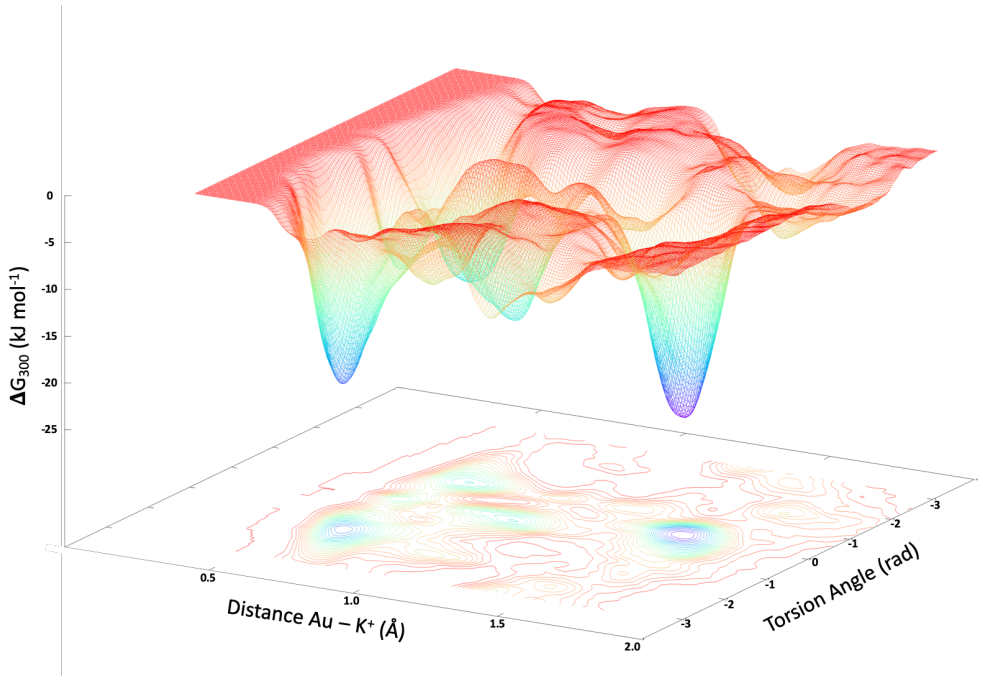
NOVEL COMPUTATIONAL METHODS FOR BIO-MECHANISTIC AND METALLODRUG DISCOVERY

PhD Thesis



Darren Dennis Wragg

Cardiff University
July 2019

Declaration

Statement 1

This thesis is being submitted in partial fulfilment of the requirements for the degree of PhD.

Signed: …… **Date: ……**

Statement 2

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently for any other degree or award (outside of any formal collaboration agreement between the University and a partner organisation).

Signed: …… **Date: ……**

Statement 3

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository (or where approved, to be available in the University's library and for inter-library loan) and for the title and summary to be made available to outside organisations, subject to the expiry of a University-approved bar on access if applicable

Signed: …… **Date: ……**

Declaration

This thesis is the result of my own independent work, except where otherwise stated, and the views expressed are my own. Other sources have been acknowledged by explicit references. The thesis has not been edited by a third party beyond what is permitted by Cardiff University's Policy on the Use of Third-Party Editors by Research Degree Students Procedure.

Signed: …… **Date: ……**

WORD COUNT: 35,329

(Excluding summary, acknowledgments, declarations, contents pages, appendices, tables, diagrams and figure, references, bibliography, footnotes and endnotes)

Table of Contents

Acknowledgments

The completion of this doctorate is the culmination of a second chance in academia. Whilst the first was a forgettable experience, the second has been one of the most rewarding experiences of my life and has come in no small part through the constant encouragement of my wife Fiona.

The biggest thanks must go to Angela and Stefano for showing such (blind) faith in taking on a fresh faced(!) straight out of an MSc (mainly) organic chemist, to take on what I now know to be a huge and complex computational chemistry project. My initial naiveté helped as I was able to approach the task without the preconceptions of what was and was not possible.

The biggest joy of computational chemistry is that you can try things out and have the "what happens if I do this" moments without the fear of blowing the lab up. You just get an incomprehensible error message instead, which, whilst annoying, saves having to explain to your supervisor why the fire brigade has turned up.

As I belong to two research groups there are quite a few people to thank. First of all, from the Casini group thank you to Dr Andreia de Almeida, my daily supervisor for the majority of my time here and her husband Tiago (unofficial group mascot). Whilst Andreia taught me about the joys and importance of aquaporins, I taught her about the subtleties of the English language and northern dialects. She proved to be such a fantastic student that she now acts as a translator to almost everyone we meet, including the natives. Thank you also to Dr Riccardo Bonsignore for his insights on both science and life from a Sicilian's perspective (mainly what you can and can't have on pizza and pasta; pineapple is a definite no, with which I agree) and meatballs should never be served with spaghetti apparently – who knew?!

I would like to thank the rest of the group; Ben Woods, Brech Aikman and Sophie Thomas, my fellow PhD students, along with Dr Margot Wenzel and Dr Samuel Meier-Menches, post-docs who spent time with our group. Also thank you to all the visiting students who spent time with our group over the past three years.

Now for the Leoni group with whom I shared an office; Dr Samuel Jobbins, Dr Timothy Flack, and Dr Duncan Hardie, all of whom were PhD students when I began and have now gone on to become doctors and experts in their field. Also, Jordan Rundle, who decided to join our group in spite of my supervision during his master's project. I'd also like to thank Riccardo Rozza, a visiting student from Palermo. I would especially like to thank them all for welcoming me into the group even though I'm older and northern; they didn't discriminate. They could not convince me that rugby is a superior sport to football, though. Even their sometimes-terrible taste in music didn't spoil the enjoyment of sharing an office with them all.

Lastly, but by no means least, I would like to thank my parents, Mary and Martin, my other parents (Fiona's parents) Margaret and Bill along with all the siblings and their partners; Peter and Kerry, Andrew and Deborah and Heather and Paul, and all my nieces and nephews. Also, my friends Chris, Craig and Steve who I abandoned to come to Wales but are always willing visitors or visitees when I need a pint or anything else really.

This experience has led me to make new friends and meet some remarkable people and has been a most enjoyable experience.

Thank you all.

Abstract

The growth of molecular modelling and free-energy calculations for drug discovery and development have provided a platform for the work undertaken in this thesis. The known affinity of certain metals for biological targets, such as gold/thiol affinity, has guided this research on the characterisation of interactions involved in both protein function inhibition and non-canonical DNA sequence stabilisation. The ability to calculate microsecond timeframes of our biological targets has allowed for greater understanding of these mechanisms at a molecular level.

The research in this thesis is organised into two sections. The first part covers the mechanism of permeation and inhibition of AQP3 *via* a number of molecular dynamic techniques, such as umbrella sampling and metadynamics. The effects of a Au(III) complex on the protein conformation of AQP3 upon metal binding were investigated using steered molecular dynamics. The pH induced gating mechanism of AQP7 and its effect on both water and glycerol permeation was studied to understand the protein conformational changes involved.

The second part investigates the stabilising effect of Au(I) N-heterocyclic carbenes (NHCs) on G-quadruplex DNA structures. The metadynamics calculated free-energy results were then compared to FRET melting assay results to determine the mechanism of interaction.

Throughout this work we have shown how the use of *in silico* methods can enhance our mechanistic knowledge and understanding of biological systems, helping validate and explain experimental results, and vice versa. We have used state of the art molecular dynamics techniques, either rarely or not yet used for such complex systems, thus, furthering the knowledge of our group, and through published work, the wider scientific community

Chapter 1 - Introduction

Section 1

1.1 Use of computational tools for the study of biological systems

The continuing increase in computational power, from institution and industry funded supercomputers to everyday desktops and laptops, has enabled the development and study of biological systems in great detail, more than ever before[1]. The advancement in software packages designed specifically for calculations on biological system, such as Gromacs[2], Amber[3] and Desmond[4], have provided new tools to enhance our understanding of biological models[5]. A number of techniques are available whose suitability will depend on the system and the reaction/interaction under investigation. For small systems with a few atoms, detailed quantum mechanical density functional theory (DFT)[6] calculations can resolve bond formation and bond breaking processes along with optimising geometries of small molecules. For larger more complex biological systems with multiple proteins and ligands all-atom molecular mechanics can be undertaken. The number of degrees of freedom on further increasing system size can be reduced by means of coarse-grained molecular dynamics, which eliminates fine interaction details and allow for a significant increase in simulation time and length scale (Figure 1.1) to be undertaken.

