASA analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

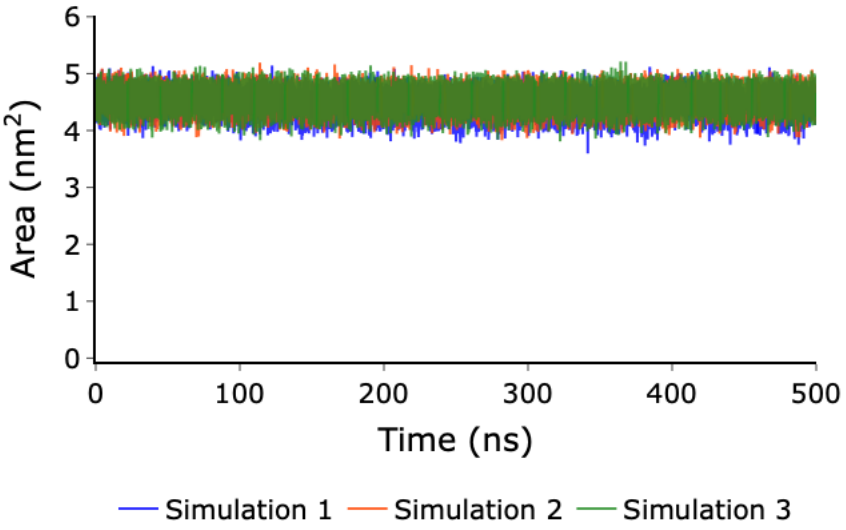


Figure S64: mCherry WT chromophore SASA analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

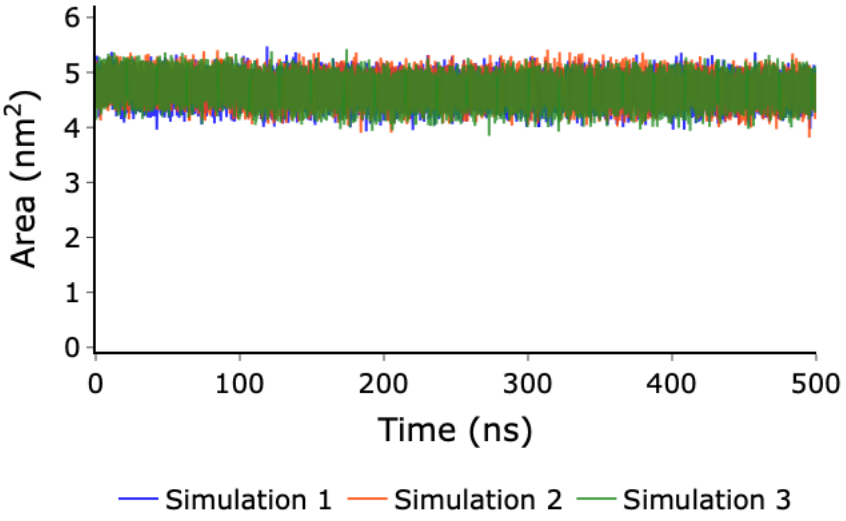


Figure S65: mCherry CRO-CN chromophore SASA analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

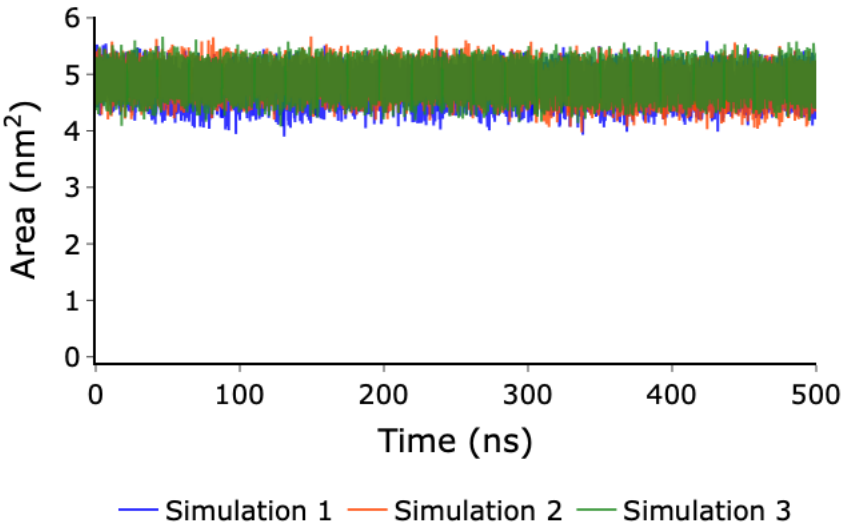


Figure S66: mCherry CRO-CC chromophore SASA analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

mCherry Rg graphs

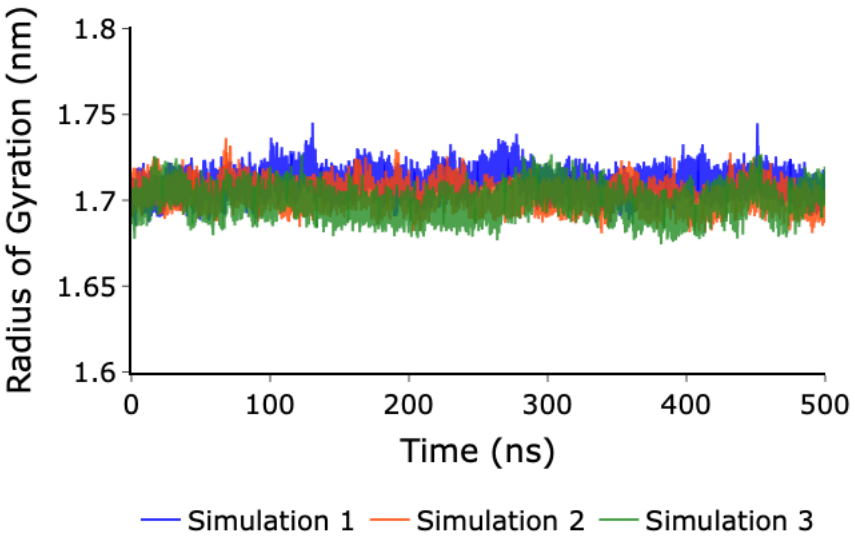


Figure S67: mCherry WT protein Rg analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

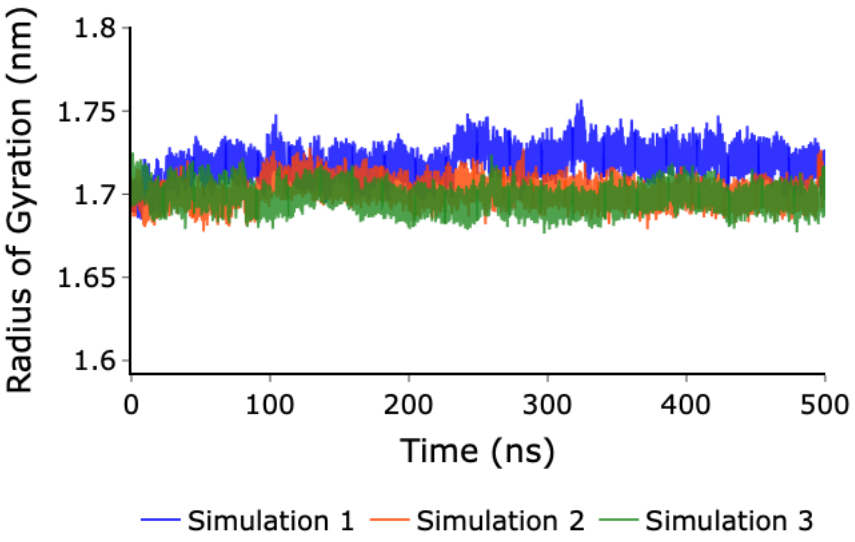


Figure S68: mCherry CRO-CN protein Rg analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

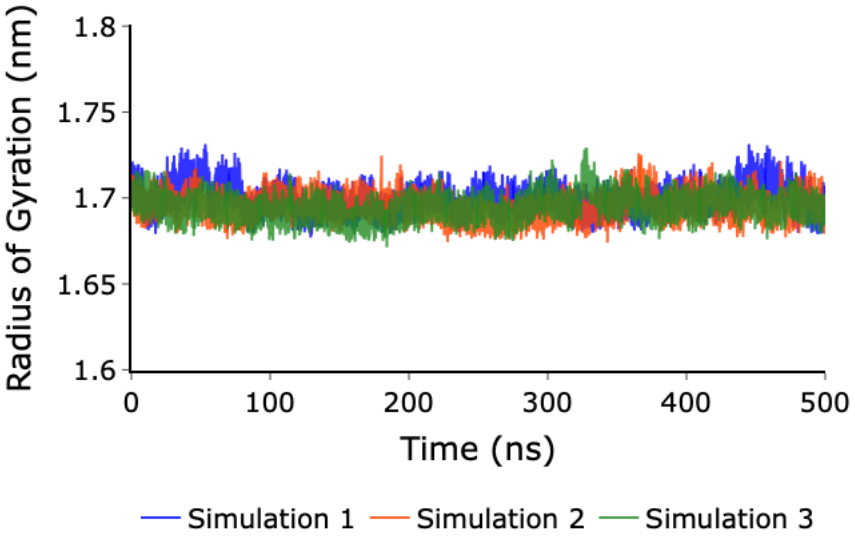


Figure S69: mCherry CRO-CC protein Rg analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

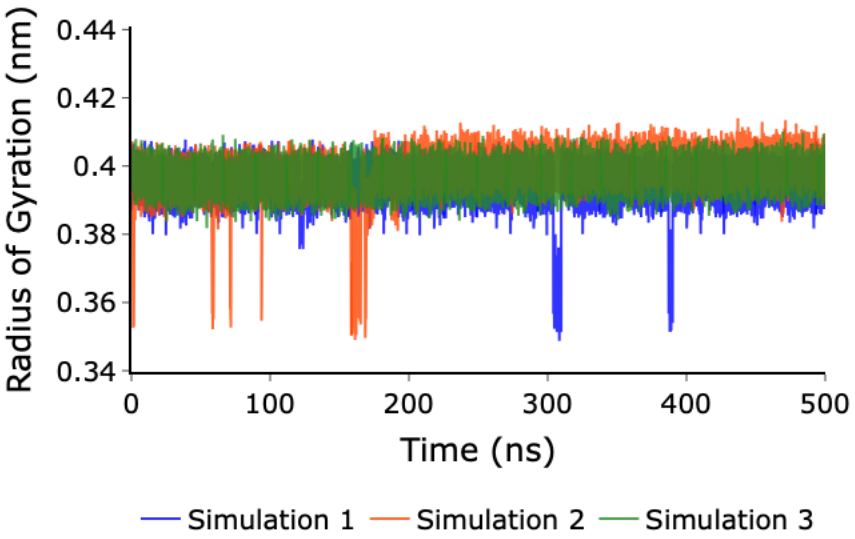


Figure S70: mCherry WT chromophore Rg analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

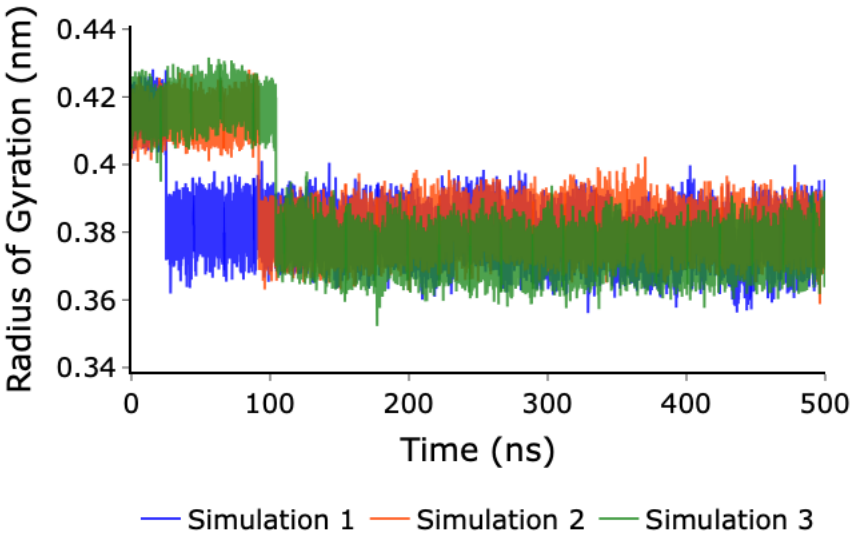


Figure S71: mCherry CRO-CN chromophore Rg analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

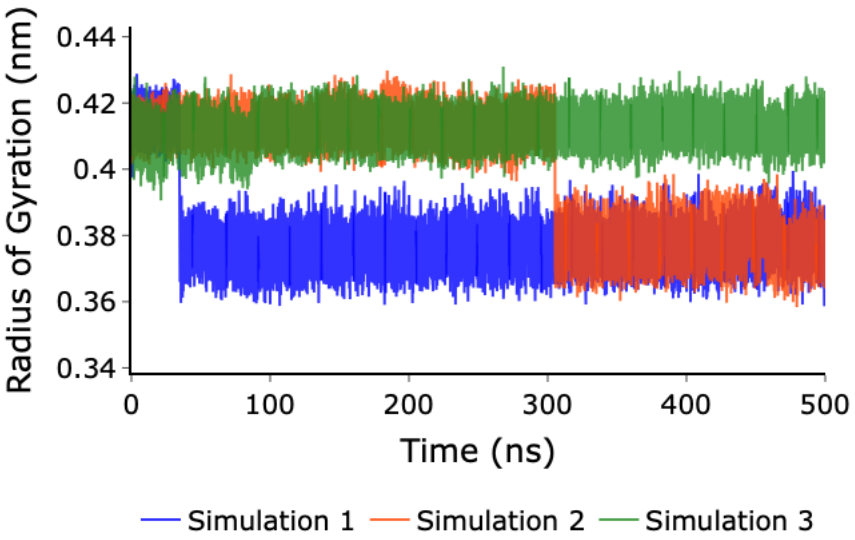


Figure S72: mCherry CRO-CC chromophore Rg analysis over three simulations. Simulation 1 (blue), Simulation 2 (orange), Simulation 3 (green).

BV dipole moment figure

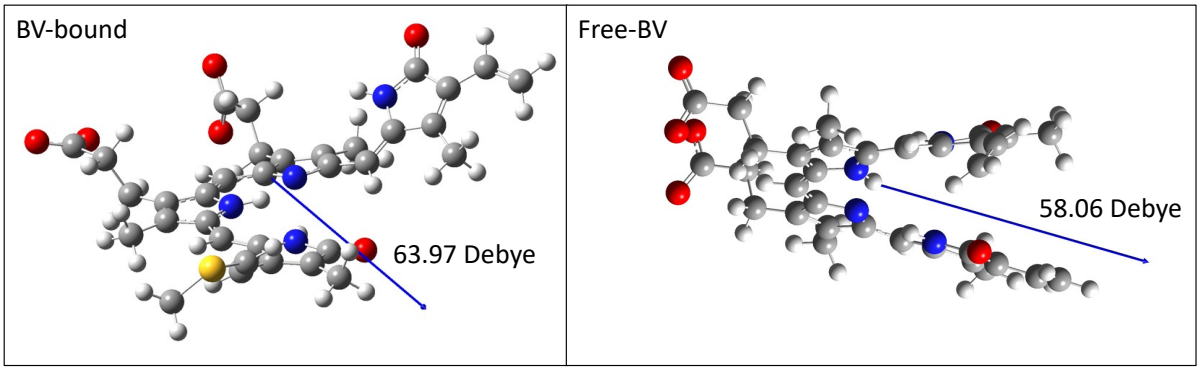
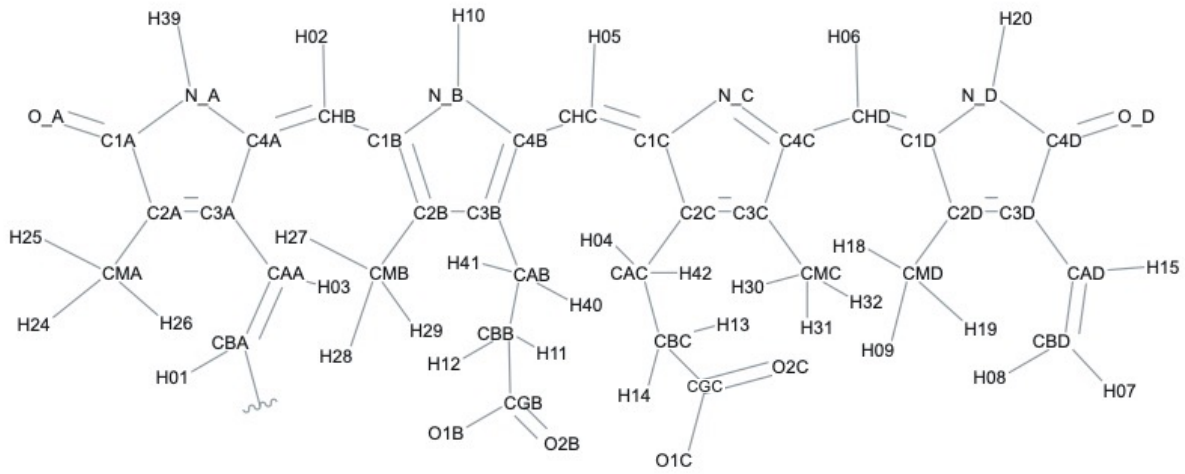


Figure S73: Dipole moments of BV-bound and Free-BV. Each box contains the dipole magnitude (in Debye; D) and the direction of the moment (blue arrow), which originates from the centre of electronic charge.