Overall, computational chemistry is able to provide insights at levels beyond that of conventional biophysical and experimental techniques[7]. For example, *in silico* experiments can provide missing or hardly accessible experimental information on the shortest timescales with atomistic levels of detail. Thus, processes at a range of timescales from bond stretching (femtosecond) to protein folding (microsecond) can be viewed and analysed using *in silico* calculations, providing information on subjects as diverse as drug screening, through to manufacturing of novel bio-materials for renewable energy[1].

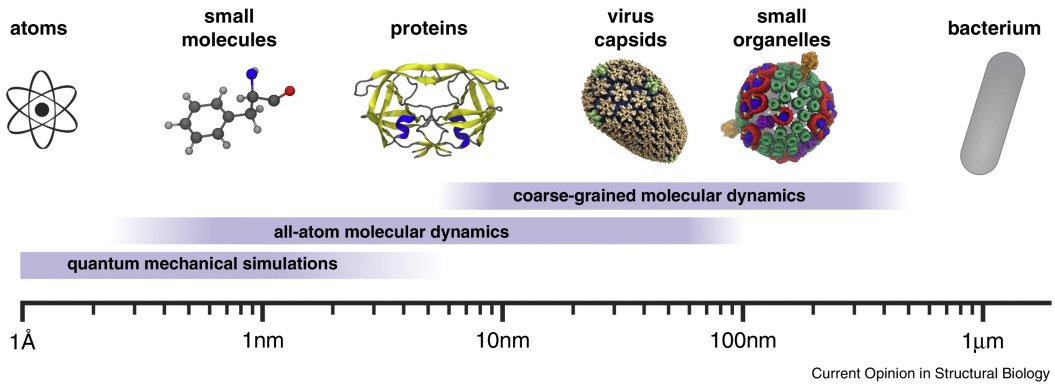


Figure 1.1 Timescales and associated theory level of biomolecular simulations based on the size of the system under investigation, from single atoms to full biological systems (Used with permission from Elsevier, License number 4366940859309).[1]

Even with the array of available analytical techniques, such as Nuclear Magnetic Resonance (NMR) and Infra-Red (IR) spectroscopies, High Performance Liquid Chromatography (HPLC), Electrospray Ionisation Mass Spectrometry (ESI-MS) and Fluorescence Resonance Energy Transfer (FRET), details on the dynamic nature of biomolecule/biomolecule (e.g. protein/protein) or biomolecule/ligand interactions in complex biosystems remain elusive. These techniques mainly provide an averaged snapshot of the overall process being investagated[7] rather than the intricacies of bond formation and electronic interactions. Thus, a combination of experimental and computational techniques can bring greater insights into the processes involved at an atomic and sub-atomic level. Moreover, whilst X-ray crystallography can identify structural conformations of biomolecules under highly specific and non-physiological conditions, this technique lacks any dynamic information. The movement of amino acid sidechains or protein loops at temperatures and concentrations that are biologically relevant cannot be ascertained with this technique, requiring therefore, assumptions gleaned from further experimental techniques in order to achieve an accurate picture. Nowadays, by using *in silico* methods in conjunction with crystallographic data and other experimental methods of structural assessment, detailed knowledge of dynamic processes can be gained, as discussed in the next sections.

1.1.2 All-atom calculations

System properties can be accessed from atomistic methods by suitable measurements. This entails knowing position and velocities of every particle as a function of time, as they are affected by interatomic forces. While for small systems quantum mechanics (DFT) represents a universal potential, large biological systems are beyond the scope of those methods and require therefore approximate model potentials. Therein, the potential energy dependency on single atom coordinates, bond stretching, angles and dihedrals is parametrically coded into a so-called force field, in which electronic degrees of freedom are effectively replaced by a set of interaction potentials. While some of the universality of the *ab initio* approach is lost, this widespread approach allows for the calculation of very large systems up to highly complex viruses scaffolds constituted by 10's of millions of atoms[1].

As the main scope of the work carried out in this thesis is based on molecular dynamics calculations enhanced by both established and state of the art enhanced sampling methods, the remainder of this chapter will briefly review important aspects of this area. Firstly, an introduction to biological relevant force fields, followed by an introduction to enhanced sampling methods, homology modelling and multiscale modelling.

This will be followed by a more in depth look at the general force field parametrisation and molecular dynamics statistical ensembles. Since the processes of interest contain activated steps, which poses specific challenges to molecular simulations, an overview of the free energy methods used in the subsequent results chapters is given here.

1.1.3 Molecular dynamics and the use of force fields

The motion of an object can be fully described by knowing its position as a function of time. Motion is affected by velocity changes, velocity changes in turn results from forces acting on an object. The acceleration of an object is therefore the second derivative of position, which is caused by a force acting on an object of mass *m*, according to Newton's second law. Molecular dynamics (MD), one widely used

atomistic method for biosystems is based on the applicability of classical laws of motions to microscopic systems. The process of performing MD simulations relies therefore on solving a set of Newton's equations of motion, one for every particle:

$$F_i = m_i a_i \qquad \text{(Eq.1)}$$

where F is the force operating on a particle i, m is its mass and a the resulting acceleration[7]. The potential from which forces are calculated roughly defines different flavours of molecular dynamics. While ab initio MD is commonly based on DFT, in classical MD, relevant for biological system, forces are evaluated based on force fields. In MD, forces acting on each atom are calculated iteratively, from which positions and velocities are updated and the system propagated[7]. The system of differential equations of motion derived from Eq.1 is solved by means of finite difference methods, efficiently implemented into algorithms, among which the most popular is the Velocity Verlet method. The latter allows solving the equation of motions on a step-by-step basis, based on a timestep, i.e. the shortest time interval in a system. A constant timestep and the total number of integration steps are explicitly set based on the system under investigation: a commonly used timestep is in the order of 1 femtosecond (10^{-15} s), with the number of steps equating to a range from picoseconds (10^{-12} s) to microseconds (10^{-6} s), depending on the type of interactions under investigation. This motion machinery, the so-called integrator, is independent from the portion that evaluate forces, and is therefore transferable among different MD calculation types.

MD allows elucidating dynamic processes which are extremely difficult to observe experimentally[7] due to too short timescales or lack of spatial resolution. This means that MD simulations can be thought of as a computational microscope[8]. The output of such calculations contains the coordinates and momenta of every atom in the system as a function of time and is termed a trajectory. The latter can be viewed as a sequence of snapshots of the system over the timeframe of the calculation[7], spaced by a timestep. Some insights can therefore be gained by inspection, like molecular docking, hydrogen bonding, π-π stacking, active site size, shape and electrostatic nature of the interactions. However, thermodynamic properties require averaging over trajectories. Free energy calculations such as molecular binding affinity, membrane permeation

events and transport mechanisms entail assessing the importance of certain configurations, which cannot be expressed as a function of single snapshots but require dedicated configuration weighting schemes. This can be achieved for example using methods such as umbrella sampling[9] and metadynamics[10] (section 1.3), both used in this work.

1.1.4 Steered molecular dynamics and enhanced sampling

The occurrence of a certain configuration or event is biased by its probability. The presence of energy barriers in a system may nonetheless prevent certain, important regions to be visited within a given simulation time. This in turn may bias time-averaged properties. This situation arises when the effective practical simulation time is shorter than the residence time in one particular state, affecting not only the overall statistics but also the very chance of observing the event of interest at all. Therefore, different approaches beyond standard MD are needed. Their focus is on enhancing the occurrence of certain events, in order to achieve a correct weighting of configurations. The occurrence of important events can be enhanced by using an external bias to increase the probability of the desired event happening within a shorter timeframe[11]. Examples of which include centre of mass (COM) pulling (section 1.3), whereby a harmonic restraint is used on the centre of mass of a specified atom or molecule to encourage it to sample the configurational space along a defined pathway in relation to the rest of the system. Examples of this are the translocation of a molecule through a lipid bilayer or a protein, as implemented for example by de Groot in investigating the free energies of small uncharged molecules passing through a membrane protein pore[12].

An enhanced free energy sampling method is metadynamics, whereby the event under investigation can be accelerated along a number of selected reaction coordinates, so called collective variables (CV). A CV can be as simple as the distance between two groups of atoms or the change in the coordination number of an ion during a reaction, or it can result from the combinations of several degrees of freedom. Metadynamics attempts to explore the free energy landscape (FES) of a system by depositing Gaussian functions biases to fill FES wells, and in doing so, it encourages the system to explore the whole FE landscape based on the CV. Metadynamics implements

therefore a coarse-grained dynamic in the space of CVs. By moving the system out of one FES well, over a saddle point and into an adjacent minimum (Figure 1.2) it allows the discovery of either lower minima states or metastable states which could provide areas for further exploration or enhancement. In biological systems, these sites can correspond to drug target sites or intermediate configurations, which allows to gain more insight into a reaction mechanism. As an example, in 2017, Moraca et al [13] reported on an advanced funnel metadynamics technique proved to be effective when used to study the interactions of G-quadruplex DNA with a stabilising organic ligand and to determine the binding energies involved.

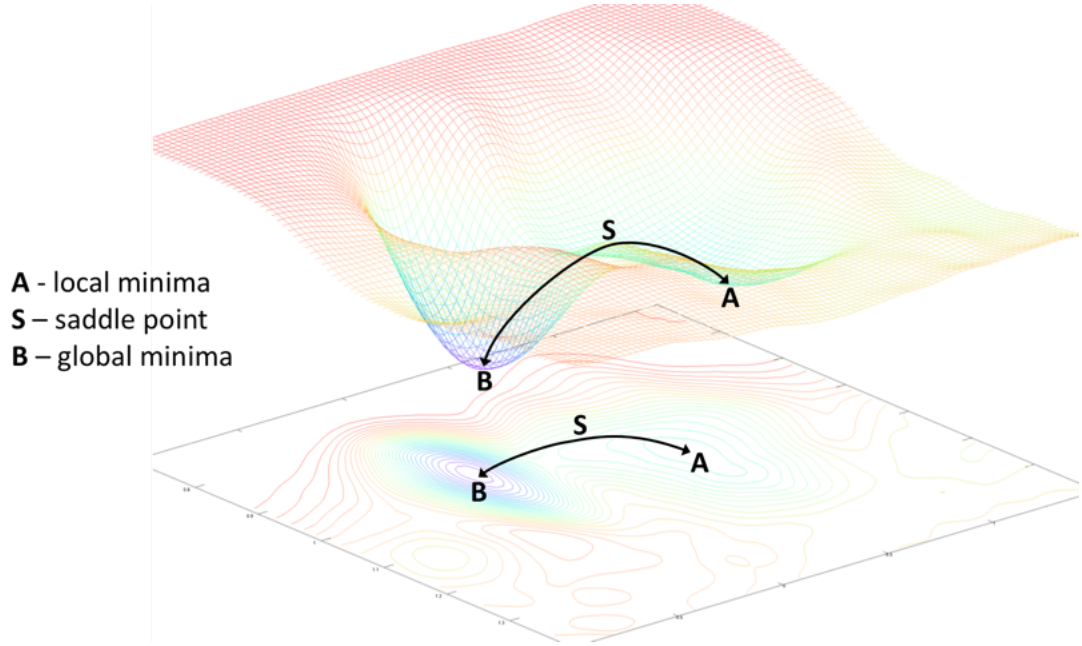


Figure 1.2 Free energy surface showing local minima and global minima with saddle point (highest point between the two minima).

1.1.5 Homology modelling

Molecular dynamics simulations are started from an initial configuration. For biological systems, this is commonly provided by an X-ray crystallography structure or homology model of the required protein or nucleic sequence being studied. Homology modelling works on the principle of evolutionary conservation[7]. Accordingly, if the structure of one isoform of a protein is known and the amino acid sequence of a related isoform can be ascertained, then it is possible to accurately model

this protein using the known isoform structure as a template. This can work between species. Therefore, if we know the structure of a viral or plant protein, which has a mammalian equivalent, the mammalian protein can be modelled. The accuracy is dependent on the sequence similarity; if the sequences have >50% similarity the homology model should have a high degree of accuracy. A reasonably accurate model can be attained with a sequence similarity above 25-30%. The knowledge of the structures in question is needed to assess the accuracy of the homology model as a small change in the amino acid sequence can alter the hydrogen bond network needed to maintain the proteins secondary structure, thus effecting the tertiary structure and protein shape.

Recently, a study by Nikolaev and co-workers[14] reported on the accuracy of a number of protein structure prediction methods to predict the structure of rhodopsins. In total 36 rhodopsin sequences were derived using the programs MEDELLER[15], I-TASSER[16] and Rosetta[17] and compared to the 24 rhodopsin structures available in the RCSB database. The quality of each model was assessed using root mean square deviation (RMSD) with the results concluding that with a sequence similarity of above 40% the predicted structures were found to be within 1.4 Å of the crystal structures[14]. It was also shown that with a sequence similarity of between 20-40% the structures were accurate to within 4.0 Å of the target. Comparison of the different homology methods used concluded that each produced accurate models overall, with the main issue being that of the flexible loop regions of rhodopsin. Of note is the need for further optimisation of the active sites when using I-TASSER or Rosetta to remove occurrences of steric clashes.

1.1.6 Multiscale methods – coarse grained (CG) and quantum mechanical/molecular mechanics (QM/MM)

Two other techniques have been widely used with biological systems, namely quantum mechanical/molecular mechanics (QM/MM) and coarse-grained modelling (CG). As with any modelling system, each one has its advantages and disadvantages. QM/MM uses a combination of molecular dynamics and quantum chemistry calculations, allowing a greater accuracy of chemical processes than classical MD, yet

with larger solvated systems than is possible with DFT. Whilst this method relies on the use of a force field to describe the majority of the atoms, for a small number of atoms within the system it allows to calculate the formation or breaking of covalent bonds. This is not possible with purely force field-based techniques, which are ill-behaved for configurations deviating from reference structures. This technique can be useful for enzymatic systems where the catalytic mechanism is of interest[7], for which the active site can be treated at DFT level, while the rest of the system and the solvent are typically described with molecular mechanics. However, QM/MM calculations are limited by the size of the system, the number of atoms needed to be described at the quantum mechanical level and, due to this, the time taken for each calculation, meaning shorter timescales compared to MM calculations, which could make it less practical to use when longer, free energy calculations are required[7]. The main constraint of a QM/MM approach is represented by the separation of the QM from the MM region, which poses problems when atoms switch from MM to QM, or vice versa. While there are adaptive QM/MM methods, the existence of two Hamiltonians, one per region remains the central problem of this approach.

At the other end of the length scale we find coarse grained modelling (CG). In this technique atoms are no longer described explicitly, instead groups of atoms are represented as pseudo atoms. In this way an amino acid, carbon chain or a number of water molecules can be represented as a single sphere in the model. Therefore, a protein containing thousands of individual atoms can be represented by a few hundred pseudo atoms. Whilst accuracy is reduced, system size and calculation timescales can be increased vastly as the number of degrees of freedom is strongly reduced. For large systems containing a number of proteins or to model the assembly of a virus capsids[1], the need for calculating each individual atom is less urgent, and so by using CG long scale biological processes can be calculated.