9.3 Parameters

All CHARMM27.ff – Section 2.6.3

Parameters 1: BV-Bound parameters




```
[ atom ]
;   id       type   charge       numb
     CAA      CHX    -0.109 0
     CAB      CMX    -0.0181      1
     CAC      CMX    -0.0292      2
     CAD      CMX    -0.109       3
     CBA      CBX    -0.21        4
     CBB      CMX    -0.1924      5
     CBC      CMX    -0.1904      6
     CBD      CMX    -0.206       7
     CGB      CGX    0.9036       8
     CGC      CGX    0.9026       9
     CHB      CFX    -0.0312      10
     CHC      CHX    0.0238       11
     CHD      CHX    -0.1876      12
     CMA      CMX    -0.0319      13
     CMB      CMX    -0.0668      14
     CMC      CMX    -0.0689      15
     CMD      CMX    -0.0589      16
     C1A      CAX    0.5985       17
     C1B      CBX    -0.1381      18
     C1C      CCX    -0.103       19
     C1D      CBX    0.029        20
     O1B      OCX    -0.8538      21
     O1C      OHX    -0.8543      22
     C2A      CAX    -0.1742      23
     C2B      CBX    -0.0843      24
     C2C      CCX    -0.0422      25
     C2D      CBX    -0.0282      26
     O2B      OHX    -0.8538      27
     O2C      OCX    -0.8543      28
     C3A      CAX    -0.003       29
     C3B      CBX    -0.0733      30
     C3C      CCX    -0.1552      31
     C3D      CBX    -0.138       32
     C4A      CAX    -0.048       33
     C4B      CBX    -0.0941      34
```

```
    C4C      CCX    0.0956       35
    C4D      CBX    0.6155       36
    N_A      N1X    -0.3975      37
    N_B      N2X    -0.1401      38
    N_C      N3X    -0.188       39
    N_D      N1X    -0.4065      40
    O_A      OCX    -0.6345      41
    O_D      OCX    -0.6185      42
    H15      HCX    0.111        43
    H08      HCX    0.1225       44
    H09      HCX    0.0557       45
    H18      HCX    0.0557       46
    H19      HCX    0.0557       47
    H01      HHX    0.111        48
    H02      HHX    0.147        49
    H03      HHX    0.136        50
    H04      HCX    0.0567       51
    H05      HHX    0.161        52
    H06      HHX    0.146        53
    H10      HNX    0.3177       54
    H11      HNX    0.0327       55
    H12      HNX    0.0327       56
    H13      HCX    0.0317       57
    H14      HCX    0.0317       58
    H20      HNX    0.3645       59
    H07      HCX    0.1225       60
    H24      HCX    0.0447       61
    H25      HCX    0.0447       62
    H26      HCX    0.0447       63
    H27      HCX    0.0607       64
    H28      HCX    0.0607       65
    H29      HCX    0.0607       66
    H30      HCX    0.0657       67
    H31      HCX    0.0657       68
    H32      HCX    0.0657       69
    H39      HNX    0.3395       70
    H40      HCX    0.0927       71
    H41      HCX    0.0927       72
    H42      HCX    0.0937       73

[ atomtypes ]
;   type     mass
    CHX      12.011000
    CMX      12.011000
    CAX      12.011000
    CBX      12.011000
    CCX      12.011000
    CFX      12.011000
    CGX      12.011000
    OCX      15.999400
    OHX      15.999400
    N1X      14.007000
    N2X      14.007000
    N3X      14.007000
    HHX      1.008000
    HNX      1.008000
    HOX      1.008000
   HCX      1.008000


[ bondtypes ]
;    i        j      func   b0           kb
    CAX      CAX    1      0.14676      310450
    CAX      N1X    1      0.13789      357820
    CAX      OCX    1      0.12183      533630
```

```
     CBX      SM     1      0.18392       180670
     CBX      HHX    1      0.10868       288360
     CBX      N2X    1      0.14015       332540
     CBX      CBX    1      0.15095       273470
     CBX      HCX    1      0.10969       276650
     CBX      N1X    1      0.13789       357820
     CBX      OCX    1      0.12183       533630
     CCX      CCX    1      0.14278       351290
     CCX      N3X    1      0.13694       369110
     CFX      CBX    1      0.13461       457980
     CFX      CAX    1      0.13656       429280
     CFX      HHX    1      0.10883       286600
     CGX      OCX    1      0.12183       533630
     CGX      OHX    1      0.13513       334800
     CHX      CBX    1      0.13461       457980
     CHX      CAX    1      0.1454        323760
     CHX      HHX    1      0.10883       286600
     CHX      CCX    1      0.13656       429280
     CMX      CMX    1      0.15375       251790
     CMX      CBX    1      0.15095       273470
     CMX      HCX    1      0.10969       276650
     CMX      CCX    1      0.15015       280160
     CMX      CGX    1      0.15241       261920
     CMX      HNX    1      0.10969       276650
     CMX      CAX    1      0.15015       280160
     N1X      HNX    1      0.10129       337400
     N2X      HNX    1      0.101         341750
     OHX      HOX    1      0.0973        310790
     CHX      HCX    1      0.0883        286600

[ angletypes ]
;    i        j      k      func   th0    cth
     CHX      CBX    SM     5      122.35 525.51
     CHX      CBX    HHX    5      122.31 410.87
     CHX      CAX    CAX    5      128.05 533.04
     CMX      CMX    CGX    5      111.04 529.69
     CMX      CMX    HNX    5      109.8  387.44
     CMX      CBX    CBX    5      123.63 536.39
     CMX      CMX    HCX    5      109.8  387.44
     CMX      CCX    CCX    5      115.97 540.57
     CMX      CBX    HCX    5      109.8  387.44
     CT2      SM     CBX    5      100.37 514.63
     CBX      CHX    CAX    5      123.32 548.11
     CBX      CHX    HHX    5      119.94 415.05
     CMX      CMX    CBX    5      111.56 530.53
     CMX      CGX    OCX    5      123.2  564.01
     CMX      CGX    OHX    5      112.73 572.37
     CMX      CMX    CCX    5      111.93 531.37
     CGX      CMX    HNX    5      108.77 392.46
     CGX      OHX    HOX    5      106.55 417.56
     CGX      CMX    HCX    5      108.77 392.46
     SM       CBX    HHX    5      116.67 361.5
     CFX      CBX    CBX    5      117    580.74
     CFX      CBX    N2X    5      124.17 570.7
     CFX      CAX    CAX    5      122.72 551.45
     CFX      CAX    N1X    5      124.2  571.53
     CHX      CCX    CCX    5      122.72 551.45
     CHX      CCX    N3X    5      123.98 574.88
     CHX      CBX    CBX    5      117    580.74
     CHX      CBX    N2X    5      124.17 570.7
     CHX      CBX    N1X    5      120.71 582.41
     CMX      CAX    CAX    5      117.76 529.69
     CAX      CAX    CAX    5      121.35 544.76
     CAX      N1X    CAX    5      123.27 541.41
     CAX      N1X    HNX    5      117.55 404.17
```

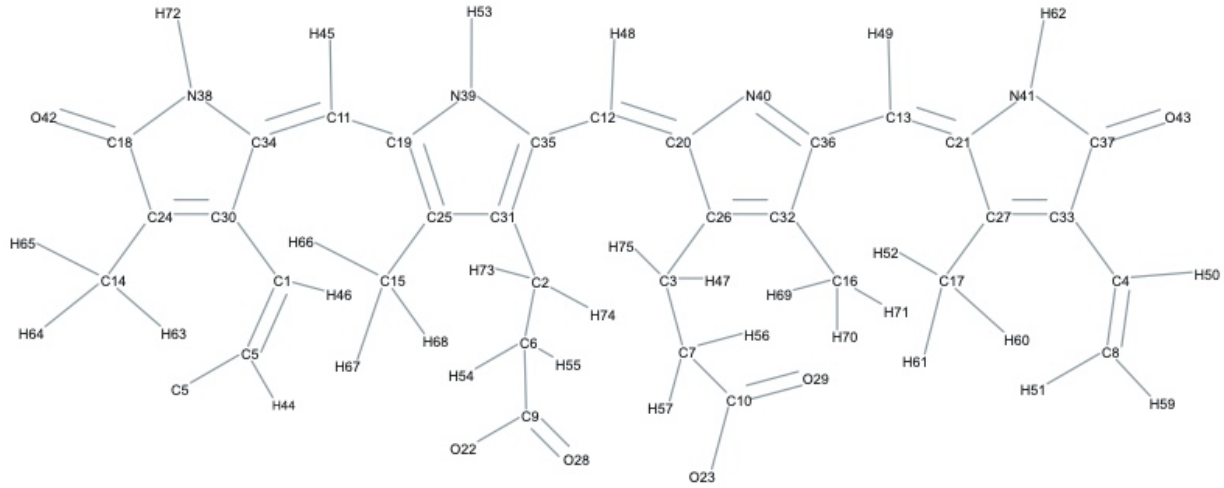
```
    CBX     CFX    CAX    5      123.32 548.1
    CBX     CFX    HHX    5      119.94 415.05
    CBX     CBX    CBX    5      121.81 579.9
    CBX     N2X    CBX    5      110.37 563.17
    CBX     N2X    HNX    5      119.28 395.81
    CCX     CHX    CBX    5      123.32 548.1
    CCX     CHX    HHX    5      114.95 419.24
    CCX     CCX    CCX    5      114.19 570.7
    CCX     N3X    CCX    5      105.49 600.82
    CBX     CHX    CCX    5      123.32 548.1
    CBX     CBX    HCX    5      110.36 393.3
    CBX     N1X    CBX    5      122.15 539.74
    CBX     N1X    HNX    5      117.9  398.32
    OCX     CGX    OHX    5      122.1  635.13
    OHX     CGX    OCX    5      122.1  635.13
    CAX     CMX    HCX    5      110.49 394.97
    CAX     CAX    N1X    5      112.7  578.23
    CAX     CAX    OCX    5      123.93 578.23
    CBX     CMX    HCX    5      110.36 393.3
    CBX     CBX    N2X    5      121.94 578.23
    CCX     CMX    HCX    5      110.49 394.97
    CCX     CCX    N3X    5      121.98 565.68
    CBX     CBX    N1X    5      114.8  559.82
    CAX     CHX    HHX    5      115.44 397.48
    CBX     CBX    OCX    5      123.2  564
    CAX     CFX    HHX    5      114.95 419.24
    N1X     CAX    OCX    5      123.05 620.91
    N1X     CBX    OCX    5      123.05 620.91
    HCX     CMX    HCX    5      107.58 329.7
    HNX     CMX    HNX    5      107.58 329.7

[ proper dihedraltypes ]
;    i       j      k      l      func   phi0   cp           mult
    CT1     CT2    SM     CBX    9      0      1.39467      3
    CHX     CBX    SM     CT2    9      180    4.6024       2
    CHX     CAX    CAX    CMX    9      180    16.736       2
    CHX     CAX    CAX    CAX    9      180    16.736       2
    CHX     CAX    CAX    CFX    9      180    16.736       2
    CHX     CAX    CAX    N1X    9      180    16.736       2
    CMX     CMX    CGX    OCX    9      180    0            2
    CMX     CMX    CGX    OHX    9      180    0            2
    CMX     CBX    CBX    CMX    9      180    7.9496       1
    CMX     CBX    CBX    CMX    9      180    27.8236      2
    CMX     CBX    CBX    CBX    9      180    27.8236      2
    CMX     CBX    CBX    CHX    9      180    27.8236      2
    CMX     CBX    CBX    N2X    9      180    27.8236      2
    CMX     CCX    CCX    CHX    9      180    16.736       2
    CMX     CCX    CCX    N3X    9      180    16.736       2
    CMX     CCX    CCX    CMX    9      180    16.736       2
    CMX     CCX    CCX    CCX    9      180    16.736       2
    CMX     CBX    CBX    HCX    9      0      0.66944      3
    CMX     CBX    CBX    N1X    9      0      0.29288      2
    CMX     CBX    CBX    N1X    9      0      0.4184       4
    CMX     CBX    CBX    OCX    9      180    0            2
    CT2     SM     CBX    HHX    9      180    4.6024       2
    CBX     CHX    CAX    CAX    9      180    4.184        2
    CBX     SM     CT2    HA     9      0      1.39467      3
    CMX     CMX    CBX    CBX    9      0      0            0
    CMX     CGX    OHX    HOX    9      180    9.6232       2
    CMX     CMX    CCX    CCX    9      0      0            0
    CMX     CMX    CBX    HCX    9      0      0.66944      3
    CGX     CMX    CMX    CBX    9      0      0.65084      3
    CGX     CMX    CMX    HCX    9      0      0.65084      3
    CGX     CMX    CMX    CCX    9      0      0.65084      3
    SMCBX   CHX    CAX    9      180    27.8236             2
```

```
SMCBX    CHX    HHX    9      180    27.8236             2
CFX      CBX    CBX    CMX    9      180    27.8236      2
CFX      CBX    CBX    CBX    9      180    27.8236      2
CFX      CBX    N2X    CBX    9      180    2.615        2
CFX      CBX    N2X    HNX    9      180    2.615        2
CFX      CAX    CAX    CAX    9      180    16.736       2
CFX      CAX    N1X    CAX    9      180    6.9036       2
CFX      CAX    N1X    HNX    9      180    6.9036       2
CHX      CCX    CCX    CCX    9      180    16.736       2
CHX      CCX    N3X    CCX    9      180    19.874       2
CHX      CBX    CBX    CBX    9      180    27.8236      2
CHX      CBX    N2X    CBX    9      180    2.615        2
CHX      CBX    N2X    HNX    9      180    2.615        2
CHX      CBX    CBX    HCX    9      0      0            0
CHX      CBX    N1X    CBX    9      180    2.7196       2
CHX      CBX    N1X    HNX    9      180    2.7196       2
CHX      CCX    CCX    CMX    9      180    16.736       2
CMX      CAX    CAX    N1X    9      180    12.029       2
CMX      CAX    CAX    OCX    9      180    12.029       2
CMX      CAX    CAX    CAX    9      180    16.736       2
CAX      CAX    CMX    HCX    9      0      0            0
CAX      CAX    CAX    CAX    9      180    16.736       2
CAX      N1X    CAX    CAX    9      180    6.9036       2
CBX      CFX    CAX    CAX    9      180    27.8236      2
CBX      CFX    CAX    N1X    9      180    27.8236      2
CBX      CBX    CMX    HCX    9      0      0            0
CBX      CBX    CMX    HCX    9      0      4.8116       1
CBX      CBX    CMX    HCX    9      180    1.58992      3
CBX      CBX    CBX    CBX    9      180    27.8236      2
CBX      N2X    CBX    CBX    9      180    2.615        2
CCX      CHX    CBX    CBX    9      180    27.8236      2
CCX      CHX    CBX    N2X    9      180    27.8236      2
CCX      CCX    CMX    HCX    9      0      0            0
CCX      CCX    CCX    CCX    9      180    16.736       2
CCX      N3X    CCX    CCX    9      180    19.874       2
CBX      CHX    CCX    CCX    9      180    4.184        2
CBX      CHX    CCX    N3X    9      180    4.184        2
CBX      CBX    CBX    HCX    9      0      0.65084      3
CBX      N1X    CBX    CBX    9      180    10.46        2
CBX      N1X    CBX    OCX    9      180    10.46        2
OCX      CGX    CMX    HNX    9      0      0            0
OCX      CGX    CMX    HNX    9      0      3.3472       1
OCX      CGX    CMX    HNX    9      180    0.33472      3
OCX      CGX    OHX    HOX    9      0      7.9496       1
OCX      CGX    OHX    HOX    9      180    9.6232       2
OHX      CGX    CMX    HCX    9      180    0            2
CAX      CAX    N1X    HNX    9      180    10.46        2
CAX      CAX    CHX    HHX    9      180    4.184        2
CAX      CAX    CAX    N1X    9      180    16.736       2
CBX      CBX    CFX    CAX    9      180    27.8236      2
CBX      CBX    CFX    HHX    9      180    27.8236      2
CBX      CBX    N2X    CBX    9      180    2.615        2
CBX      CBX    N2X    HNX    9      180    2.615        2
CBX      CBX    CBX    N2X    9      180    27.8236      2
CCX      CMX    CMX    HCX    9      0      0.65084      3
CCX      CCX    CHX    HHX    9      180    27.8236      2
CCX      CCX    N3X    CCX    9      180    19.874       2
CCX      CCX    CCX    N3X    9      180    16.736       2
CBX      CBX    CHX    CCX    9      180    27.8236      2
CBX      CBX    CHX    HHX    9      180    27.8236      2
CBX      CBX    N1X    HNX    9      180    2.7196       2
CBX      CBX    CBX    N1X    9      0      0.29288      2
CBX      CBX    CBX    N1X    9      0      0.4184       4
CBX      CBX    CBX    OCX    9      180    0            2
OHX      CGX    CMX    HNX    9      180    0            2
OCX      CGX    CMX    HCX    9      0      0            0
```

```
     OCX      CGX    CMX    HCX    9      0      3.3472       1
     OCX      CGX    CMX    HCX    9      180    0.33472      3
     CAX      CHX    CBX    HHX    9      180    27.8236      2
     CAX      CAX    CAX    OCX    9      180    12.029       2
     CAX      CAX    CFX    HHX    9      180    27.8236      2
     CBX      CMX    CMX    HNX    9      0      0.65084      3
     CBX      CMX    CMX    HCX    9      0      0.66944      3
     CAX      CFX    CBX    N2X    9      180    27.8236      2
     CAX      N1X    CAX    OCX    9      180    10.46        2
     CCX      CHX    CBX    N1X    9      180    27.8236      2
     N1X      CAX    CFX    HHX    9      180    27.8236      2
     N2X      CBX    CFX    HHX    9      180    27.8236      2
     N2X      CBX    CHX    HHX    9      180    27.8236      2
     N3X      CCX    CHX    HHX    9      180    27.8236      2
     N3X      CCX    CHX    HHX    9      180    4.184        2
     N1X      CBX    CHX    HHX    9      180    27.8236      2
     N1X      CBX    CBX    HCX    9      0      0            0
     N1X      CBX    CBX    HCX    9      180    0            2
     OCX      CAX    N1X    HNX    9      0      8.368        1
     OCX      CAX    N1X    HNX    9      180    10.46        2
     OCX      CBX    CBX    HCX    9      0      0            0
     OCX      CBX    CBX    HCX    9      0      3.3472       1
     OCX      CBX    CBX    HCX    9      180    0.33472      3
     OCX      CBX    N1X    HNX    9      0      8.368        1
     OCX      CBX    N1X    HNX    9      180    10.46        2
     HCX      CBX    CMX    HCX    9      0      0.6276       3
     HCX      CBX    CBX    HCX    9      0      0.6276       3
     HHX      CBX    CHX    HHX    9      180    27.8236      2
     HCX      CMX    CMX    HCX    9      0      0.6276       3
     HNX      CMX    CMX    HCX    9      0      0.6276       3

[ improper dihedraltypes ]
;    i        j      k      l      func   q0     cq
     CHX      HHX    CBX    SM     2      0.00   4.6024
     CMX      CCX    CCX    CCX    2      0.00   4.6024
     CBX      CAX    CHX    HHX    2      0.00   4.6024
     CMX      OCX    CGX    OHX    2      0.00   4.6024
     CAX      CMX    CAX    CAX    2      0.00   4.6024
     CAX      CAX    N1X    HNX    2      0.00   4.6024
     CBX      CBX    CBX    CMX    2      0.00   4.6024
     CBX      CAX    CFX    HHX    2      0.00   4.6024
     CBX      CBX    N2X    HNX    2      0.00   4.6024
     CBX      CCX    CHX    HHX    2      0.00   4.6024
     CAX      N1X    CAX    OCX    2      0.00   43.932
     CBX      CFX    CBX    N2X    2      0.00   4.6024
     CCX      CHX    CCX    N3X    2      0.00   4.6024
     CBX      CHX    CBX    N1X    2      0.00   4.6024
     CAX      CFX    CAX    N1X    2      0.00   4.6024
     CBX      CHX    CBX    N2X    2      0.00   4.6024
     CBX      N1X    CBX    OCX    2      0.00   43.932
     CAX      CAX    CAX    CHX    2      0.00   4.6024
     CBX      CBX    N1X    HNX    2      0.00   4.6024
     CHX      CAX    CAX    CAX    2      0.00   4.60240
     CMX      CBX    CBX    CBX    2      0.00   4.60240
     CMX      CCX    CCX    CCX    2      0.00   4.60240
     CBX      CAX    CHX    HHX    2      0.00   4.60240
     CMX      OCX    CGX    OHX    2      0.00   4.60240
     CMX      OCX    CGX    OHX    2      0.00   4.60240
     CMX      CBX    CBX    CBX    2      0.00   4.60240
     CMX      CCX    CCX    CCX    2      0.00   4.60240
     CAX      CMX    CAX    CAX    2      0.00   4.60240
     CAX      CAX    N1X    HNX    2      0.00   4.60240
     CBX      CAX    CFX    HHX    2      0.00   4.60240
     CBX      CBX    N2X    HNX    2      0.00   4.60240
     CBX      CCX    CHX    HHX    2      0.00   4.60240
```

```
CAX      N1X    CAX    OCX    2      0.00   43.9320
CBX      CFX    CBX    N2X    2      0.00   4.60240
CCX      CHX    CCX    N3X    2      0.00   4.60240
CBX      CHX    CBX    N1X    2      0.00   4.60240
CAX      CFX    CAX    N1X    2      0.00   4.60240
CBX      CHX    CBX    N2X    2      0.00   4.60240
CCX      CHX    CCX    N3X    2      0.00   4.60240
CBX      N1X    CBX    OCX    2      0.00   43.9320
CBX      CCX    CHX    HHX    2      0.00   4.60240
CBX      CBX    N1X    HNX    2      0.00   4.60240
```