1.2 Force fields, an introduction

The use of force field offers a handy method to evaluate the potential energy of a system, without explicitly performing an electronic energy calculation. Therein, the "exact" energy is expressed as a (truncated) expansion of interatomic potential terms[18]. A force field typically contains terms accounting for bonds, angles and can include some finer electronic effects in parametric form[19]. In this way forces can be evaluated with enough efficiency to allow for long MD of large systems, which are typical for biological problems. Therein, atoms and molecules are mapped onto 'ball and spring' models with parameters defining atom size, bond length, angle, dihedral angle and partial charges[18]. As individual terms are obtained from quantum chemical calculations on small molecules, within the forcefield atom types are used to refer to parameter sets for, say, a carbon in different environments such as a methyl or carboxylic acid[18]. Therefore, within the calculation these molecular environments will react and interact consistently. This highlights one of the main drawbacks in the use of force fields, that is the transferability of parameters. While, to some extent, similar molecular fragments can be expected to be described by similar parameters, there are no means to decide when and under which conditions a force field will behave less reliably[18].

The force field energies are calculated as the sum of the energy functions pertaining to different terms;

$$E_{FF} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{el} + E_{cross} \qquad \text{(Eq.2)}^{18}$$

Each term refers to the energies involved in the distortion of a molecule; E_{str} describes bond stretching, E_{bend} describes bond angle bending, E_{tors} describes torsional rotation, E_{vdw} and E_{el} describe the *non-bonded* interactions between atoms and E_{cross} describes the coupling of E_{str} + E_{bend} + E_{tors}[18]. Energy parameters can be obtained from experimental techniques and *ab initio* calculations. Figure 1.3 summarises the force field terms and their meaning.

By applying these energy parameters to the nuclear coordinates of a model system, such as a protein, forces can be evaluated, and the optimised geometry of the system can be calculated.

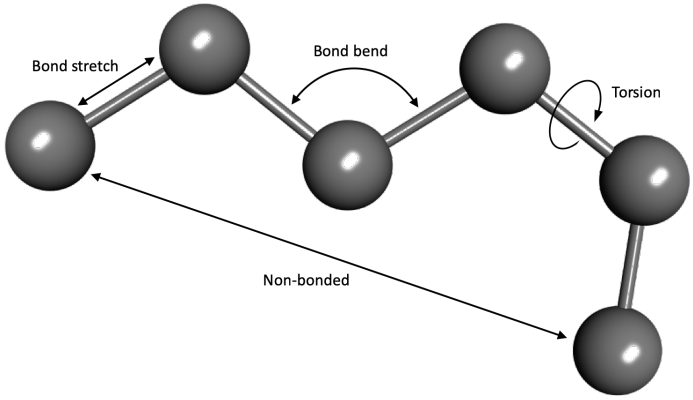


Figure 1.3 Representation of the main energy terms parametrised within a force field (based on Jensen, 2007, figure 2.1)[18]

1.2.1 Bonded parameters

Bond stretching and angle can both be controlled by a simple harmonic potential between either two (1,2 stretch), three (1,3 angle) covalently bonded atoms. At the reference bond length or angle the energy is at its lowest, by compressing, stretching or bending beyond these set parameters results in a rapid increase in energy. The individual terms for bond stretch, bond angle and dihedral can be seen below (Eq.3 and 4). In the case of the dihedrals, the torsional rotation around the bond means the function to describe it must be periodic as the energy should return to the same value. Therefore, a Fourier series with an appropriate number of periodic terms is used (Eq.5).

$$E = \sum_{\text{bonds}} K_r (r - r_{eq})^2 \qquad \text{(Eq.3)}$$

$$E = \sum_{\text{angles}} K_\theta (\theta - \theta_{eq})^2 \qquad \text{(Eq.4)}$$

$$E = \sum_{\text{dihedrals}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] \qquad \text{(Eq.5)}$$

1.2.2 Non-bonded parameters

The non-bonded terms described in the force field are split into two main types, van der Waals and electrostatic interactions.

Van der Waals interactions model the short range attractive and repulsive forces between non-bonded atoms. As the distance between atoms grows shorter the van der Waals energy increases greatly due to the repulsive nature of the atomic orbitals. As the distance between atoms increases an induced dipole-dipole interaction creates an attractive force between them. This attractive force changes as function of the inverse sixth power of the distance between the atoms. Therefore, at longer distances the van der Waals energy goes to zero.

The Lennard-Jones potential(L-J)[20] is the most common function found in force fields to calculate the van der Waals energy due to is relative simplicity in requiring on two parameters (ε and σ);

$$V_{LJ} = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] = \varepsilon\left[\left(\frac{r_{min}}{r_0}\right)^{12} - 2\left(\frac{r_{min}}{r_0}\right)^{6}\right] \qquad \text{(Eq.6)}^{18}$$

ε is the depth of the potential energy well, r is the distance between atoms, σ is the distance where the interatomic potential energy is zero and r_0 is the distance at which the inter atomic distance is at its minimum (Figure 1.4).

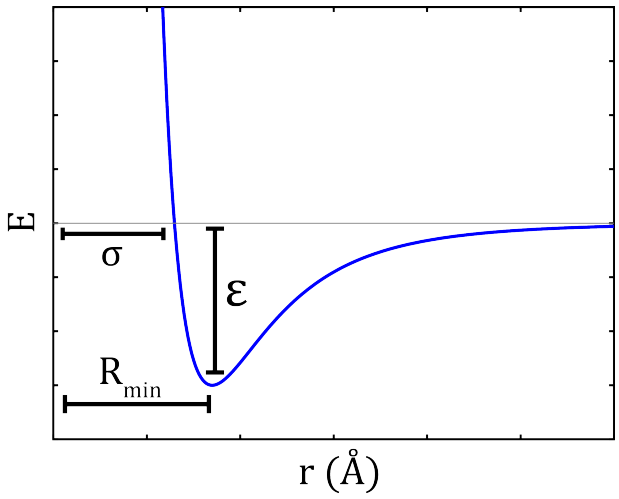


Figure 1.4 Lennard-Jones plot

The r^6 and r^{12} terms correspond to the attractive and repulsive components of the components respectively. The r^6 relating to the long term attractive component of the L-J potential can be justified theoretically as it relates to the induced dipole moment when two atoms are in close proximity, known as the van der Waals dispersion interaction, and obeys the scaling law; $\propto r^{-6}$ [19]. Whereas, r^{12} relating to the short range repulsive component, due to overlapping of electron orbitals, is purely a simplification used to increase the efficiency of the computational chemistry calculation for larger systems[19]. Rather than calculating the true repulsion term, such as that used in the Buckingham[21] or Morse[22] potential, the r^{12} term can be derived by calculating the square of the r^6 term. In large biomolecular systems containing hundreds of thousands of atoms and above the increased calculation speed obtained from the Lennard-Jones potential outweighs the need for a more accurate repulsive term.

Electrostatic interactions between pairs of atoms obey Coulomb's Law:

$$F = k_e \frac{q_i \times q_j}{r^2} \qquad \text{(Eq.7)}^{18}$$

where q_i and q_j are the magnitude of the charges of each atom, r is the distance between the charges and k_e is Coulomb's constant (9×10^9 N m^2 C^{-2}), as the distance between atoms increase the force between atoms decreases. Calculating the Coulombic interaction between each atom pair in a large simulation is computationally expensive,

especially so for simulations using periodic boundary conditions. The most common solution is to use the Ewald summation to calculate the electrostatic interactions.
The Ewald summation calculates the short-range and long-range contributions separately, the short-range contribution is calculated as a point charge on each atom, whereas the long-range contribution is calculated using a Fourier transform.

$$\mathcal{U}q = k_e \frac{q_i \times q_j}{r^2} \qquad \text{(Eq.8)}$$

The complete equation for the Amber force field can be seen below (Eq.9), combining all the above parameters.

$$E_{\text{total}} = \sum_{\text{bonds}} K_r(r - r_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}}\right] \qquad \text{(Eq.9)}$$

1.2.3 The Verlet Algorithm

The above terms allow us to calculate both bonded and non-bonded interactions and the energies involved. Finally, Newton's equations of motion need to be integrated. The simplest numerical integration scheme is a Taylor series expansion (Eq10) for solving the equations of motion for any given body (from an atom to a planet the only difference is a suitable timestep to describe the motion).

$$f(x) = a_0 + a_1(x - c) + a_2(x - c)^2 + a_3(x - c)^3 + \cdots a_n(x - c)^n \qquad \text{(Eq.10)}$$

If we apply this to Newtons second law, we get:

$$r_i(t + \Delta t) = r_i(t) + \Delta t v_i(t) + \frac{\Delta t^2}{2m_i} F_i(t) \qquad \text{(Eq.11)}^{23}$$

where r_i is the position of the given particle, t is time, v is velocity, m is mass and $F_i(t)$ is the short form of the Taylor expansion, $F_i(r_1(t),.....,r_N(t))$.
This can also be written for $r_i(t-\Delta t)$ to give a velocity-independent term:

$$r_i(t-\Delta t) = r_i(t) - \Delta t v_i(t) + \frac{\Delta t^2}{2m_i} F_i(t) \qquad \text{(Eq.12)}^{23}$$

and when equations 11 and 12 are combined and rearranged we get the Verlet algorithm (Eq.13)

$$r_i(t+\Delta t) = 2r_i(t) - r_i(t-\Delta t) + \frac{\Delta t^2}{m_i} F_i(t) \qquad \text{(Eq.13)}^{23}$$

The Verlet algorithm allows for the generation of a trajectory based on initial coordinates and velocities of a body over a given appropriate timestep, in the case of atomistic simulations the timestep is usually in femtoseconds. As the Verlet algorithm only calculates position, a second equation (Eq.14) is needed to generate the velocities. This can be done from any point in the trajectory.

$$v(t) = \frac{x(t+\Delta t) - x(t-\Delta t)}{2\Delta t} \qquad \text{(Eq.14)}^{23}$$

Whilst the Verlet algorithm is computationally simplistic it can be erroneous in its velocity approximations with the error increasing with larger timesteps. Two adaptations have been devised to reduce this discrepancy; the Leap Frog algorithm and the Velocity Verlet algorithm.
With the Leap Frog algorithm force is calculated first based on the new position, then velocity is calculated half step ahead based on position and force before calculating the final position and the process starting again for the next time step. This helps reduce the error in approximating the velocity.
In Velocity Verlet the next position is calculated based on velocity and force, then velocity is calculated a half step based on the previous force. Force is then calculated based on the new position and the velocity is advance a further half step to bring position, velocity and fore to the same timestep. In this way the velocity is not approximated and is calculated from only known values of position and force.

1.3 Statistical ensembles

The "measurement" conditions of any molecular dynamics system can be defined using a statistical thermodynamic ensemble[24]. A statistical ensemble describes the probability distribution of the state of a system based on the average of certain conserved variables over infinite time. These variable properties, such as pressure, energy, volume or temperature will average and therefore provide values for the thermodynamic quantity of interest. Each ensemble can be used based on the properties of the system under investigation and so are analogues to the reaction conditions of a chemical process[18] (Figure 1.5).
The most commonly used ensembles are;

Microcanonical ensemble or *NVE* ensemble:

In the microcanonical[24] ensemble the system is completely isolated with no transfer of energy between it and its surroundings. The total number of particles (*N*), total volume (*V*) and total energy (*E*) are constant. Temperature cannot be defined as this requires an interaction with its surroundings

Canonical ensemble or *NVT* ensemble:

In the canonical[24] ensemble the system energy can be transferred between the system and its surroundings. The total number of particles (*N*), volume (*V*), and temperature (*T*) remain constant. The system can be thought to be immersed in a large heat bath at a set temperature and a high enough heat capacity to remain stable no matter the energy transfer from the system to its surroundings. In this way the total energy of the system can be calculated.

Grand canonical ensemble or *μVT* ensemble:

In the grand canonical[24] ensemble both energy and particles can be transferred between system and its surroundings or a second system. The volume (*V*) of the system remains constant. Again, a heat bath controls the temperature (*T*) and final

thermodynamic variable is chemical potential (μ) can be used to calculate the rate of a chemical reaction or phase transition through changes in the particle numbers within the system.

Isothermal-isobaric ensemble or *NpT* ensemble:

In the isothermal-isobaric[25] ensemble, like the canonical and grand canonical ensembles, energy can be transferred between the system and its surroundings via a heat band to control the temperature(T). Also, the volume of the system can change, with the pressure (p) of the system being controlled by a barostat, so as to match the pressure exerted by its surroundings. The number of particles (N) remain constant within the system. As the system maintains a constant temperature and pressure the Isothermal-isobaric ensemble or *NpT* ensemble can be used to for free energy calculations of chemical reaction or integration as the Gibbs free energy describes the maximum amount of work undertaken within a system under the conditions of constant pressure and temperature.

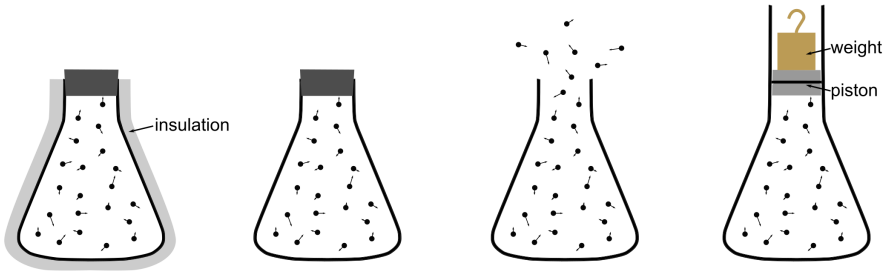


Figure 1.5 Graphical representation of statistical thermodynamic ensembles.

1.4 Free-energy Methods

Molecular dynamics (MD) propagate positions and velocities of interacting atoms in finite timesteps to incrementally build a relevant trajectory, based on a statistical-mechanical ensemble, for the system under investigation. Ideally, each trajectory would be ergodic, visiting all points on the potential energy surface (PES) and thus cover every possibility within the system in question. As MD trajectories may fail to explore all the relevant regions of configuration space due to the presence large energy barriers separating low energy metastable regions[11], only a small section of the overall configuration space is sampled by unbiased MD methods.

To overcome such barriers and visit important configuration regions, enhanced sampling methods have been proposed to help exploring the whole system. Since in this way the correct weight of configurations can be obtained, the free energy can be evaluated. The free energy of a system can help ascertain the chemical processes involved and the probability of a system's configuration or conformation at a given time[26].

By increasing the likelihood of sampling, the rare event under investigation, methods such as umbrella sampling and metadynamics can explore relevant areas of configuration space, which can provide greater understanding at an atomistic level to the processes involved. Umbrella sampling and metadynamics were methods of choice in this work. Umbrella sampling is an established method of calculating the free-energy of a biological system based on trajectory configurations, which are restrained by an external harmonic potential. The second method, metadynamics, is a state of the art free-energy method which, whilst not yet an established method for use with large bio-systems, has proved useful in fragment-based docking studies[27]. With the potential to elucidate the mechanistic nature of our systems, we feel metadynamics is the ideal free-energy method to explore further. One advantage of metadynamics is the presence of a dynamic trajectory, which is not present in Umbrella sampling.

1.4.1 Umbrella Sampling

Umbrella sampling, developed by Torrie and Valleau[28], is a method to evaluate the free energy barrier between two states. Therein, the reaction coordinated is restrained within a narrow set of values with a biasing potential. The measure that differentiates the two states, such as a ligand interacting with a protein active site or the folded and unfolded states of a protein during activation, defines the reaction coordinate. Umbrella sampling uses a bias potential to counter the free energy along the reaction coordinate pathway and in effect flatten the free energy surface[28]. One single bias potential is unlikely to provide an accurate picture of the overall energy landscape along the reaction coordinate; therefore, it is advantageous to combine the results of a number of overlapping trajectory slices, each with individual bias potentials, in order to reconstruct an accurate picture of the free energy surface. The reaction coordinate can be split into overlapping windows, each accounting for a section of the pathway along the reaction coordinate (Figure 1.6).