Parameters 2: Free-BV parameters (produced via CHARMM-GUI)



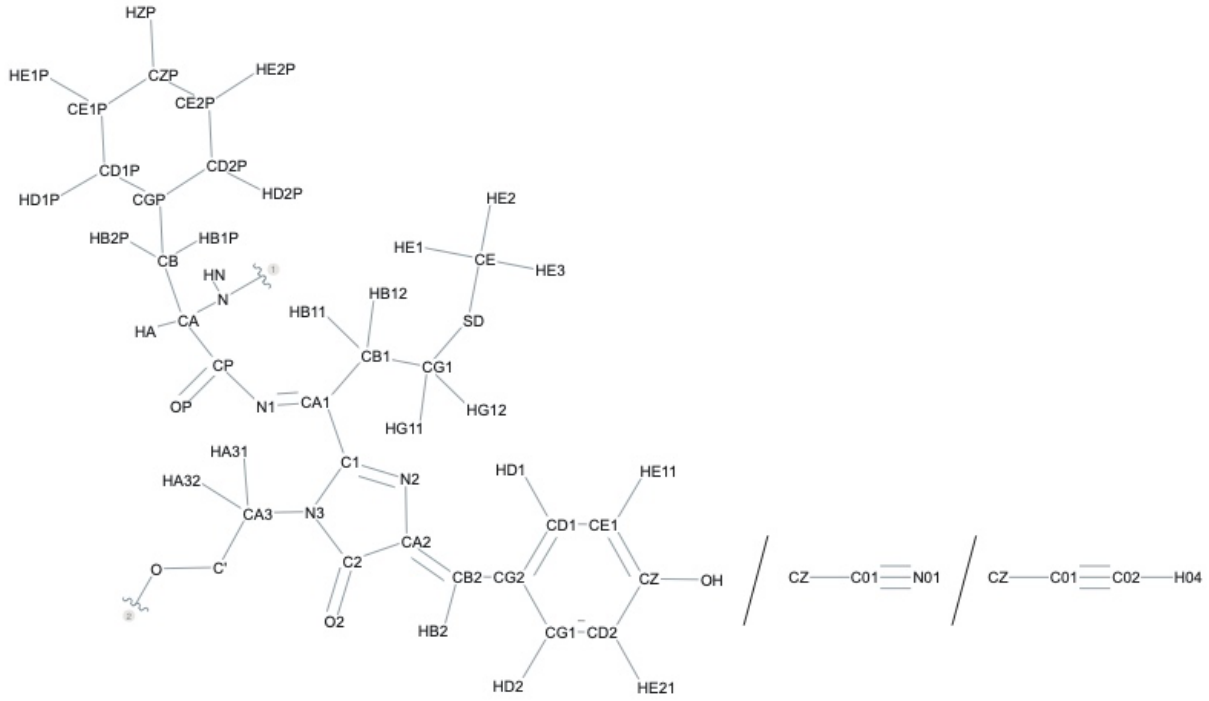

```
  [ atoms ]
;   id        type      charge numb
    C1        CG2DC1   -0.155   0
    C2        CG321    -0.271   1
    C3        CG321    -0.348   2
    C4        CG2DC1   -0.155   3
    C5        CG2DC3   -0.520   4
    C6        CG321    -0.275   5
    C7        CG321    -0.275   6
    C8        CG2DC3   -0.520   7
    C9        CG2O3     0.618   8
    C10       CG2O3     0.618   9
    C11       CG2DC1    0.092  10
    C12       CG2DC1    0.103  11
    C13       CG2DC1    0.262  12
    C14       CG331    -0.156  13
    C15       CG331    -0.203  14
    C16       CG331    -0.154  15
    C17       CG331    -0.206  16
    C18       CG2R53    0.673  17
    C19       CG2R51   -0.243  18
    C20       CG2510    0.088  19
    C21       CG2510    0.150  20
    O22       OG2D2    -0.760  21
    O23       OG2D2    -0.760  22
    C24       CG2R51   -0.242  23
    C25       CG2R51   -0.411  24
    C26       CG2R51    0.429  25
    C27       CG2R51   -0.156  26
    O28       OG2D2    -0.760  27
    O29       OG2D2    -0.760  28
    C30       CG2R51    0.097  29
    C31       CG2R51    0.359  30
    C32       CG2R51   -0.496  31
    C33       CG2R51    0.137  32
    C34       CG2510    0.109  33
    C35       CG2R51   -0.307  34
    C36       CG2R52   -0.183  35
     C37      CG2R53    0.556  36
    N38       NG2R51   -0.535  37
    N39       NG2R51    0.079  38
    N40       NG2R50   -0.211  39
```

```
    N41      NG2R51   -0.535  40
    O42      OG2D1    -0.570  41
    O43      OG2D1    -0.570  42
    H44      HGA5      0.210  43
    H45      HGA4      0.149  44
    H46      HGA4      0.204  45
    H47      HGA2      0.090  46
    H48      HGA4      0.149  47
    H49     HGA4      0.149  48
    H50      HGA4      0.204  49
    H51     HGA5      0.210  50
    H52     HGA3      0.090  51
    H53     HGP1      0.246  52
    H54     HGA2      0.090  53
    H55     HGA2      0.090  54
    H56      HGA2      0.090  55
    H57     HGA2      0.090  56
    H58     HGA5      0.210  57
    H59     HGA5      0.210  58
    H60     HGA3      0.090  59
    H61      HGA3      0.090  60
    H62     HGP1      0.313  61
    H63     HGA3      0.090  62
    H64     HGA3      0.090  63
    H65     HGA3      0.090  64
    H66     HGA3      0.090  65
    H67     HGA3      0.090  66
    H68     HGA3      0.090  67
    H69     HGA3      0.090  68
    H70     HGA3      0.090  69
    H71     HGA3      0.090  70
    H72     HGP1      0.313  71
    H73     HGA2      0.090  72
    H74     HGA2      0.090  73
    H75     HGA2      0.090  74


[ bondtypes ]
;      i        j  func           b0           kb
  CG2DC1   CG2R52     1   0.14253999    288896.84
  CG251O   CG2R51     1   0.14800000    213384.00
  CG251O   NG2R51     1   0.14100000    167360.00


[ angletypes ]
;      i        j        k  func  th0          cth
  CG251O   CG2DC1   CG2R52     5   122.000000   242.671997
  CG2R52   CG2DC1     HGA4     5   120.000000   267.776001
  CG2R51   CG2R53   NG2R51     5   107.000000   418.399994
  CG2DC1   CG2R51   NG2R51     5   153.479996   292.880005
  CG2DC1   CG251O   CG2R51     5   119.000000   334.720001
  CG2R51   CG251O   NG2R50     5   110.349998   987.591370
  CG2DC1   CG251O   NG2R51     5   130.000000   334.720001
  CG2R51   CG251O   NG2R51     5   111.000000   970.687988
   CG331   CG2R51   CG2R53     5   130.000000   167.360001
   CG321   CG2R51   CG251O     5   130.000000   383.254395
  CG251O   CG2R51   CG2R51     5   116.500000   836.799988
   CG331   CG2R51   CG251O     5   130.000000   383.254395
  CG2DC1   CG2R51   CG251O     5   154.350006   291.875854
   CG331   CG2R51   CG2R52     5   130.000000   167.360001
  CG2DC1   CG2R51   CG2R53     5   154.350006   291.875854
  CG2DC1   CG2R52   CG2R51     5   154.350006   291.875854
  CG2DC1   CG2R52   NG2R50     5   153.479996   292.880005
  CG2R53   NG2R51   CG251O     5   107.500000  1087.839966
  CG251O   NG2R51     HGP1     5   127.000000   251.039993
  CG251O   NG2R50   CG2R52     5   101.000000   836.799988
```

```
[ proper dihedraltypes ]
;    i        j        k        l func          phi0        cp          mult
  CG2DC3   CG2DC1   CG2R51   CG251O     9   180.000000     0.209200     1
  CG2DC3   CG2DC1   CG2R51   CG251O     9   180.000000     3.221680     2
  CG2DC3   CG2DC1   CG2R51   CG251O     9     0.000000     1.799120     4
    HGA4   CG2DC1   CG2R51   CG251O     9   180.000000     3.472720     2
  CG2R51    CG321    CG321    CG2O3     9     0.000000     0.836800     3
   CG321    CG321   CG2R51   CG251O     9     0.000000     0.836800     1
   CG321    CG321   CG2R51   CG251O     9     0.000000     1.129680     2
    HGA2    CG321   CG2R51   CG251O     9     0.000000     0.000000     3
  CG2DC3   CG2DC1   CG2R51   CG2R53     9   180.000000     0.209200     1
  CG2DC3   CG2DC1   CG2R51   CG2R53     9   180.000000     3.221680     2
  CG2DC3   CG2DC1   CG2R51   CG2R53     9     0.000000     1.799120     4
    HGA4   CG2DC1   CG2R51   CG2R53     9   180.000000     3.472720     2
  CG251O   CG2DC1   CG2R51   NG2R51     9   180.000000     8.326160     1
  CG251O   CG2DC1   CG2R51   NG2R51     9   180.000000     5.564720     2
  CG251O   CG2DC1   CG2R51   NG2R51     9     0.000000     0.418400     4
    HGA4   CG2DC1   CG2R51   NG2R51     9   180.000000     0.041840     2
  CG2R51   CG2DC1   CG251O   CG2R51     9   180.000000    16.819679     2
  CG2R51   CG2DC1   CG251O   NG2R51     9   180.000000    18.869841     2
    HGA4   CG2DC1   CG251O   CG2R51     9   180.000000    16.317600     2
    HGA4   CG2DC1   CG251O   NG2R51     9   180.000000    22.551760     2
  CG2R52   CG2DC1   CG251O   CG2R51     9   180.000000    16.819679     2
  CG2R52   CG2DC1   CG251O   NG2R51     9   180.000000    18.869841     2
  CG251O   CG2DC1   CG2R52   CG2R51     9   180.000000     0.209200     1
  CG251O   CG2DC1   CG2R52   CG2R51     9   180.000000     3.221680     2
  CG251O   CG2DC1   CG2R52   CG2R51     9     0.000000     1.799120     4
  CG251O   CG2DC1   CG2R52   NG2R50     9   180.000000     8.326160     1
  CG251O   CG2DC1   CG2R52   NG2R50     9   180.000000     5.564720     2
  CG251O   CG2DC1   CG2R52   NG2R50     9     0.000000     0.418400     4
    HGA4   CG2DC1   CG2R52   CG2R51     9   180.000000     3.472720     2
    HGA4   CG2DC1   CG2R52   NG2R50     9   180.000000     0.041840     2
    HGA3    CG331   CG2R51   CG2R53     9     0.000000     0.000000     3
    HGA3    CG331   CG2R51   CG2R52     9     0.000000     0.000000     3
    HGA3    CG331   CG2R51   CG251O     9     0.000000     0.000000     3
  NG2R51   CG2R53   CG2R51    CG331     9   180.000000    12.552000     2
  NG2R51   CG2R53   CG2R51   CG2R51     9   180.000000    16.736000     2
   OG2D1   CG2R53   CG2R51    CG331     9   180.000000     0.000000     2
  CG2R51   CG2R53   NG2R51   CG251O     9   180.000000     8.368000     2
  CG2R51   CG2R53   NG2R51     HGP1     9   180.000000     1.255200     2
   OG2D1   CG2R53   NG2R51   CG251O     9   180.000000    10.460000     2
  CG2DC1   CG2R51   CG2R51    CG331     9   180.000000     0.000000     2
  CG2DC1   CG2R51   NG2R51   CG2R51     9   180.000000     0.083680     2
  CG2DC1   CG2R51   NG2R51     HGP1     9   180.000000     0.167360     2
  CG2DC1   CG251O   CG2R51    CG321     9   180.000000     3.347200     2
  CG2DC1   CG251O   CG2R51   CG2R51     9   180.000000     0.836800     2
  NG2R50   CG251O   CG2R51    CG321     9   180.000000    12.552000     2
  NG2R50   CG251O   CG2R51   CG2R51     9   180.000000    35.563999     2
  CG2DC1   CG251O   NG2R50   CG2R52     9   180.000000     7.447520     2
  CG2R51   CG251O   NG2R50   CG2R52     9   180.000000     0.334720     2
  CG2DC1   CG251O   CG2R51    CG331     9   180.000000     3.347200     2
  NG2R51   CG251O   CG2R51    CG331     9   180.000000    12.552000     2
  NG2R51   CG251O   CG2R51   CG2R51     9   180.000000    16.736000     2
  CG2DC1   CG251O   NG2R51   CG2R53     9   180.000000     4.184000     2
  CG2DC1   CG251O   NG2R51     HGP1     9   180.000000     1.673600     2
  CG2R51   CG251O   NG2R51   CG2R53     9   180.000000     2.092000     2
  CG2R51   CG251O   NG2R51     HGP1     9   180.000000     1.673600     2
   CG331   CG2R51   CG2R51   CG251O     9   180.000000    17.991199     2
  CG2R53   CG2R51   CG2R51   CG2DC1     9   180.000000     0.000000     2
  CG2R53   CG2R51   CG2R51   CG251O     9   180.000000    62.759998     2
   CG321   CG2R51   CG2R51   CG2R52     9   180.000000    17.991199     2
  CG251O   CG2R51   CG2R51   CG2R52     9   180.000000    62.759998     2
   CG331   CG2R51   CG2R51   CG2R53     9   180.000000    17.991199     2
  CG251O   CG2R51   CG2R51   CG2DC1     9   180.000000     0.000000     2
  CG2DC1   CG2R51   CG251O   CG2DC1     9   180.000000     3.347200     2
```

```
  CG2DC1   CG2R51   CG251O   NG2R51     9   180.000000     0.000000     2
   CG331   CG2R51   CG2R52   CG2DC1     9   180.000000     0.000000     2
   CG331   CG2R51   CG2R52   NG2R50     9   180.000000    12.552000     2
  CG2R51   CG2R51   CG2R52   CG2DC1     9   180.000000     0.000000     2
  CG2DC1   CG2R51   CG2R53   NG2R51     9   180.000000     0.000000     2
  CG2DC1   CG2R51   CG2R53    OG2D1     9   180.000000     5.020800     2
  CG2DC1   CG2R52   NG2R50   CG251O     9   180.000000     3.807440     2
  CG2R51   CG2R52   NG2R50   CG251O     9   180.000000    58.576000     2

[ improper dihedraltypes ]
;    i        j        k          l       func.   q0     cq
  CG2R53   CG2R51   NG2R51    OG2D1     2     0.000000   753.119995
```