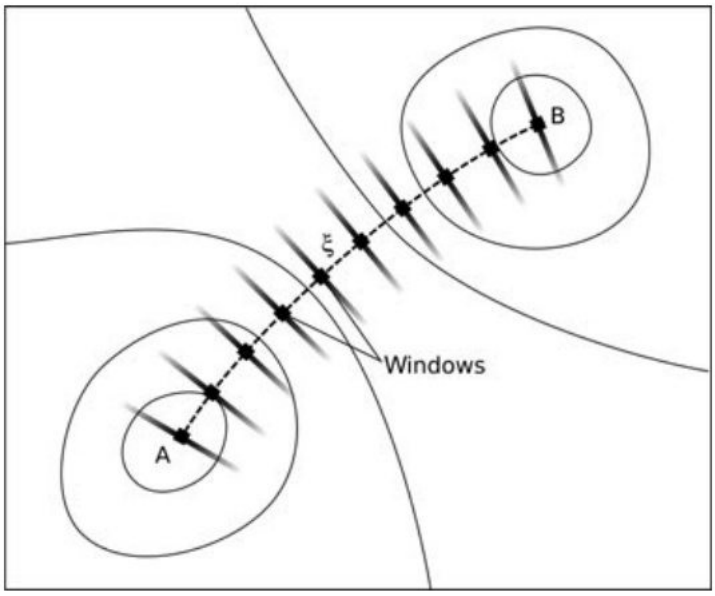


Figure 1.6 Schematic of how the reaction coordinate between minimas A and B is split into individual windows.

Bias potentials are simple harmonic springs used to restrain the trajectory to reference point within the window. The bias potential can either pull the system slowly along

the reaction coordinate within the window or maintain it a stationary position, allowing for conformational changes without directional progression. In both cases the system is allowed to explore the free energy within a set, restricted phase space. Due to the trajectory slicing, there is no dynamical information in the sequence of configurations that can be collected.

The statistics for each window can then be combined and analysed using an unbiasing technique such as weighted histogram analysis method (WHAM) or umbrella integration.

Weighted histogram analysis method (WHAM)

Weighted histogram analysis method (WHAM) is based on the work of Ferrenberg and Swendsen[29] on Multiple Histogram equations and further developed by Kumar *et al*[30] for use with biomolecule free energy calculations. This involves calculating the relative states of interest from a number of individual windows along a reaction coordinate, such as those calculated by umbrella sampling. From these states of interest histograms of the most probable molecular states can be obtained. The histogram probabilities from each umbrella window can then be combined to produce a potential of mean force (PMF) of the energies relating to the reaction coordinate pathway.

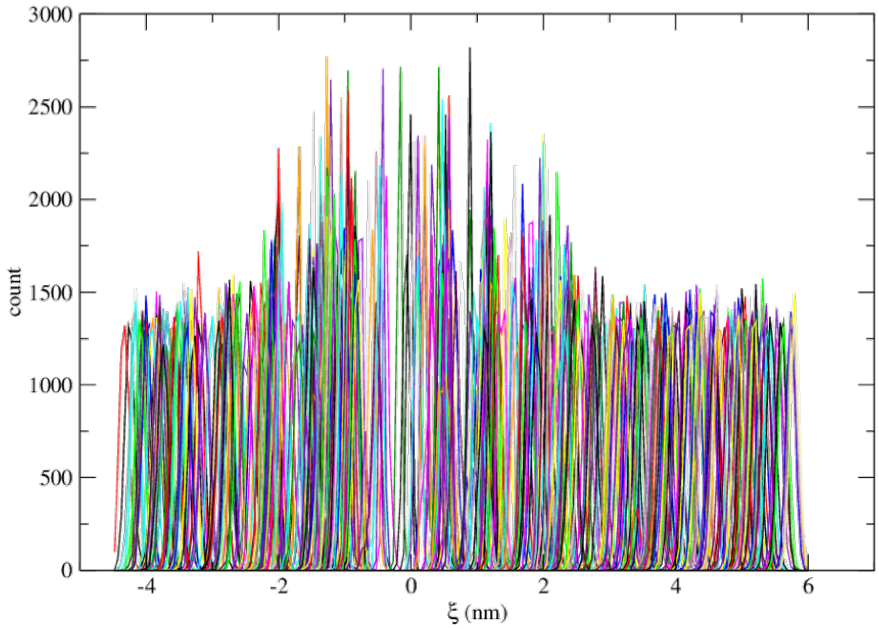


Figure 1.7 Histogram of each of 500 windows overplayed, showing the overlap of each.

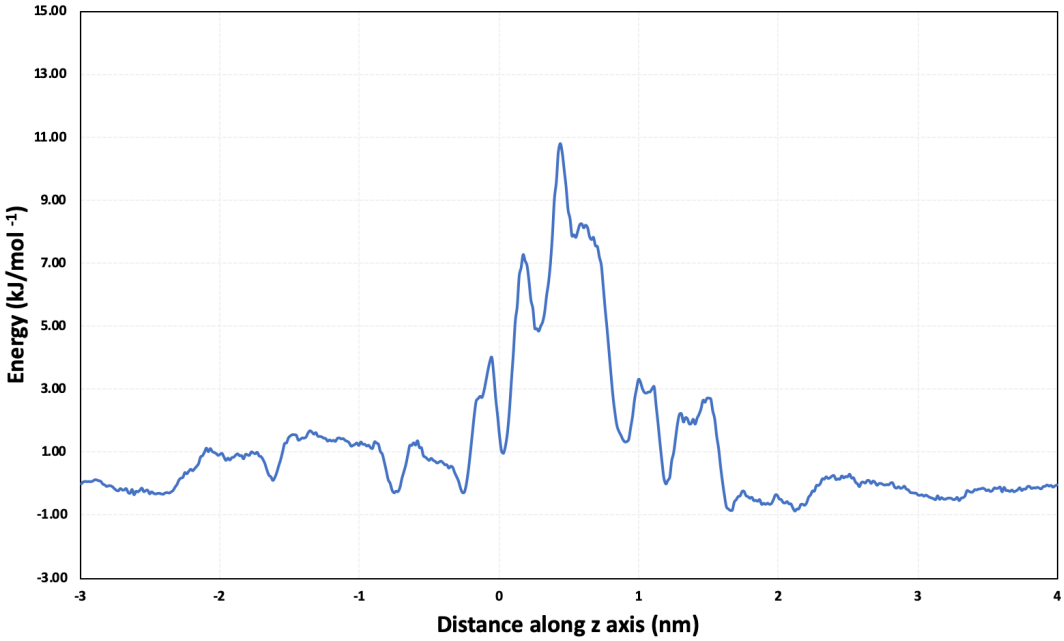


Figure 1.8 Example of a resolved potential of mean force (PMF) from an umbrella sampling calculation of H_2O_2 permeation of AQP3.

1.4.2 Metadynamics

Metadynamics is a technique originally developed by Laio and Parrinello[10] to efficiently explore the free energy surface (FES) of a system based on a number of selected reaction coordinates, so-called collective variables (CVs). CVs are able to describe the key aspect of the system under investigation, such as how the distance between two groups within the simulation changes and the effect such movement has on the FES. A single CV can be used or a combination thereof, which is capable of describing the dynamics of the system in a way than can then be interpreted by the user. A history-dependent potential is included so as to discourage the system from revisiting a previously described state. In doing so, metadynamics can push the system to explore the whole FES, overcome energy barriers and find more stable conformations or metastable conformations, which remains elusive if other methods are used (Figure 1.7).

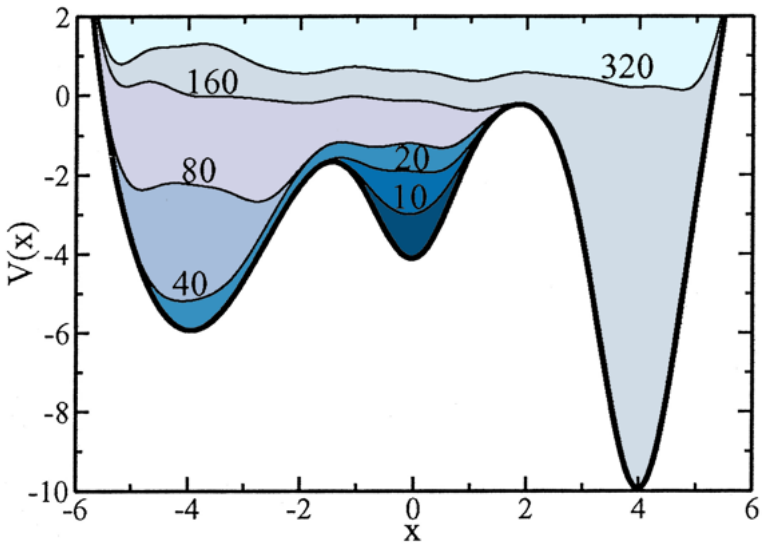


Figure 1.7. One-dimensional free energy surface (FES) showing three metastable energy wells and the gaussian deposition filling the wells and overcoming the energy barriers. (V(x) equating to the free energy and x dependant on the collective variable (CV used).(Adapted from Fig. 1, Laio and Parrinello, 2002[10]).