Parameters 3: mCherry chromophore parameters (WT, CRO-CC, CRO-CN)




```
[ atoms ] ;   WT CRO [CCA]
;   id type charge   numb
     N       NH1    -0.47  0
     HN       H      0.31   1
     CA       CT1    0.07   2
     HA       HB     0.09   3
     CB       CT2   -0.18   4
     HB1P     HA     0.09   5
     HB2P     HA     0.09   6
     CGP      CA     0.00   7
     CD1P     CA    -0.115  8
     HD1P     HP     0.115  9
     CE1P     CA    -0.115  10
     HE1P     HP     0.115  11
     CZP      CA    -0.115  12
     HZP      HP     0.115  13
     CD2P     CA    -0.115  14
     HD2P     HP     0.115  15
     CE2P     CA    -0.115  16
     HE2P     HP     0.115  17
     CP       C      0.51   18
     OP       O     -0.51   19
     N1       NQ     -0.40  20
     CA1      CA1Q   0.40   21
     C1       C1Q    0.5    22
     N2       N2Q    -0.6   23
     N3       N3Q    -0.57  24
     C2       C2Q    0.57   25
     O2       O2Q    -0.57  26
     Ca2      CA2Q   0.1    27
     Cb2      CB2Q   -0.14  28
     Hb2      HB2Q   0.21   29
     Cg2      CG2Q   -0.09  30
     Cd1      CDQ    -0.08  31
     Hd1      HDQ    0.14   32
     Cd2      CDQ    -0.08  33
     Hd2      HDQ    0.14   34
     Ce1      CEQ    -0.28  35
```

```
     He11     HEQ    0.1    36
     Ce2      CEQ    -0.28  37
     He21     HEQ    0.1    38
     Cz       CZQ    0.45   39
     Oh       OHQ    -0.62  40
     CB1      CT2    -0.18  41
     HB11     HA     0.09   42
     HB12     HA     0.09   43
     CG1      CT2    -0.14  44
     HG11     HA     0.09   45
     HG12     HA     0.09   46
     SD       S      -0.09  47
     CE       CT3    -0.22  48
     HE1      HA     0.09   49
     HE2      HA     0.09   50
     HE3      HA     0.09   51
     CA3      CT2    -0.18  52
     HA31     HB     0.09   53
     HA32     HB     0.09   54
     C        C      0.51   55
     O        O      -0.51  56

[ atoms ] ;   CRO-CN [CCN]
;   id type charge  numb
     N      NH1    0.007    0
     HN     H      0.063    1
     CA     CT1    -0.06    2
     HA     HB     0.063    3
     CB     CT2    -0.06    4
     HB1P   HA     0.062    5
     HB2P   HA     0.062    6
     CGP    CA     -0.06    7
     CD1P   CA     -0.06    8
     HD1P   HP     0.062    9
     CE1P   CA     -0.06    10
     HE1P   HP     0.062    11
     CZP    CA     -0.06    12
     HZP    HP     0.062    13
     CD2P   CA     -0.06    14
     HD2P   HP     0.062    15
     CE2P   CA     -0.06    16
     HE2P   HP     0.062    17
     CP     C      -0.06    18
     OP     O       0.02    19
     N1     NQ     0.007    20
     CA1    CA1Q   -0.06    21
     C1     C1Q    -0.06    22
     N2     N2Q    0.007    23
     N3     N3Q    0.007    24
     C2     C2Q    -0.06    25
     O2     O2Q     0.02    26
     Ca2    CA2Q   -0.06    27
     Cb2    CB2Q   -0.06    28
     Hb2    HB2Q   0.062    29
     Cg2    CG2Q   -0.06    30
     Cd1    CDQ    -0.06    31
     Hd1    HDQ    0.062    32
     Cd2    CDQ    -0.06    33
     Hd2    HDQ    0.062    34
     Ce1    CEQ    -0.06    35
     He11   HEQ    0.062    36
     Ce2    CEQ    -0.06    37
     He21   HEQ    0.062    38
     Cz     CZQ    -0.06    39
     C01    CN     -0.06    40
     N01    NC     0.007    41
```

```
     CB1   CT2    -0.06    42
     HB11  HA     0.062    43
     HB12  HA     0.062    44
     CG1   CT2    -0.06    45
     HG11  HA     0.062    46
     HG12  HA     0.062    47
     SD    S      0.037    48
     CE    CT3    -0.06    49
     HE1   HA     0.062    50
     HE2   HA     0.062    51
     HE3   HA     0.062    52
     CA3   CT2    -0.06    53
     HA31  HB     0.062    54
     HA32  HB     0.062    55
     C     C      -0.06    56
     O     O      0.02     57

[ atoms ] ;   CRO-CN [CCC]
;   id  type charge    numb
     N        NH1    -0.16114     0
     HN       H      0.06226      1
     CA       CT1    -0.06004     2
     HA       HB     0.06226      3
     CB       CT2    -0.06004     4
     HB1P     HA     0.06226      5
     HB2P     HA     0.06226      6
     CGP      CA     -0.06004     7
     CD1P     CA     -0.06004     8
     HD1P     HP     0.06226      9
     CE1P     CA     -0.06004     10
     HE1P     HP     0.06226      11
     CZP      CA     -0.06004     12
     HZP      HP     0.06226      13
     CD2P     CA     -0.06004     14
     HD2P     HP     0.06226      15
     CE2P     CA     -0.06004     16
     HE2P     HP     0.06226      17
     CP       C      -0.06004     18
     OP       O      0.02036      19
     N1       NQ     0.00726      20
     CA1      CA1Q   -0.06004     21
     C1       C1Q    0.13736      22
     N2       N2Q    0.00726      23
     N3       N3Q    0.00726      24
     C2       C2Q    -0.08934     25
     O2       O2Q    0.02036      26
     CA2      CA2Q   -0.06004     27
     CB2      CB2Q   -0.06004     28
     HB2      HB2Q   0.06226      29
     CG2      CG2Q   -0.06004     30
     CD1      CDQ    -0.06004     31
     HD1      HDQ    0.06226      32
     CD2      CDQ    -0.06004     33
     HD2      HDQ    0.06226      34
     CE1      CEQ    -0.06004     35
     HE11     HEQ    0.06229      36
     CE2      CEQ    -0.06004     37
     HE21     HEQ    0.06229      38
     CZ       CZQ    -0.06004     39
     C01      CCQ    -0.06004     40
     C02      CCQ    -0.06004     41
     H04      HEQ    0.06226      42
     CB1      CT2    -0.06004     43
     HB11     HA     0.06226      44
     HB12     HA     0.06226      45
     CG1      CT2    -0.06004     46
```

```
     HG11     HA     0.06226      47
     HG12     HA     0.06226      48
     SD       S      0.03696      49
     CE       CT3    -0.06004     50
     HE1      HA     0.06226      51
     HE2      HA     0.06226      52
     HE3      HA     0.06226      53
     CA3      CT2    -0.06004     54
     HA31     HB     0.06226      55
     HA32     HB     0.06226      56
     C        C      -0.06004     57
     O        O      0.02036      58

[ bondtypes ]
;    i        j      func    b0     kb
     C1Q      C      1      0.149  296227.2
     N3Q      CT2    1      0.144  331372.8
     N3Q      C1Q    1      0.139  334720.0
     N3Q      C2Q    1      0.141  334720.0
     C2Q      O2Q    1      0.124  714627.2
     C1Q      N2Q    1      0.13   334720.0
     C2Q      CA2Q   1      0.146  343088.0
     N2Q      CA2Q   1      0.14   334720.0
     CA2Q     CB2Q   1      0.139  418400.0
     CB2Q     HB2Q   1      0.11   301666.4
     CB2Q     CG2Q   1      0.141  365681.6
     CG2Q     CDQ    1      0.143  255224.0
     CG2Q     CDQ    1      0.143  255224.0
     CDQ      HDQ    1      0.108  284512.0
     CDQ      HDQ    1      0.108  284512.0
     CDQ      CEQ    1      0.135  255224.0
     CDQ      CEQ    1      0.135  255224.0
     CEQ      HEQ    1      0.108  284512.0
     CEQ      HEQ    1      0.108  284512.0
     CEQ      CZQ    1      0.1455 255224.0
     CEQ      CZQ    1      0.1455 255224.0
     CZQ      OHQ    1      0.125  704585.6
     C1Q      CA1Q   1      0.149  296227.2
     CT2      CA1Q   1      0.149  209200.0
     CA1Q     NQ     1      0.128  404174.4
     C        NQ     1      0.1141 382417.6
     CT1      CA1Q   1      0.149  209200.0
     CEQ      NOQ    1      0.1469 209370
     NOQ      ONQ    1      0.1231 302700
     CZQ      CN     1      0.1424 327905.1
     CZQ      CCQ    1      0.1431 265610
     CCQ      CCQ    1      0.1214 573180
     CCQ      HEQ    1      0.1066 328390

[ angletypes ]
;    i        j       k     func    th0    cth
     O        C       C1Q    5      119    669.44
     CT2      C       C1Q    5      120    669.44
     N2Q      C1Q     C      5      125    334.72
     N2Q      C1Q     N3Q    5      113.3  1087.8
     C1Q      N2Q     CA2Q   5      106.6  1087.8
     N3Q      C1Q     C      5      121.7  334.72
     C1Q      N3Q     C2Q    5      107.9  1087.8
     C1Q      N3Q     CT2    5      128.3  334.72
     N2Q      CA2Q    C2Q    5      108.3  1087.8
     N2Q      CA2Q    CB2Q   5      129.5  383.25
     C2Q      N3Q     CT2    5      123.8  334.72
     N3Q      C2Q     O2Q    5      124    418.44
     N3Q      C2Q     CA2Q   5      103    1087.8
```

```
     N3Q      CT2    HB     5      108    401.66
     N3Q      CT2    C      5      107    418.42
     O2Q      C2Q    CA2Q   5      133    368.19
     C2Q      CA2Q   CB2Q   5      122.7  383.25
     CA2Q     CB2Q   CG2Q   5      133.2  1087.8
     CA2Q     CB2Q   HB2Q   5      112    351.45
     CB2Q     CG2Q   CDQ    5      120    383.25
     CB2Q     CG2Q   CDQ    5      120    383.25
     HB2Q     CB2Q   CG2Q   5      115    351.45
     CG2Q     CDQ    CEQ    5      122    334.72
     CG2Q     CDQ    CEQ    5      122    334.72
     CDQ      CG2Q   CDQ    5      116    334.72
     CDQ      CEQ    CZQ    5      122    334.72
     CDQ      CEQ    CZQ    5      122    334.72
     CEQ      CZQ    CEQ    5      115    334.72
     HEQ      CEQ    CZQ    5      120    251.04
     HEQ      CEQ    CDQ    5      120    251.04
     HDQ      CDQ    CEQ    5      120    251.04
     HDQ      CDQ    CG2Q   5      120    251.04
     OHQ      CZQ    CEQ    5      120    378.23
     HEQ      CEQ    CZQ    5      120    251.04
     HEQ      CEQ    CDQ    5      120    251.04
     HDQ      CDQ    CEQ    5      120    251.04
     HDQ      CDQ    CG2Q   5      120    251.04
     OHQ      CZQ    CEQ    5      120    378.23
     CT2      CA1Q   C1Q    5      120    669.44
     N3Q      C1Q    CA1Q   5      121.7  334.72
     N2Q      C1Q    CA1Q   5      125    334.72
     O        CA1Q   C1Q    5      119    669.44
     NQ       CA1Q   CT2    5      121.0  585.76
     NQ       CA1Q   C1Q    5      118.4  585.76
     CA1Q     NQ     C      5      121.9  279.49
     NQ       C      O      5      122.5  669.44
     NQ       C      CT1    5      113.6  669.44
     CT2      CT2    CA1Q   5      108.0  435.13
     HA       CT2    CA1Q   5      109.5  276.14
     HA       CT1    CA1Q   5      109.5  276.14
     CT3      CT1    CA1Q   5      108.0  435.13
     OH1      CT1    CA1Q   5      110.1  633.45
     NQ       CA1Q   CT1    5      121.0  585.76
     CT1      CA1Q   C1Q    5      120    669.44
     ONQ      NOQ    ONQ    5      125.0  1096.4
     CEQ      NOQ    ONQ    5      117.7  829.10
     CDQ      CEQ    NOQ    5      119.0  553.29
     CZQ      CEQ    NOQ    5      117.8  553.71
     CCQ      CCQ    HEQ    5      179.6  342.17
     CZQ      CCQ    CCQ    5      179.5  454.22
     CEQ      CZQ    CCQ    5      120.2  525.93


[ proper dihedraltypes ]
;    i        j        k    l      func   phi0    cp       mult
     C1Q      N3Q    CT2    HB     9      0      0.133888      3
     C1Q      N3Q    CT2    C      9      0      0.133888      3
     C2Q      N3Q    CT2    HB     9      180    0.133888      3
     C2Q      N3Q    CT2    C      9      180    0.133888      3
     C1Q      C      CT2    HA     9      0.00   0.2092        3
     C1Q      C      CT2    CT2    9      0.00   0.2092        3
     N3Q      CT2    C      NH1    9      180.0  12.9704       1
     N3Q      CT2    C      O      9      0.00   2.092         1
     CT2      CT2    CA1Q   NQ     9      0.0    5.8576        1
     HA       CT2    CA1Q   NQ     9      180.0  0.0           3
     NH1      CT1    C      NQ     9      0.0    1.6736        1
     CT1      C      NQ     CA1Q   9      0.00   6.6944        1
     CT1      C      NQ     CA1Q   9      180.0  10.46         2
     NQ       C      CT1    HB     9      0.00   0.0           1
```

```
    NQ      C     CT1   CT2   9     0.00   0.0          1
    O       C     NQ    CA1Q  9     180.0  10.46        2
    C1Q     CA1Q  CT2   CT2   9     0.00   0.0          1
    C1Q     CA1Q  CT2   HA    9     0.00   0.0          1
    NQ      CA1Q  CT1   CT3   9     0.00   5.8576       1
    HA      CT1   CA1Q  NQ    9     180.0  0.0          3
    NQ      CA1Q  CT1   OH1   9     0.00   5.8576       1
    C1Q     CA1Q  CT1   OH1   9     0.00   5.8576       1
    C1Q     CA1Q  CT1   CT3   9     0.00   5.8576       1
    HA      CT1   CA1Q  C1Q   9     180.0  0.0          3
    X       CA2Q  CB2Q  X     9     180    16.3176      2
    X       CB2Q  CG2Q  X     9     180    11.2968      2
    X       C1Q   C     X     9     180    6.6944       2
    X       CDQ   CG2Q  X     9     180.0  12.9704      2
    X       CEQ   CDQ   X     9     180.0  12.9704      2
    X       CZQ   CEQ   X     9     180.0  12.9704      2
    X       N2Q   CA2Q  X     9     180.0  58.576       2
    X       N3Q   C2Q   X     9     180.0  58.576       2
    X       CA2Q  C2Q   X     9     180.0  58.576       2
    X       C1Q   N3Q   X     9     180.0  58.576       2
    X       C1Q   N2Q   X     9     180.0  58.576       2
    X       C1Q   CA1Q  X     9     180.0  6.6944       2
    X       CA1Q  NQ    X     9     180.0  24.647944    1
    X       CA1Q  NQ    X     9     180.0  1.1506       2
    X       CA1Q  NQ    X     9     180.0  -17.602088   3
    X       CA1Q  NQ    X     9     180.0  -8.20064     4
    NOQ     CEQ   CDQ   HDQ   9     180    15.167       2
    CDQ     CEQ   NOQ   ONQ   9     180    3.01248      2
    CZQ     CEQ   NOQ   ONQ   9     180    2.5104       2
    CG2Q    CDQ   CEQ   NOQ   9     180    15.167 2
    CEQ     CZQ   CEQ   NOQ   9     180    4.184        2
    OHQ     CZQ   CEQ   NOQ   9     180    4.184        2
    CZQ     CCQ   CCQ   HEQ   9     180    0.0          2
    CEQ     CZQ   CCQ   CCQ   9     180    0.0          2
    CDQ     CEQ   CZQ   CCQ   9     180    15.167       2
    CCQ     CZQ   CEQ   HEQ   9     180    15.167       2

[ improper dihedraltypes ]
;    i       j     k     l     func   q0     cq
    CA2Q    X     X     CB2Q  2      0.0    418.4
    CB2Q    X     X     HB2Q  2      0.0    251.04
    C1Q     X     X     N3Q   2      0.0    418.4
    C       X     X     C1Q   2      0.0    418.4
    N3Q     X     X     C2Q   2      0.0    418.4
    O2Q     X     X     C2Q   2      0.0    418.4
    N3Q     X     X     CT2   2      0.0    418.4
    N2Q     X     X     C1Q   2      0.0    418.4
    CEQ     ONQ   NOQ   ONQ   2      0.00   4.6024
    CZQ     CDQ   CEQ   NOQ   2      0.00   4.6024
    CEQ     CEQ   CZQ   CCQ   2      180    4.6024
```

9.4 Gaussian09 scripts

BV-bound

1. **HF/3-21g Geometry optimisation**

```
%chk=BV-bound.chk
%mem=100GB
# opt hf/3-21g scrf=(solvent=water) geom=connectivity int=acc2e=11

BV-bound

-1 1
 C              -1   -8.09465920  -38.92838005  -30.21685792
 S              -1   -8.96665920  -39.80638005  -28.88585792
 C              -1    6.49599015  -31.61876477  -24.16380899
 C              -1    6.39657306  -32.39156526  -25.53638527
 N              -1    6.11242604  -33.81019603  -25.25901152
 C              -1    4.90271664  -34.34763810  -25.09858963
 N              -1    3.74777590  -33.64137063  -25.28323292
 N              -1    4.79494051  -35.61773031  -24.71702564
 C              -1    1.54103126  -47.07325112  -12.61687761
 C              -1    0.88003126  -47.03125112  -13.98787761
 C              -1    1.19003126  -46.33125112  -15.16287761
 N              -1   -0.20596874  -47.82325112  -14.28787761
 C              -1   -0.56596874  -47.62525112  -15.57287761
 N              -1    0.29203126  -46.72525112  -16.11887761
 C               0    2.53955429  -42.26887866  -23.53568054
 C              -1   -2.68912793  -48.33012243  -19.12374276
 C               0   -7.20455745  -40.16006646  -27.02901029
 C               0    0.17703186  -38.02835812  -25.57260341
 N               0   -1.12890679  -42.73838994  -23.32056425
 C               0    2.95655429  -41.78487866  -22.12768054
 C               0   -4.00530488  -48.23084734  -19.10881993
 C               0   -8.45965920  -40.49238005  -27.39885792
 C               0    0.78203186  -37.60535812  -24.22260341
 N               0   -0.43472403  -45.54513985  -20.25482869
 O               0   -0.00897365  -46.74550361  -18.31828341
 C               0    4.36055429  -41.27287866  -22.12868054
 C               0    1.93203186  -36.67735812  -24.45960341
 N               0   -2.42661690  -40.70516665  -24.88433195
 C               0   -0.04828264  -40.87586190  -24.29646828
 C               0   -1.51947319  -44.78269008  -22.26201808
 C               0   -4.48438353  -39.77021312  -25.80788066
 C               0    1.62354689  -44.98973246  -22.23595925
 C               0   -3.65641354  -46.99940771  -21.56569457
 C               0   -8.87872797  -41.98984877  -24.94733839
 C               0   -2.74534516  -37.32268229  -26.50809099
 C               0   -0.03247807  -42.06957155  -23.66614496
 C               0   -1.44562584  -45.61595880  -21.17707116
 C               0   -6.47211203  -41.75277612  -23.80221323
 C               0   -3.10851057  -39.66120391  -25.47397859
 O               0    5.14455429  -41.31987866  -23.04268054
 O               0    1.72303186  -35.63335812  -24.98760341
 C               0    1.13855429  -42.72087866  -23.34468054
 C               0   -2.41975322  -46.63802525  -20.84359558
 C               0   -7.48572797  -41.48184877  -24.87433839
 C               0   -2.29534516  -38.58768229  -25.81809099
 O               0    4.72855429  -40.81687866  -21.05268054
```

O	0	3.01203186	−37.05135812	−24.15160341
C	0	0.74454689	−43.90173246	−22.75295925
C	0	−1.97724153	−47.23011639	−19.69654679
C	0	−6.79560402	−40.70852269	−25.75055307
C	0	−1.01296814	−38.89435812	−25.39460341
C	0	−0.62113037	−43.83599727	−22.76872598
C	0	−0.72704192	−46.52400563	−19.33764772
C	0	−5.46019657	−40.51051744	−25.24022019
C	0	−1.14360891	−40.16190034	−24.85047123
N	0	−5.25885906	−41.12469687	−24.05840395
O	0	−6.74711203	−42.46077612	−22.79521323
H	−1	5.58888780	−31.75299451	−23.61240379
H	−1	6.64833037	−30.57644194	−24.35157371
H	−1	7.36540499	−32.35008237	−25.98865053
H	−1	5.54691402	−31.99876601	−26.05473675
H	−1	6.89726597	−34.42537785	−25.18432804
H	−1	3.89126036	−36.02804441	−24.59453264
H	−1	5.61761895	−36.16198403	−24.55274520
H	−1	2.86248479	−34.08728838	−25.15125935
H	−1	3.78776600	−32.67853580	−25.55034690
H	−1	0.91813914	−46.57293210	−11.90512885
H	−1	1.67758451	−48.09148785	−12.31780043
H	−1	1.98081991	−45.62156417	−15.28896384
H	−1	0.27146027	−46.40265339	−17.06519021
H	−1	−1.38545403	−48.09575728	−16.07484156
H	−1	−7.03872353	−39.04421320	−30.08847358
H	−1	−8.38567947	−39.33507930	−31.16279816
H	0	3.13324081	−43.12451720	−23.78127981
H	0	2.52810014	−41.42061938	−24.18777333
H	0	−2.22207146	−49.01315894	−18.44534805
H	0	−6.77368129	−39.66626867	−27.87482912
H	0	−0.17200086	−37.13101266	−26.03934455
H	0	0.91669250	−38.63072037	−26.05733205
H	0	2.94991690	−42.64879543	−21.49640901
H	0	2.32973651	−40.94892705	−21.89706318
H	0	−4.59669013	−49.02016124	−18.69391387
H	0	−4.48082064	−47.36106044	−19.51162946
H	0	−8.85789624	−41.17037172	−26.67316202
H	0	0.03057596	−37.04139499	−23.71058815
H	0	1.18572404	−38.48923274	−23.77460980
H	0	0.34712996	−44.92177043	−20.24410939
H	0	−2.76409334	−41.59389070	−24.57404303
H	0	0.76181856	−40.18936604	−24.42822368
H	0	−2.46343644	−44.33096825	−22.48513414
H	0	−4.79627424	−39.17253084	−26.63878383
H	0	1.01912733	−45.77612370	−21.83450462
H	0	2.25866658	−44.60046541	−21.46784848
H	0	2.22356747	−45.37329557	−23.03455467
H	0	−4.14862533	−47.79875284	−21.05220170
H	0	−4.30553307	−46.14988069	−21.60867928
H	0	−3.41100635	−47.31140059	−22.55934227
H	0	−9.07859683	−42.60641420	−24.09598682
H	0	−9.55936061	−41.16429705	−24.95751509
H	0	−9.00225180	−42.56498357	−25.84112910
H	0	−1.90605912	−36.67330707	−26.64524335
H	0	−3.48198781	−36.83141555	−25.90732932
H	0	−3.16695191	−37.56668279	−27.46077746
H	0	−4.44315959	−41.14312830	−23.48022175
H	−1	−8.34617301	−37.88891346	−30.18294599

```
H              -1    2.49188325  -46.58484860  -12.66424289
H              -1    7.31815104  -32.00294602  -23.59692608

1 2 1.0 72 1.0 73 1.0 105 1.0
2 22 1.0
3 4 1.0 58 1.0 59 1.0 107 1.0
4 5 1.0 60 1.0 61 1.0
5 6 1.5 62 1.0
6 8 1.5 7 1.5
7 65 1.0 66 1.0
8 63 1.0 64 1.0
9 10 1.0 67 1.0 68 1.0 106 1.0
10 12 1.5 11 1.5
11 14 1.5 69 1.0
12 13 1.5
13 14 1.5 71 1.0
14 70 1.0
15 42 1.0 20 1.0 74 1.0 75 1.0
16 21 2.0 49 1.0 76 1.0
17 22 2.0 50 1.0 77 1.0
18 51 1.0 23 1.0 78 1.0 79 1.0
19 36 1.5 52 1.5
20 26 1.0 80 1.0 81 1.0
21 82 1.0 83 1.0
22 84 1.0
23 27 1.0 85 1.0 86 1.0
24 37 1.0 53 1.0 87 1.0
25 53 2.0
26 40 2.0 46 2.0
27 41 2.0 47 2.0
28 39 1.5 55 1.5 88 1.0
29 55 1.0 36 2.0 89 1.0
30 37 2.0 52 1.0 90 1.0
31 39 1.0 54 2.0 91 1.0
32 48 1.0 92 1.0 93 1.0 94 1.0
33 43 1.0 95 1.0 96 1.0 97 1.0
34 44 1.0 98 1.0 99 1.0 100 1.0
35 45 1.0 101 1.0 102 1.0 103 1.0
36 42 1.5
37 43 1.0
38 44 1.0 57 2.0 56 1.0
39 45 1.5
40
41
42 48 1.5
43 49 1.0
44 50 2.0
45 51 1.5
46
47
48 52 1.5
49 53 1.0
50 54 1.0
51 55 1.5
52
53
54 56 1.0
55
56 104 1.0
```

57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107

2. **B3LYP/6-31g Geometry optimisation**

```
%chk=BV-bound.chk
%mem=100GB
# opt b3lyp/6-31g scrf=(solvent=water) guess=read geom=checkpoint
int=acc2e=11

BV-bound

-1 1
```

3. **B3LYP/6-31g+(d,p) Geometry optimisation**

```
%chk=BV-bound.chk
%mem=100GB
# opt b3lyp/6-31+g(d,p) scrf=(solvent=water) guess=read geom=checkpoint
int=acc2e=11

BV-bound

-1 1
```

4. **B3LYP/6-31g+(d,p) Analysis – Frequency and TD-DFT (combined for conciseness)**

```
%chk=BV-bound.chk
%mem=100GB
# freq=raman td=(nstates=10) b3lyp/6-31+g(d,p) scrf=(solvent=water)
geom=connectivity int=acc2e=11

BV-bound

-1 1
 C                  5.34661500    8.73621500    1.49828100
 S                  3.74692100    9.00389500    2.31812900
 C                  9.41768700   -8.18657300    1.44718800
 C                  9.60723800   -7.22676300    0.20867900
 N                  8.27971400   -6.78618800   -0.25366500
 C                  7.60299900   -5.73260700    0.20459400
 N                  8.12583800   -4.85280400    1.10979000
 N                  6.35401200   -5.52679000   -0.20535500
 C                -10.34925000   -6.12852200    2.77272400
 C                 -9.58551500   -4.90119500    2.29449500
 C                 -8.29172100   -4.73774900    1.77864000
 N                -10.17485600   -3.65711100    2.25458900
 C                 -9.29991300   -2.76062900    1.75376600
 N                 -8.15103000   -3.41657300    1.44678900
 C                 -0.41309700   -2.90404100   -2.28507500
 C                 -8.02039400    0.95672600    0.28499200
 C                  2.61827500    7.09038300    0.49758900
 C                  4.43990800   -0.62716500   -0.29095400
 N                 -0.81530900    0.70741100   -1.17089400
 C                 -0.56667400   -3.93122900   -1.14143400
 C                 -8.85704700    1.94918700   -0.07673000
 C                  2.48279900    8.06952500    1.41040200
 C                  4.84177700   -1.05085600    1.14900000
 N                 -4.65181100   -0.37821700   -0.23668700
```