Metadynamics works by depositing Gaussian functions onto the FES, analogous to filling a mould with sand where you can control the size of grain and pour rate. A fast deposition rate with a large Gaussian will allow a quick exploration of the FES but will fail to differentiate and explore states smaller than the gaussian width. This could result in missing possibly crucial states due to its coarse nature, whereas filling with a small gaussian may not allow escaping from a single FE well within the simulation time. A balance needs to be found based on the system and the reaction/interaction under investigation.

The choice and number of CVs is dependent on the system under investigation. Therefore, there are no set rules to state which CV will be best for any given calculation. Instead, the CV choice should satisfy three ideal properties[31];

- Clearly distinguish between initial state, any intermediates and the final state
- Describe all the relevant slow events for the process of interest
- The number of CV's used should not be too large, so as to be detrimental to the time taken to fill the FES

Examples of collective variables are:

- Geometric CV's – distances between atoms or molecules, angles, torsion angles
- Coordination number
- Potential energy of the system – changes in entropy and enthalpy
- Path – such as the pathway between two energy minima of the FES

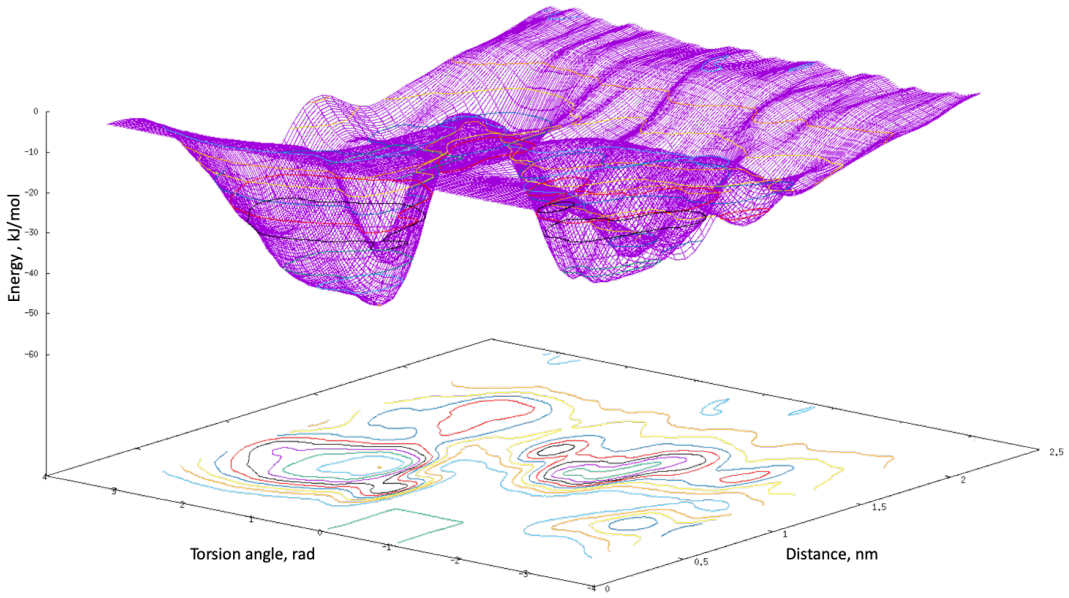


Figure 1.8. Example of a two-dimensional free energy surface with centre of mass (COM distance and torsion angle as the collective variables (CVs).

By following these principles and using the simplest CV or collection of CVs possible generally affords the best results. A certain degree of trial and error is necessary to find the correct combination of CV's, Gaussian size and deposition rate to fully explore the system under investigation and allow an accurate interpretation of the FES. As the calculation is run, a history dependent bias potential is constructed based on the chosen collective variables, by the deposition of Gaussian functions, the summation of which is described as the metadynamics potential V

$$V(s,t) = \sum_{k\tau<t} W(k\tau) exp\left(-\sum_{i=1}^{d} \frac{(s_i - s_i^{(0)}(k\tau))^2}{2{\sigma_i}^2}\right) \qquad \text{(Eq.15)}^{10}$$

Where τ is the deposition rate in MD steps, W is the Gaussian height and σ is the Gaussian width. As previously stated, the choice of these parameters will affect how the system is biased in conjunction with the chosen CV's, and again a certain element of trial and error will be needed to determine the most appropriate values.

One issue with metadynamics is achieving convergence and knowing when to end the calculation. By continually depositing Gaussians of equal size the free energy can be overestimated and the system may fail to converge fully but fluctuates around the final value with an averaged error value proportional to the square root of the bias

potential deposition rate[32]. One way of solving this fluctuating error is to use an adaptive bias, i.e. rescaling the Gaussian height 'on the fly' based on the previously deposited gaussians and temperature known as well-tempered metadynamics.

$$\mathcal{W} = \omega_0 \tau_G e^{-\frac{\Delta V(\theta,t)}{k_B \Delta T}} \qquad \text{(Eq.16)}^{11}$$

ω_0 is the starting height of the gaussian, ΔT is the term that controls how far away the trajectory biasing deviates from the free-energy minima based on a temperature different controlled by a biasing factor. Effectively well-tempered metadynamics samples the CV at a higher temperature $T + \Delta T$ yet reconstructs the free-energy surface corresponding the target temperature.

Whilst exploring the entirety of the system is advantageous and sometimes essential to fully understand its dynamics, in some cases this is unnecessary such as when the binding site is well established. In this case the history dependent deposition results in only exploring the active site briefly before exploring the rest of the system and, due to the discouragement of revisiting minima, provides only a single binding/unbinding event in a potentially microsecond simulation. For this reason, a development of metadynamics specifically to address this issue was devised, Funnel Metadynamics[33]. In Funnel Metadynamics a funnel shaped constraint is applied around the active site to dissuade the ligand from exploring outside of the area of interest. The funnel can be adapted using simple geometric terms such as R_{cyl} for the radius of the cylinder forming the pathway for the ligand and Z_{cc} and α to determine the width and funnel angle to full incorporate the protein active site (Figure 1.9). The size and shape of the funnel must be chosen so as not to affect the binding site with an external bias.[33]

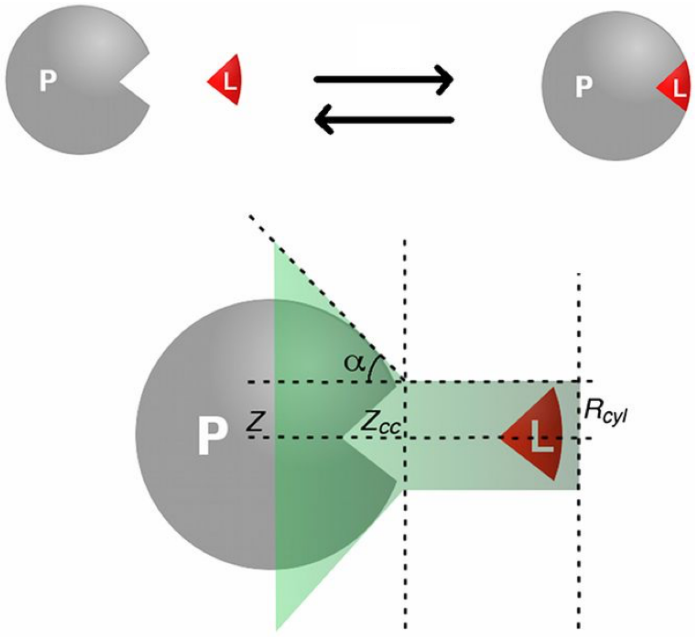


Figure 1.9. Schematic showing how the funnel incorporates the active site of the protein. (Adapted from Fig.1, Limongelli et al (2013)[33]).