O	-6.41330500	-1.49335400	0.80950000
C	-0.11999400	-5.36917400	-1.50154400
C	5.53033700	-2.41941100	1.15222800
N	1.64651700	1.82062000	-0.55167000
C	1.31294400	-0.53591200	-1.08710600
C	-3.20255200	1.27677200	-1.29694300
C	2.55899100	4.09492400	-0.05970200
C	-3.33368000	-1.68537500	-2.65147400
C	-5.85511700	3.05491200	-1.09778900
C	-0.30332300	8.24947800	-0.41139200
C	5.20853500	2.43795500	0.29228400
C	-0.04000600	-0.45472800	-1.37865100
C	-4.40938200	0.89580100	-0.78509600
C	-0.53914800	5.73810900	-1.00785000
C	2.66444700	2.68087100	-0.21103900
O	0.32612900	-5.58589800	-2.68767300
O	6.65289800	-2.55147100	1.76892900
C	-0.87363800	-1.52536500	-1.90810500
C	-5.67433900	1.64200200	-0.64329300
C	0.24556600	6.86216400	-0.45754100
C	3.85352700	1.90113800	-0.06762700
O	-0.23776000	-6.24086800	-0.56241300
O	4.92424700	-3.36809400	0.52044100
C	-2.15721200	-1.01381300	-1.99457400
C	-6.61141100	0.80514600	-0.06520700
C	1.48170600	6.38334600	-0.08983900
C	3.52422200	0.56079700	-0.34302000
C	-2.08663000	0.36286400	-1.48601800
C	-5.93831300	-0.48392400	0.23024800
C	1.52160800	4.93026500	-0.36236200
C	2.13249200	0.52395200	-0.65786600
N	0.27936800	4.60830500	-0.92150400
O	-1.69917400	5.74564700	-1.48762500
H	8.86740000	-7.67841800	2.21129800
H	10.37632600	-8.47150900	1.82761200
H	10.03285300	-7.81346300	-0.57842600
H	10.11857600	-6.35651500	0.56378800
H	7.84723400	-7.32572000	-0.97606800
H	5.83993700	-4.73892000	0.13373800
H	5.92944500	-6.15995200	-0.85251400
H	7.58017500	-4.07633100	1.42497500
H	9.05416800	-4.98534900	1.45711800
H	-10.45470200	-6.09074200	3.83684400
H	-11.31799400	-6.14509300	2.31867400
H	-7.55221300	-5.50275600	1.66554000
H	-7.33523500	-3.00209800	1.04346400
H	-9.47987200	-1.71419000	1.62154700
H	5.31333800	7.82358500	0.94069600
H	5.54875500	9.55177500	0.83578800
H	-0.98159500	-3.26723300	-3.14999000
H	0.63506800	-2.89144900	-2.60474800
H	-8.40122000	0.14912600	0.90506300
H	3.60701400	6.76015400	0.18894200
H	5.35509900	-0.42089300	-0.86328800
H	3.96744900	-1.49404800	-0.76190700
H	-1.61482700	-3.97378700	-0.81267700
H	0.00772900	-3.60610600	-0.26210600
H	-9.88994300	1.94982300	0.25602700
H	-8.55904900	2.77296000	-0.71549600

```
H                  1.50670400    8.37165900    1.77508200
H                  5.50068900   -0.31058600    1.61099300
H                  3.93410100   -1.12300700    1.76372300
H                 -3.94959400   -1.08932100   -0.10970000
H                  0.63922300    1.96430400   -0.64546300
H                  1.79461400   -1.49955200   -1.21532400
H                 -3.04173100    2.32463300   -1.53215000
H                  3.44135700    4.56858000    0.35297700
H                 -4.08554200   -0.95654600   -2.96761700
H                 -3.83940200   -2.41361900   -2.00418700
H                 -3.00612500   -2.23465500   -3.54230700
H                 -6.39638500    3.10373800   -2.05221400
H                 -6.43880100    3.62704800   -0.36801900
H                 -4.89744100    3.56274300   -1.23875600
H                 -0.99286600    8.41216200   -1.24643600
H                 -0.87060300    8.43716600    0.51083700
H                  0.49513500    8.99692300   -0.46020500
H                  5.99735900    1.74221700   -0.00929800
H                  5.31318900    2.59874800    1.37367500
H                  5.41162700    3.39558300   -0.19943900
H                 -0.00010500    3.72035500   -1.30711900
H                  6.11959600    8.67567900    2.23566700
H                 -9.81176300   -7.01222600    2.49872400
H                  8.88083000   -9.06048400    1.14229000

1 2 1.0 72 1.0 73 1.0 105 1.0
2 22 1.0
3 4 1.0 58 1.0 59 1.0 107 1.0
4 5 1.0 60 1.0 61 1.0
5 6 1.5 62 1.0
6 7 1.5 8 1.5
7 65 1.0 66 1.0
8 63 1.0 64 1.0
9 10 1.0 67 1.0 68 1.0 106 1.0
10 11 1.5 12 1.5
11 14 1.5 69 1.0
12 13 1.5
13 14 1.5 71 1.0
14 70 1.0
15 20 1.0 42 1.0 74 1.0 75 1.0
16 21 2.0 49 1.0 76 1.0
17 22 2.0 50 1.0 77 1.0
18 23 1.0 51 1.0 78 1.0 79 1.0
19 36 1.0 52 1.5
20 26 1.0 80 1.0 81 1.0
21 82 1.0 83 1.0
22 84 1.0
23 27 1.0 85 1.0 86 1.0
24 37 1.0 53 1.5 87 1.0
25 53 2.0
26 40 2.0 46 2.0
27 41 1.5 47 1.5
28 39 1.5 55 1.0 88 1.0
29 36 1.5 55 1.5 89 1.0
30 37 2.0 52 1.0 90 1.0
31 39 1.5 54 2.0 91 1.0
32 48 1.0 92 1.0 93 1.0 94 1.0
33 43 1.0 95 1.0 96 1.0 97 1.0
34 44 1.0 98 1.0 99 1.0 100 1.0
```

```
35 45 1.0 101 1.0 102 1.0 103 1.0
36 42 1.0
37 43 1.0
38 44 1.0 56 1.0 57 2.0
39 45 1.5
40
41
42 48 2.0
43 49 2.0
44 50 2.0
45 51 1.5
46
47
48 52 1.0
49 53 1.0
50 54 1.0
51 55 1.5
52
53
54 56 1.0
55
56 104 1.0
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
```

94
95
96
97
98
99
100
101
102
103
104
105
106
107

Free-BV

1. **HF/3-21g Geometry optimisation**

```
%chk=Free-BV.chk
%mem=100GB
# opt hf/3-21g scrf=(solvent=water) geom=connectivity int=acc2e=11

Free-BV

-2 1
 C                  6.10721354   -2.15686580   -0.42365440
 C                  1.67131740    4.39452603    0.53249447
 C                 -3.14465336    3.03331553    0.86601971
 C                 -6.54734362   -5.20131137    0.62536509
 C                  6.88583603   -2.99054993    0.29829976
 C                  1.58407640    5.26413903   -0.71754553
 C                 -3.85646836    3.62533353   -0.34635129
 C                 -6.33115962   -6.35880337    0.06007109
 C                  1.29530940    6.71861603   -0.41877153
 C                 -5.08314436    4.43885353    0.00836671
 C                  3.99899202    0.05475150   -0.49551871
 C                 -0.40438777    2.16685441    0.01886948
 C                 -2.89474156   -1.92034461    0.59190686
 C                  4.66921552   -4.58985176   -1.64647815
 C                  4.54231469    3.05020174    0.19606516
 C                 -4.91290399    0.46746657    0.93799119
 C                 -4.35720067   -4.00063683    2.49217747
 C                  2.64660496   -3.07877153   -1.50543344
 C                  3.01220343    1.04924697   -0.26394362
 C                 -1.40415777    1.28259741    0.25249948
 C                 -4.11371686   -2.49253646    0.33972499
 O                  0.76128040    6.99508903    0.67520047
 O                 -6.17673036    4.10017853   -0.50191429
 C                  3.97507452   -3.49538076   -0.83701315
 C                  3.21303969    2.37811274    0.04939416
 C                 -2.67688536    1.63229853    0.61961471
 C                 -4.74488067   -3.52568383    1.12083847
 O                  1.62456340    7.56513103   -1.25768253
 O                 -4.93875136    5.40318053    0.79725071
 C                  4.70246447   -2.35165187   -0.72571139
 C                  1.96126140    2.95966303    0.20547247
```

```
C                 -3.43484499    0.48820057    0.66592519
C                 -5.76994505   -4.02977210    0.36450651
C                  3.82923807   -1.21050215   -0.93462782
C                  1.00417672    1.99012969   -0.01451393
C                 -2.57134097   -0.56587824    0.44757077
C                 -5.89764141   -3.17084216   -0.84288968
N                  2.80715441   -1.68162866   -1.70194213
N                  1.65452806    0.80975136   -0.31246023
N                 -1.37239924   -0.09595492    0.11668383
N                 -4.97369832   -2.15655377   -0.69474035
O                  1.67008296   -3.73519253   -1.81939044
O                 -6.63323941   -3.30832816   -1.80618468
H                 -5.51805862   -6.47138537   -0.65768691
H                 -6.96318462   -7.21201237    0.30552309
H                 -7.36852662   -5.12361337    1.33824809
H                  6.70207016   -4.04449004    0.27985136
H                  5.02831602    0.35151850   -0.29201371
H                  6.57783254   -1.25454580   -0.81548240
H                 -3.85604636    3.00802353    1.69095771
H                 -0.73780977    3.18575041   -0.17669852
H                 -2.10087356   -2.58277161    0.93749586
H                 -3.71521467   -4.87711583    2.40469647
H                  1.21481706   -0.07368864   -0.52844923
H                  0.75301740    4.88706903   -1.31406453
H                  2.53347440    5.20130103   -1.24965453
H                 -4.19416536    2.78864553   -0.95766529
H                 -3.15573036    4.27072853   -0.87710529
H                 -3.82020367   -3.20848683    3.01403547
H                 -5.25408167   -4.26303383    3.05383947
H                 -4.93269732   -1.31243277   -1.24726635
H                  4.39078952   -5.56581076   -1.24962515
H                  4.36174652   -4.52113576   -2.69047315
H                  5.74976952   -4.46434576   -1.57603615
H                  4.39797969    4.05820574    0.58621316
H                  5.16444869    2.47944174    0.88546416
H                  5.03218069    3.10463774   -0.77618484
H                 -5.38185099    1.32710057    0.45980419
H                 -5.34425199   -0.45006443    0.53795619
H                 -5.08611299    0.51171157    2.01305619
H                  2.23678641   -1.11933166   -2.31774313
H                  0.70561040    4.43398903    1.03593947
H                  2.46686240    4.77655403    1.17189947
H                 -2.28197836    3.65471753    1.10588571
H                  7.68673232   -2.59738168    0.88896368

1 5 2.0 30 1.0 49 1.0
2 6 1.0 31 1.0 72 1.0 73 1.0
3 7 1.0 26 1.0 50 1.0 74 1.0
4 8 2.0 33 1.0 46 1.0
5 47 1.0 75 1.0
6 9 1.0 55 1.0 56 1.0
7 10 1.0 57 1.0 58 1.0
8 44 1.0 45 1.0
9 22 1.5 28 1.5
10 23 1.5 29 1.5
11 19 1.0 34 2.0 48 1.0
12 20 2.0 35 1.0 51 1.0
13 21 2.0 36 1.0 52 1.0
14 24 1.0 62 1.0 63 1.0 64 1.0
```

15 25 1.0 65 1.0 66 1.0 67 1.0
16 32 1.0 68 1.0 69 1.0 70 1.0
17 27 1.0 53 1.0 59 1.0 60 1.0
18 24 1.0 38 1.0 42 2.0
19 25 1.5 39 1.5
20 26 1.5 40 1.5
21 27 1.0 41 1.0
22
23
24 30 2.0
25 31 1.5
26 32 1.5
27 33 2.0
28
29
30 34 1.0
31 35 1.5
32 36 1.5
33 37 1.0
34 38 1.0
35 39 1.5
36 40 1.5
37 41 1.0 43 2.0
38 71 1.0
39 54 1.0
40
41 61 1.0
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73

```
 74
 75
```

2. **B3LYP/6-31g Geometry optimisation**

```
%chk=Free-BV.chk
%mem=100GB
# opt b3lyp/6-31g scrf=(solvent=water) guess=read geom=checkpoint
int=acc2e=11

Free-BV

-2 1
```

3. **B3LYP/6-31g+(d,p) Geometry optimisation**

```
%chk=Free-BV.chk
%mem=100GB
# opt b3lyp/6-31+g(d,p) scrf=(solvent=water) guess=read geom=checkpoint
int=acc2e=11

Free-BV

-2 1
```

4. **B3LYP/6-31g+(d,p) Analysis – Frequency and TD-DFT (combined for conciseness) – no constraints therefore able to use checkpoint for analysis**

```
%chk=Free-BV.chk
%mem=100GB
# freq=raman td=(nstates=10) b3lyp/6-31+g(d,p) scrf=(solvent=water)
guess=read geom=checkpoint int=acc2e=11

Free-BV

-2 1
```

mCherry WT chromophore

1. HF/3-21g Geometry optimisation

```
%chk=CRO-WT.chk
%mem=100GB
# opt hf/3-21g scrf=(solvent=water) geom=connectivity int=acc2e=11

CRO-WT

-1 1
 C                -1    8.02682800    2.88994100   -1.08729900
 C                 0    5.81110100    1.36050300    0.57752200
 C                 0    3.49838200   -1.32564300   -0.12843800
 C                -1    7.16054300   -1.38043600   -0.43258300
 C                -1    5.02811100    2.59125700    1.02929300
 C                 0    2.29167500   -1.93987300   -0.30501300
 C                 0    0.92971800   -1.56060600   -0.15714100
 C                 0    0.45258100   -0.28993600    0.25495200
 C                 0   -0.05701100   -2.52928500   -0.47867300
 C                 0   -0.88690600   -0.01633400    0.32703000
 C                 0   -1.39643200   -2.27040000   -0.40856600
 C                 0   -1.88680100   -0.99078600   -0.00746600
 O                 0   -3.14660900   -0.72681300    0.04656900
 C                 0    5.09528700    0.10065100    0.26603400
 N                 0    7.05751800    1.31785300    0.40970200
 C                 0    4.74587500   -1.99559800   -0.45373200
 N                 0    3.81355900   -0.02854700    0.31572700
 O                 0    4.94151600   -3.15294200   -0.87193500
 N                 0    5.73127200   -1.06657600   -0.21589300
 C                 0    8.04730300    2.23640500    0.26323800
 O                 0    8.91777300    2.43964100    1.10406200
 H                -1    7.23906200   -2.18419700   -1.19743200
 H                -1    7.63569100   -0.52603100   -0.87167400
 H                -1    5.64473500    3.55453900    0.97087600
 H                -1    4.16455400    2.71595000    0.30554700
 H                -1    7.03031500    3.29618800   -1.25898400
 H                 0    2.41620600   -2.94979000   -0.65852200
 H                -1    7.59497900   -1.71144500    0.51153100
 H                -1    8.74827700    3.70525300   -1.04619600
 H                -1    8.28722300    2.21492400   -1.90192800
 H                 0    0.26704000   -3.50526600   -0.79337000
 H                 0   -2.11355100   -3.02741900   -0.66240300
 H                 0    1.16378000    0.47213700    0.51033800
 H                 0   -1.22546200    0.95394200    0.63701000
 H                -1    4.70449800    2.41243800    2.05529200
 N                -1   -7.67032200    3.55736700   -0.89680800
 C                -1   -8.44876000    2.86830000    0.14765200
 C                -1   -7.62465000    1.88889900    0.89722600
 O                -1   -6.76756000    1.27110000    0.32134500
 C                -1   -9.75154500    2.28563200   -0.53393600
 H                -1   -7.78145600    3.11673900   -1.82586100
 H                -1   -8.65640500    3.60921000    0.79368800
 H                -1  -10.03232200    2.88823500   -1.40753300
 H                -1   -9.47317100    1.24154400   -0.89199600
 H                -1  -10.58715300    2.28524700    0.16543300
 H                -1   -7.09274700    4.36787200   -0.72407100
 H                -1   -7.79538500    1.72935700    1.96289300
 C                -1   -6.64194800   -2.88526200    2.03540400
```

```
C               -1   -6.41006400   -2.73143300    0.50497300
O               -1   -5.96507900   -1.45260100    0.12995800
H               -1   -7.55710300   -3.40789800    2.13647300
H               -1   -5.59266900   -3.45337500    0.13753700
H               -1   -7.37777800   -2.92143600   -0.06984400
H                0   -5.02277900   -1.30060900    0.25582500
H               -1   -5.84399300   -3.45072100    2.51806100
H               -1   -6.65954600   -1.90684800    2.51476300
O               -1   -5.67266800    0.08585100   -2.01366700
H                0   -6.16335100    0.03993500   -1.19628600
H                0   -4.80438200   -0.26900700   -1.85178700

1 20 1.0 26 1.0 29 1.0 30 1.0
2 5 1.0 14 1.0 15 2.0
3 6 2.0 16 1.0 17 1.0
4 19 1.0 22 1.0 23 1.0 28 1.0
5 25 1.0 35 1.0
6 7 1.5 27 1.0
7 8 1.5 9 1.5
8 10 2.0 33 1.0
9 11 2.0 31 1.0
10 12 1.5 34 1.0
11 12 1.5 32 1.0
12 13 1.5
13
14 17 2.0 19 1.0
15 20 1.5
16 18 2.0 19 1.5
17
18
19
20 21 2.0
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36 37 1.0 41 1.0 46 1.0
37 38 1.0 40 1.0 42 1.0
38 39 2.0 47 1.0
39
40 43 1.0 45 1.0
41
42
43
44
45
46
47
```

```
 48 49 1.0 51 1.0 55 1.0 56 1.0
 49 50 1.0
 50 54 1.0
 51
 52
 53
 54
 55
 56
 57 58 1.0 59 1.0
 58
 59
```

2. **B3LYP/6-31g Geometry optimisation**

```
%chk=CRO-WT.chk
%mem=100GB
# opt b3lyp/6-31g scrf=(solvent=water) guess=read geom=checkpoint
int=acc2e=11

CRO-WT

-1 1
```

3. **B3LYP/6-31g+(d,p) Geometry optimisation**

```
%chk=CRO-WT.chk
%mem=100GB
# opt b3lyp/6-31+g(d,p) scrf=(solvent=water) guess=read geom=checkpoint
int=acc2e=11

CRO-WT

-1 1
```

4. **B3LYP/6-31g+(d,p) Analysis – Frequency and TD-DFT (combined for conciseness)**

```
%chk=CRO-WT.chk
%mem=100GB
# freq=raman td=(nstates=10) b3lyp/6-31+g(d,p) scrf=(solvent=water)
geom=connectivity int=acc2e=11

CRO-WT

-1 1
 C                 -8.23395293   -0.44632168   -4.06938234
 C                 -6.37095293   -1.17132168   -1.72538234
 C                 -4.23324362    0.70352876    0.44751528
 C                 -7.95824362    0.93852876    0.10451528
 C                 -5.41295293   -2.15332168   -2.42938234
 C                 -3.08424362    1.21752876    1.14251528
 C                 -1.72620636    0.89536317    1.03322607
 C                 -1.07753014   -0.02122319    0.19711088
 C                 -0.89179586    1.60721177    1.90330117
 C                  0.31302908   -0.06476946    0.04074982
```

```
C                  0.57930981    1.53272032    1.80922832
C                  1.15813013    0.75176979    0.80171489
O                  2.54658884    0.57581564    0.83668014
C                 -5.79781475   -0.27668880   -0.78328646
N                 -7.65595293   -1.29932168   -1.95238234
C                 -5.54724362    1.26952876    0.88151528
N                 -4.46024362   -0.26747124   -0.53748472
O                 -5.85124362    2.07752876    1.78151528
N                 -6.52924362    0.59952876    0.04951528
C                 -8.41895293   -1.27632168   -2.79038234
O                 -9.52695293   -1.55432168   -2.42338234
H                 -8.09924362    1.98552876    0.45251528
H                 -8.29624362    0.93152876   -0.91248472
H                 -5.89695293   -2.55032168   -3.38838234
H                 -4.50395293   -1.55632168   -2.74938234
H                 -7.18595293   -0.46732168   -4.36738234
H                 -3.30824362    1.98952876    1.87851528
H                 -8.48024362    0.24552876    0.76551528
H                 -8.84595293   -0.86532168   -4.86738234
H                 -8.53595293    0.58367832   -3.88238234
H                 -1.30179586    2.24621177    2.68630117
H                  1.22030981    2.06572032    2.51122832
H                 -1.67753014   -0.73822319   -0.36188912
H                  0.75302908   -0.74576946   -0.68725018
H                 -5.16895293   -2.97932168   -1.76038234
N                  7.38753026   -0.72489146   -2.40387448
C                  8.01553026   -1.32789146   -1.21487448
C                  7.03453026   -1.55589146   -0.12587448
O                  6.13653026   -0.77189146    0.03712552
C                  9.26653026   -0.43889146   -0.83287448
H                  7.48953026    0.30410854   -2.42387448
H                  8.28253026   -2.23889146   -1.54387448
H                  9.66753026    0.06510854   -1.72187448
H                  8.87553026    0.35410854   -0.11587448
H                 10.05753026   -1.04989146   -0.39887448
H                  6.91053026   -1.25689146   -3.11787448
H                  7.12753026   -2.42789146    0.52312552
C                  5.38917817   -0.39314630    4.45612295
C                  5.25917817    0.89385370    3.59212295
O                  5.00217817    0.63585370    2.23512295
H                  6.22317817   -0.23214630    5.08812295
H                  4.37617817    1.54085370    3.94712295
H                  6.22317817    1.50385370    3.63112295
H                  4.04417817    0.79485370    2.15812295
H                  4.50117817   -0.57314630    5.06312295
H                  5.50717817   -1.26714630    3.81612295
O                  5.02180832    1.83297307   -0.13426935
H                  5.47180832    1.04897307   -0.44926935
H                  4.23680832    1.50097307    0.30073065

1 20 1.0 26 1.0 29 1.0 30 1.0
2 5 1.0 14 1.0 15 2.0
3 6 2.0 16 1.0 17 1.0
4 19 1.0 22 1.0 23 1.0 28 1.0
5 24 1.0 25 1.0 35 1.0
6 7 1.0 27 1.0
7 8 1.5 9 1.5
8 10 1.5 33 1.0
9 11 1.5 31 1.0
```

```
10 12 1.5 34 1.0
11 12 1.5 32 1.0
12 13 1.0
13
14 17 2.0 19 1.0
15 20 1.0
16 18 2.0 19 1.0
17
18
19
20 21 2.0
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36 37 1.0 41 1.0 46 1.0
37 38 1.0 40 1.0 42 1.0
38 39 2.0 47 1.0
39
40 43 1.0 44 1.0 45 1.0
41
42
43
44
45
46
47
48 49 1.0 51 1.0 55 1.0 56 1.0
49 50 1.0 52 1.0 53 1.0
50 54 1.0
51
52
53
54
55
56
57 58 1.0 59 1.0
58
59
```

mCherry CRO-CN-NC chromophore

1. HF/3-21g Geometry optimisation

```
%chk=CRO-CN-NC.chk
# opt hf/3-21g scrf=(solvent=water) geom=connectivity int=acc2e=11

CRO-CN-NC

0 1
 C                -1    5.45788800   -1.73951500    1.35598500
 C                 0    2.57266800   -0.93850400   -0.16493400
 C                 0   -0.23095900    1.34776200   -0.00041800
 C                -1    3.33922000    2.29307600   -0.16638400
 C                -1    1.85933500   -2.32341900   -0.24758800
 C                 0   -1.51723000    1.66191100    0.05275400
 C                 0   -2.67875600    0.77062500    0.03948600
 C                 0   -2.56955600   -0.61876600   -0.03544600
 C                 0   -3.94633900    1.34659100    0.10448000
 C                 0   -3.70039200   -1.40648500   -0.04400500
 C                 0   -5.08148600    0.56269400    0.09615700
 C                 0   -4.95936600   -0.81848100    0.02209700
 C                 0    1.65535700    0.25571800   -0.10865900
 N                 0    3.80217100   -0.71549500   -0.08053400
 C                 0    0.85215200    2.36283200    0.01695000
 N                 0    0.38643500    0.07921200   -0.08443100
 O                 0    0.77013300    3.57485200    0.06575300
 N                 0    2.01760700    1.62466200   -0.04169600
 C                 0    4.98924100   -1.44600300   -0.05786900
 O                 0    5.58416200   -1.74704300   -1.06980100
 H                 0    3.37996200    3.27121600    0.36170100
 H                -1    4.02925400    1.66145300    0.35666900
 H                -1    2.64897200   -3.13776800   -0.08993700
 H                -1    1.15623900   -2.37906100    0.64004300
 H                -1    4.59461500   -2.22731900    1.80812000
 H                -1   -1.73380900    2.71267200    0.11327100
 H                 0    3.55337900    2.42564800   -1.22753400
 H                -1    6.30580700   -2.42346900    1.36393300
 H                -1    5.70693600   -0.84483700    1.92548400
 H                -1   -4.04465200    2.41203800    0.16175400
 H                 0   -6.05048500    1.01481700    0.14648700
 H                 0   -1.60015300   -1.06303400   -0.08617100
 H                 0   -3.61160500   -2.47198500   -0.10151200
 H                 0    1.35442800   -2.46768600   -1.20333600
 C                -1   -6.12946100   -1.63812400    0.01386500
 N                 0   -7.06294800   -2.29361300    0.00731500

 1 19 1.0 25 1.0 28 1.0 29 1.0
 2 5 1.0 13 1.0 14 2.0
 3 6 2.0 15 1.0 16 1.0
 4 18 1.0 21 1.0 22 1.0 27 1.0
 5 24 1.0 34 1.0
 6 7 1.0 26 1.0
 7 8 1.5 9 1.5
 8 10 2.0 32 1.0
 9 11 2.0 30 1.0
 10 12 1.5 33 1.0
 11 12 1.5 31 1.0
 12 35 1.5
```

```
 13 16 2.0 18 1.0
 14 19 1.0
 15 17 2.0 18 1.5
 16
 17
 18
 19 20 2.0
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35 36 3.0
 36
```

2. **B3LYP/6-31g Geometry optimisation**

```
%chk=CRO-CN-NC.chk
%mem=100GB
# opt b3lyp/6-31g scrf=(solvent=water) guess=read geom=checkpoint
int=acc2e=11

CRO-CN-NC

0 1
```

3. **B3LYP/6-31g+(d,p) Geometry optimisation**

```
%chk=CRO-CN-NC.chk
%mem=100GB
# opt b3lyp/6-31+g(d,p) scrf=(solvent=water) guess=read geom=checkpoint
int=acc2e=11

CRO-CN-NC

0 1
```

4. B3LYP/6-31g+(d,p) Analysis – Frequency and TD-DFT (combined for conciseness)

```
%chk=CRO-CN-NC.chk
%mem=100GB
# freq=raman td=(nstates=10) b3lyp/6-31+g(d,p) scrf=(solvent=water)
geom=connectivity int=acc2e=11

CRO-CN-NC

0 1
 C                  5.35170794   -1.51333991    1.53517954
 C                  2.60776200   -0.96837114   -0.43382810
 C                 -0.57529160    0.90268664   -0.05352155
 C                  2.89493969    2.31795880    0.00125696
 C                  2.18151315   -2.42698315   -0.69549063
 C                 -1.87910910    1.20062203    0.02951286
 C                 -3.07222359    0.33920865   -0.03854690
 C                 -3.11351449   -1.02667054   -0.19122070
 C                 -4.26836496    1.01773338    0.07656173
 C                 -4.29670073   -1.70597898   -0.00791429
 C                 -5.46475818    0.32980632    0.20009572
 C                 -5.49544511   -1.04898083    0.24912960
 C                  1.44834105    0.09284897   -0.22439951
 N                  3.87312725   -0.68113620   -0.51349535
 C                  0.35201723    2.07369481    0.01087216
 N                  0.18720979   -0.26109050   -0.24375924
 O                  0.14239541    3.30134837    0.07572319
 N                  1.67931382    1.49860996   -0.10318703
 C                  5.19214450   -1.02691544    0.08711286
 O                  6.11218936   -0.48068639   -0.45583140
 H                  2.70048775    3.23202784    0.60446539
 H                  3.58490731    1.73863107    0.58169321
 H                  3.05389884   -3.13434050   -0.47158220
 H                  1.36615188   -2.67249829    0.05283986
 H                  4.51512423   -2.16169229    1.79463293
 H                 -2.09328818    2.25962825    0.17265381
 H                  3.24832734    2.58863517   -0.99448681
 H                  6.28402681   -2.06852654    1.63243454
 H                  5.36706805   -0.65647075    2.20794435
 H                 -4.30678331    2.10767848    0.06922025
 H                 -6.40831471    0.87257410    0.25374629
 H                 -2.22712126   -1.59355829   -0.47320566
 H                 -4.29847452   -2.79416757   -0.06437333
 H                  1.86028083   -2.54072634   -1.73146624
 C                 -6.86213216   -1.75776617    0.21184912
 N                 -7.89158475   -2.29165643    0.18376772

 1 19 1.0 25 1.0 28 1.0 29 1.0
 2 5 1.0 13 1.0 14 2.0
 3 6 2.0 15 1.0 16 1.0
 4 18 1.0 21 1.0 22 1.0 27 1.0
 5 23 1.0 24 1.0 34 1.0
 6 7 1.0 26 1.0
 7 8 1.5 9 1.5
 8 10 1.5 32 1.0
 9 11 1.5 30 1.0
 10 12 1.5 33 1.0
 11 12 1.5 31 1.0
 12 35 1.0
```

```
 13 16 2.0 18 1.0
 14 19 1.0
 15 17 2.0 18 1.0
 16
 17
 18
 19 20 2.0
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35 36 3.0
 36
```

mCherry CRO-CN-IC chromophore

1. **HF/3-21g Geometry optimisation**

```
%chk=CRO-CN-IC.chk
%mem=100GB
# opt hf/3-21g scrf=(solvent=water) geom=connectivity int=acc2e=11

CRO-CN-IC

0 1
 C                -1   35.41604785   38.80466047   35.30983685
 C                 0   36.97320000   36.11590000   37.50440000
 C                 0   38.19270000   37.59480000   40.66630000
 C                -1   34.62280000   37.25980000   39.57500000
 C                -1   38.23800000   35.60520000   36.78450000
 C                 0   39.06000000   38.05050000   41.71900000
 C                 0   40.52890000   38.02910000   41.83750000
 C                 0   41.43800000   37.57270000   40.92830000
 C                 0   40.99450000   38.54810000   43.00250000
 C                 0   42.75000000   37.82900000   41.09760000
 C                 0   42.35900000   38.74800000   43.17990000
 C                 0   43.23110000   38.46850000   42.20170000
 C                 0   37.17970000   36.81460000   38.91220000
 N                 0   36.25714785   36.77956047   36.66123685
 C                 0   36.73030000   37.74420000   40.93580000
 N                 0   38.33770000   36.97060000   39.40980000
 O                 0   36.10290000   38.14650000   41.93570000
 N                 0   36.08210000   37.21780000   39.74870000
 C                 0   36.51094785   38.20256047   36.20173685
 O                 0   37.52174785   38.80706047   36.43003685
 H                -1   34.18220000   38.09290000   40.16470000
 H                -1   34.46240000   37.52450000   38.54870000
```

```
H               -1   38.02130000   35.46360000   35.66950000
H               -1   39.01620000   36.42660000   36.86070000
H               -1   34.44604785   38.40866047   35.60753685
H                0   38.54360000   38.48670000   42.57400000
H               -1   34.18460000   36.30760000   39.87620000
H               -1   35.41354785   39.88886047   35.41763685
H               -1   35.60874785   38.54306047   34.26923685
H                0   40.32460000   38.81250000   43.82160000
H                0   42.73360000   39.12790000   44.12940000
H                0   41.12390000   36.98200000   40.06810000
H                0   43.45060000   37.51320000   40.32480000
H               -1   38.56450000   34.66370000   37.22810000
C                0   44.73447129   38.54928811   42.52565686
N                0   45.86688083   38.61014149   42.76967632

1 19 1.0 25 1.0 28 1.0 29 1.0
2 5 1.0 13 1.0 14 2.0
3 6 2.0 15 1.0 16 1.0
4 18 1.0 21 1.0 22 1.0 27 1.0
5 23 1.0 24 1.0 34 1.0
6 7 1.0 26 1.0
7 8 1.5 9 1.5
8 10 1.5 32 1.0
9 11 1.5 30 1.0
10 12 1.5 33 1.0
11 12 1.5 31 1.0
12 35 1.0
13 16 2.0 18 1.0
14 19 1.0
15 17 2.0 18 1.0
16
17
18
19 20 2.0
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35 36 3.0
36
```

2. **B3LYP/6-31g Geometry optimisation**

```
%chk=CRO-CN-IC.chk
%mem=100GB
# opt b3lyp/6-31g scrf=(solvent=water) guess=read geom=checkpoint
int=acc2e=11

CRO-CN-IC

0 1
```

3. **B3LYP/6-31g+(d,p) Geometry optimisation**

```
%chk=CRO-CN-IC.chk
%mem=100GB
# opt b3lyp/6-31+g(d,p) scrf=(solvent=water) guess=read geom=checkpoint
int=acc2e=11

CRO-CN-IC

0 1
```

4. **B3LYP/6-31g+(d,p) Analysis – Frequency and TD-DFT (combined for conciseness)**

```
%chk=CRO-CN-NC.chk
%mem=100GB
# freq=raman td=(nstates=10) b3lyp/6-31+g(d,p) scrf=(solvent=water)
geom=connectivity int=acc2e=11

CRO-CN-NC

0 1
 C                  5.32708800   -1.57514300    1.69418000
 C                  3.06769300   -0.85690900   -1.28063700
 C                  0.10005900    1.06428400   -0.19707500
 C                  3.65338900    2.25574400   -0.23726600
 C                  2.51345100   -2.20743400   -1.77831900
 C                 -1.25733300    1.43165800    0.10387700
 C                 -2.51182200    0.66562800   -0.00392700
 C                 -2.67418900   -0.62663500   -0.41025700
 C                 -3.61579300    1.36256700    0.36875600
 C                 -3.85925900   -1.24479800   -0.23943600
 C                 -4.84232000    0.71717400    0.48005700
 C                 -4.95264800   -0.59973700    0.25841100
 C                  2.01444800    0.19778600   -0.74117700
 N                  4.06832500   -1.08360100   -0.49868600
 C                  1.11155800    2.14708100   -0.00218200
 N                  0.77383800   -0.07005000   -0.69582600
 O                  0.99900800    3.34472600    0.32684200
 N                  2.37676100    1.53887100   -0.37070400
 C                  3.98560500   -1.52773600    0.94920400
 O                  2.97478900   -1.90307900    1.47515800
 H                  3.58393200    3.03529900    0.55229100
 H                  4.35102600    1.53404500    0.13895000
 H                  3.34904200   -2.98960000   -1.79380300
 H                  1.75697900   -2.55786700   -1.00952400
```

```
H                  5.97908700   -0.78724100    1.31941500
H                 -1.37912300    2.44891200    0.47579900
H                  3.92942400    2.70511500   -1.19188500
H                  5.15717100   -1.42851600    2.76036300
H                  5.80058200   -2.54345800    1.53112300
H                 -3.57529000    2.43154700    0.58145600
H                 -5.73378700    1.28574500    0.74123500
H                 -1.86210700   -1.16972600   -0.89306500
H                 -3.94622700   -2.29707900   -0.50949600
H                  2.08743800   -2.09042500   -2.77562900
C                 -6.36437700   -1.21437200    0.22933400
N                 -7.42775700   -1.67734300    0.20743200

1 19 1.0 25 1.0 28 1.0 29 1.0
2 5 1.0 13 1.0 14 2.0
3 6 1.5 15 1.0 16 1.0
4 18 1.0 21 1.0 22 1.0 27 1.0
5 24 1.0 34 1.0
6 7 1.0 26 1.0
7 8 2.0 9 2.0
8 10 2.0 32 1.0
9 11 1.5 30 1.0
10 12 2.0 33 1.0
11 12 2.0 31 1.0
12 35 1.0
13 16 2.0 18 1.0
14 19 1.0
15 17 2.0 18 1.0
16
17
18
19 20 2.0
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35 36 3.0
36
```

mCherry CRO-CC-NC chromophore

1. **HF/3-21g Geometry optimisation**

```
%chk=CRO-CC-NC.chk
# opt hf/3-21g scrf=(solvent=water) geom=connectivity int=acc2e=11

CRO-CC-NC

0 1
 C                -1   15.55892451   -6.71278759    4.55160716
 C                 0   12.83428207   -6.01528476    2.56517217
 C                 0    9.74453181   -3.99405357    2.94647316
 C                -1   13.27910721   -2.74901761    3.01834509
 C                -1   12.33862304   -7.45015263    2.29465224
 C                 0    8.45644374   -3.63367548    3.02745042
 C                 0    7.22318648   -4.43586128    2.95154172
 C                 0    7.11619595   -5.79733162    2.79156664
 C                 0    6.06099935   -3.70073207    3.06689511
 C                 0    5.90092081   -6.41939811    2.96800577
 C                 0    4.83229621   -4.33046034    3.18348079
 C                 0    4.73465998   -5.70637072    3.22517138
 C                 0   11.72703732   -4.90017908    2.77696021
 N                 0   14.10248812   -5.79107924    2.49110928
 C                 0   10.72731586   -2.86973346    3.01960662
 N                 0   10.45029578   -5.19244588    2.75223268
 O                 0   10.57725549   -1.63370766    3.09033554
 N                 0   12.02551450   -3.50791094    2.90621163
 C                 0   15.42751842   -6.21178672    3.10568279
 O                 0   16.37460187   -5.70803931    2.56818032
 H                -1   13.12736583   -1.82969667    3.62581355
 H                -1   13.93841810   -3.36408624    3.59763956
 H                -1   13.17501122   -8.20011779    2.51725832
 H                -1   11.51005419   -7.65967202    3.03941106
 H                -1   14.69111261   -7.32113174    4.80531821
 H                 0    8.29341897   -2.56627649    3.17556755
 H                -1   13.64821731   -2.49069835    2.02502578
 H                -1   16.46293558   -7.31302042    4.64852956
 H                -1   15.61376394   -5.86112863    5.22891286
 H                 0    6.07548382   -2.61018283    3.06518382
 H                 0    3.91600133   -3.74285969    3.23736470
 H                 0    7.97491795   -6.40508205    2.50906654
 H                 0    5.84657018   -7.50591719    2.90581520
 H                -1   12.01539648   -7.54287715    1.25720339
 C                 0    3.33534046   -6.34785106    3.18036233
 C                 0    2.24950486   -6.84562315    3.14559168
 H                 0    1.28542823   -7.28757811    3.11471999

 1 19 1.0 25 1.0 28 1.0 29 1.0
 2 5 1.0 13 1.0 14 2.0
 3 6 2.0 15 1.0 16 1.0
 4 18 1.0 21 1.0 22 1.0 27 1.0
 5 23 1.0 24 1.0 34 1.0
 6 7 1.0 26 1.0
 7 8 1.5 9 1.5
 8 10 1.5 32 1.0
 9 11 1.5 30 1.0
 10 12 1.5 33 1.0
 11 12 1.5 31 1.0
```

```
 12 35 1.0
 13 16 1.5 18 1.0
 14 19 1.0
 15 17 2.0 18 1.0
 16
 17
 18
 19 20 2.0
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35 36 3.0
 36 37 1.0
 37
```

2. **B3LYP/6-31g Geometry optimisation**

```
%chk=CRO-CC-NC.chk
%mem=100GB
# opt b3lyp/6-31g scrf=(solvent=water) guess=read geom=checkpoint
int=acc2e=11

CRO-CC-NC

0 1
```

3. **B3LYP/6-31g+(d,p) Geometry optimisation**

```
%chk=CRO-CC-NC.chk
%mem=100GB
# opt b3lyp/6-31+g(d,p) scrf=(solvent=water) guess=read geom=checkpoint
int=acc2e=11

CRO-CC-NC

0 1
```

4. B3LYP/6-31g+(d,p) Analysis – Frequency and TD-DFT (combined for conciseness)

```
%chk=CRO-CC-NC.chk
%mem=100GB
# freq=raman td=(nstates=10) b3lyp/6-31+g(d,p) scrf=(solvent=water)
geom=connectivity int=acc2e=11

CRO-CC-NC

0 1
 C                  5.55892451   -1.71278759    1.55160716
 C                  2.83428207   -1.01528476   -0.43482783
 C                 -0.25546819    1.00594643   -0.05352684
 C                  3.27910721    2.25098239    0.01834509
 C                  2.33862304   -2.45015263   -0.70534776
 C                 -1.54355626    1.36632452    0.02745042
 C                 -2.77681352    0.56413872   -0.04845828
 C                 -2.88380405   -0.79733162   -0.20843336
 C                 -3.93900065    1.29926793    0.06689511
 C                 -4.09907919   -1.41939811   -0.03199423
 C                 -5.16770379    0.66953966    0.18348079
 C                 -5.26534002   -0.70637072    0.22517138
 C                  1.72703732    0.09982092   -0.22303979
 N                  4.10248812   -0.79107924   -0.50889072
 C                  0.72731586    2.13026654    0.01960662
 N                  0.45029578   -0.19244588   -0.24776732
 O                  0.57725549    3.36629234    0.09033554
 N                  2.02551450    1.49208906   -0.09378837
 C                  5.42751842   -1.21178672    0.10568279
 O                  6.37460187   -0.70803931   -0.43181968
 H                  3.12736583    3.17030333    0.62581355
 H                  3.93841810    1.63591376    0.59763956
 H                  3.17501122   -3.20011779   -0.48274168
 H                  1.51005419   -2.65967202    0.03941106
 H                  4.69111261   -2.32113174    1.80531821
 H                 -1.70658103    2.43372351    0.17556755
 H                  3.64821731    2.50930165   -0.97497422
 H                  6.46293558   -2.31302042    1.64852956
 H                  5.61376394   -0.86112863    2.22891286
 H                 -3.92451618    2.38981717    0.06518382
 H                 -6.08399867    1.25714031    0.23736470
 H                 -2.02508205   -1.40508205   -0.49093346
 H                 -4.15342982   -2.50591719   -0.09418480
 H                  2.01539648   -2.54287715   -1.74279661
 C                 -6.66465954   -1.34785106    0.18036233
 C                 -7.75049514   -1.84562315    0.14559168
 H                 -8.71457177   -2.28757811    0.11471999

 1 19 1.0 25 1.0 28 1.0 29 1.0
 2 5 1.0 13 1.0 14 2.0
 3 6 2.0 15 1.0 16 1.0
 4 18 1.0 21 1.0 22 1.0 27 1.0
 5 23 1.0 24 1.0 34 1.0
 6 7 1.0 26 1.0
 7 8 1.5 9 1.5
 8 10 1.5 32 1.0
 9 11 1.5 30 1.0
 10 12 1.5 33 1.0
 11 12 1.5 31 1.0
```

```
 12 35 1.0
 13 16 1.5 18 1.0
 14 19 1.0
 15 17 2.0 18 1.0
 16
 17
 18
 19 20 2.0
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35 36 3.0
 36 37 1.0
 37
```

mCherry CRO-CC-IC chromophore

1. **HF/3-21g Geometry optimisation**

```
%chk=CRO-CC-IC.chk
# opt hf/3-21g scrf=(solvent=water) geom=connectivity int=acc2e=11

CRO-CC-IC

0 1
 C                -1   35.41604785   38.80466048   35.30983685
 C                 0   36.97320000   36.11590000   37.50440000
 C                 0   38.19270000   37.59480000   40.66630000
 C                -1   34.62280000   37.25980000   39.57500000
 C                -1   38.23800000   35.60520000   36.78450000
 C                 0   39.06000000   38.05050000   41.71900000
 C                 0   40.52890000   38.02910000   41.83750000
 C                 0   41.43800000   37.57270000   40.92830000
 C                 0   40.99450000   38.54810000   43.00250000
 C                 0   42.75000000   37.82900000   41.09760000
 C                 0   42.35900000   38.74800000   43.17990000
 C                 0   43.23110000   38.46850000   42.20170000
 C                 0   37.17970000   36.81460000   38.91220000
 N                 0   36.25714785   36.77956048   36.66123685
 C                 0   36.73030000   37.74420000   40.93580000
 N                 0   38.33770000   36.97060000   39.40980000
 O                 0   36.10290000   38.14650000   41.93570000
 N                 0   36.08210000   37.21780000   39.74870000
 C                 0   36.51094785   38.20256048   36.20173685
 O                 0   37.52174785   38.80706048   36.43003685
 H                -1   34.18220000   38.09290000   40.16470000
```

```
H                -1   34.46240000   37.52450000   38.54870000
H                -1   38.02130000   35.46360000   35.66950000
H                -1   39.01620000   36.42660000   36.86070000
H                -1   34.44604785   38.40866048   35.60753685
H                 0   38.54360000   38.48670000   42.57400000
H                -1   34.18460000   36.30760000   39.87620000
H                -1   35.41354785   39.88886048   35.41763685
H                -1   35.60874785   38.54306048   34.26923685
H                 0   40.32460000   38.81250000   43.82160000
H                 0   42.73360000   39.12790000   44.12940000
H                 0   41.12390000   36.98200000   40.06810000
H                 0   43.45060000   37.51320000   40.32480000
H                -1   38.56450000   34.66370000   37.22810000
C                 0   44.73447129   38.54928811   42.52565686
C                 0   45.90104836   38.61197758   42.77703897
H                 0   46.93681261   38.66763745   43.00023262

1 19 1.0 25 1.0 28 1.0 29 1.0
2 5 1.0 13 1.0 14 2.0
3 6 2.0 15 1.0 16 1.0
4 18 1.0 21 1.0 22 1.0 27 1.0
5 23 1.0 24 1.0 34 1.0
6 7 1.0 26 1.0
7 8 1.5 9 1.5
8 10 1.5 32 1.0
9 11 1.5 30 1.0
10 12 1.5 33 1.0
11 12 1.5 31 1.0
12 35 1.0
13 16 2.0 18 1.0
14 19 1.0
15 17 2.0 18 1.0
16
17
18
19 20 2.0
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35 36 3.0
36 37 1.0
37
```

2. **B3LYP/6-31g Geometry optimisation**

```
%chk=CRO-CC-IC.chk
%mem=100GB
# opt b3lyp/6-31g scrf=(solvent=water) guess=read geom=checkpoint
int=acc2e=11

CRO-CC-IC

0 1
```

3. **B3LYP/6-31g+(d,p) Geometry optimisation**

```
%chk=CRO-CC-IC.chk
%mem=100GB
# opt b3lyp/6-31+g(d,p) scrf=(solvent=water) guess=read geom=checkpoint
int=acc2e=11

CRO-CC-IC

0 1
```

4. **B3LYP/6-31g+(d,p) Analysis – Frequency and TD-DFT (combined for conciseness)**

```
%chk=CRO-CC-IC.chk
%mem=100GB
# freq=raman td=(nstates=10) b3lyp/6-31+g(d,p) scrf=(solvent=water)
geom=connectivity int=acc2e=11

CRO-CC-IC

0 1
 C                  6.16612900   -1.23520500    0.34739600
 C                  2.60128100   -1.00170000   -0.61758400
 C                 -0.09622100    1.26985600   -0.02230600
 C                  3.35538900    2.25421700   -0.71430500
 C                  1.97716500   -2.40090000   -0.86311500
 C                 -1.39627200    1.58925700    0.23429200
 C                 -2.59487200    0.77807700    0.16024500
 C                 -2.60636800   -0.58268500   -0.22824600
 C                 -3.82565700    1.38813000    0.49596600
 C                 -3.79668000   -1.29431400   -0.27617600
 C                 -5.01851800    0.67910800    0.44928900
 C                 -5.02145800   -0.67644300    0.06126500
 C                  1.72401400    0.20351100   -0.52029500
 N                  3.78913600   -0.76670200   -0.18318700
 C                  1.01611900    2.25007100    0.09593900
 N                  0.43001200    0.04509100   -0.42267200
 O                  0.98255700    3.45492200    0.34289500
 N                  2.16838100    1.51578500   -0.19112200
 C                  4.68830600   -1.51446600    0.53464100
 O                  4.30944800   -2.29293000    1.41795500
 H                  3.66054900    3.13659200   -0.11072200
 H                  4.20842100    1.60552000   -0.71820900
 H                  2.73408200   -3.16274800   -1.25948100
 H                  1.80739100   -2.65233400    0.22946700
```

```
H                   6.42602000   -0.49668600   -0.40945100
H                  -1.54738100    2.62204400    0.54242900
H                   3.01495600    2.56918000   -1.70173100
H                   6.60018300   -0.94217400    1.30343100
H                   6.55910700   -2.20834900    0.05327400
H                  -3.83744900    2.43187600    0.79609700
H                  -5.95267700    1.16465400    0.71053800
H                  -1.67449800   -1.06830200   -0.48946900
H                  -3.79226700   -2.33700800   -0.57605800
H                   1.06278800   -2.43358400   -1.45714600
C                  -6.24380200   -1.41413300    0.00878800
C                  -7.28224000   -2.04089000   -0.03661500
H                  -8.19586100   -2.59278300   -0.07598900

1 19 1.0 25 1.0 28 1.0 29 1.0
2 5 1.0 13 1.0 14 2.0
3 6 2.0 15 1.0 16 1.0
4 18 1.0 21 1.0 22 1.0 27 1.0
5 24 1.0 34 1.0
6 7 1.0 26 1.0
7 8 1.5 9 1.5
8 10 1.5 32 1.0
9 11 1.5 30 1.0
10 12 1.5 33 1.0
11 12 1.5 31 1.0
12 35 1.5
13 16 2.0 18 1.0
14 19 1.5
15 17 2.0 18 1.0
16
17
18
19 20 2.0
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35 36 3.0
36 37 1.0
37
```