By effectively reducing the sampling space with the funnel constraint, the ligand/protein binding/unbinding mechanism can be calculated multiple times within a single simulation allowing for more accurate and faster convergence of the system.

Section 2

2.1 Computational methods and drug discovery

With the cost of developing a new therapeutic drug estimated at between $400 million and $2 billion[34] and a large proportion of this expenditure spent on the synthesis and testing of lead analogues[35], a method to more efficiently determine and optimise main compounds could reduce the time and costs substantially.

Computer aided drug design (CADD) has, over the past 30 years, evolved into a major tool in the development of new therapeutic small molecules. The use of three-dimensional protein and DNA structures, derived from crystallography and NMR, allows for virtual high throughput screening (vHTS) for hit to lead and lead optimisation. An example of the potential CADD has for efficiently identifying future therapeutics is the case of tyrosine phosphatase-1B inhibitors; virtual screening identified 127 compounds showing effective inhibition out of an initial screening of 365, an approximate hit rate of 35%. Whereas traditional high-throughput screening identified 81 compounds out of a library of 400,000, a hit rate of 0.021%[35].

In silico methods are usually classified as structure-based or ligand-based. The former relies on knowledge of the target structure, such as a protein active site, and the interactions within it. The latter is based on the knowledge of the activity of small molecules and other chemically similar compounds to construct quantitative structure-activity relation (QSAR) models and are generally used when little structural knowledge of the target is known.

In both cases the main aim is to develop a compound, which binds tightly within the active site and shows a quantifiable reduction in free-energy. Therefore, in silico methods that can reliably calculate the free-energy of binding and interaction are essential. The methods discussed previously (umbrella sampling and metadynamics) are two such computational techniques, which offer advantages in terms of computationally efficiency and reliability.

Virtual screening libraries are built up of drug-like ligands with certain desirable physiochemical properties for the target of interest. A common 'drug-likeness' evaluation is Lipinski's rule of five[36], which states an orally active drug should not violate more than one of the criteria: 1) maximum of five hydrogen bond donors, 2) no more than ten oxygen or nitrogen atoms, 3) a molecular mass of less than 500 Da and 5) an octanol-water partition coefficient no greater than five. These libraries are used in docking studies where molecular docking algorithms are used to rapidly screen and score the ligands based on certain criteria relating to the free-energy of each ligand within the target site such as; solvent effects, conformational changes, interactions and geometry[37]. Unfortunately, due to the nature of high-throughput virtual screening, in docking the target is usually static and based on a crystal structure, which represents only a snapshot pose of the active site and does not take into account the flexibility of the protein and possible conformational changes induced on activation as well as failing to take into account solvent entropy contributions[38].
For the screening of large numbers of compounds docking studies provide the most efficient method to identify potential lead compounds but due to the issues previously mentioned the development of more accurate active docking where the flexibility of the target can be taken into account.

The next stage consists in taking the identified compounds and, through structure activity relationship studies, design a highly selective inhibitor for the given target. This requires further *in silico* techniques, based around free-energy calculation methods, to further enhance the knowledge base of drug design. Such methods allow for the flexibility of active sites and, with the correct choice of collective variables, provide a more complete enumeration of a biological systems conformational states than can be obtained by docking alone.

2.2 Metals in Medicine

The used of coordination and organometallic metal complexes for therapy is one of the emerging areas of drug design. Since the discovery of the Pt(II) complex cisplatin (Figure. 2.1) by Rosenberg in 1965[39] for the treatment of cancer, a number of Pt(II)

and Pt(IV) derivatives have been further developed, including the clinically used drugs carboplatin[40] and oxaliplatin[41]. The postulated mechanism of action involves direct DNA damaging *via* formation of coordination bonds between the Pt(II) ions and the DNA nucleobases. Thus, cisplatin is a pro-drug which, following hydrolysis of the chloride ligands inside the cells, can further exchange the water/hydroxide ligands with nitrogen atoms in nucleobases, causing DNA structural lesions eventually leading to cellular apoptosis (Figure 2.1).

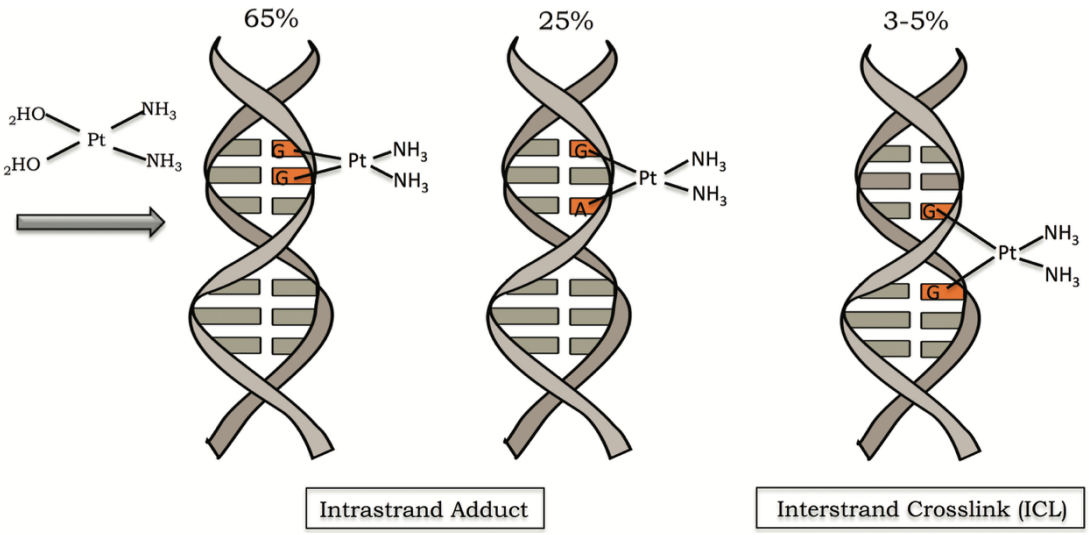


Figure 2.1. Cisplatin binding modes with DNA, intrastrand DNA adducts and interstrand crosslinking with the frequency of each. (Adapted from Figure 1, Rocha et al (2018)[42]).

Following the clinical success of cisplatin and analogues, several other experimental anticancer metallodrugs have been developed, based on ruthenium, osmium, iridium, palladium and gold. Among them, the gold(I)-based compound auranofin[43] (Figure. 2.2) was approved for the treatment of rheumatoid arthritis in 1985 and it is now in clinical trials for the treatment of different cancers alone or in combination therapy. These different metal-based drugs can exploit novel mechanisms of action with respect to cisplatin and other classical organic compounds, and some of them have been reported to target proteins instead of nucleic acids targets[44].

On the other hand, another key area of drug discovery is the use of small organic ligands featuring a metal-binding pharmacophore (MBP) capable of forming strong interactions with metals within the catalytic site of metalloenzymes and metalloproteins. As an example, matrix metalloproteinases (MMPs)[45,46] contain a

catalytically important Zn^{2+} centre within the active site and coordinating to a conserved HExxHxxGxxH binding motif[45]. Due to their role in a number of pathological conditions, MMPs are ideal drug candidates, and several compounds have been developed so far with the aim to selectively inhibit their activity also for cancer treatment. As an example, ABT-518 (Figure. 2.2) was found to be both a potent, orally bioavailable inhibitor with good selectivity for both MMP-2 and MMP-9 over MMP-1[47]. ABT-518 underwent Phase I clinical trials achieving 48% inhibition of B16 melanoma tumour growth in several animal models[47].

Similarly, the magnesium dependent HIV integrase (HIV-IN) features an active site containing highly conserved catalytic motif coordinating to Mg^{2+} divalent metal ions[48]. The metal ions are essential for the catalytic activity of HIV integrase, by stabilising the negative charge during phosphate hydrolysis, during its role in integrating viral DNA into the host DNA. By sequestering these metal ions using organic chelators such as Raltegravir[48] (Figure. 2.2), it is possible to cause displacement of viral 3'-OH of HIV-1 intergrase and prevent integration of the viral DNA.

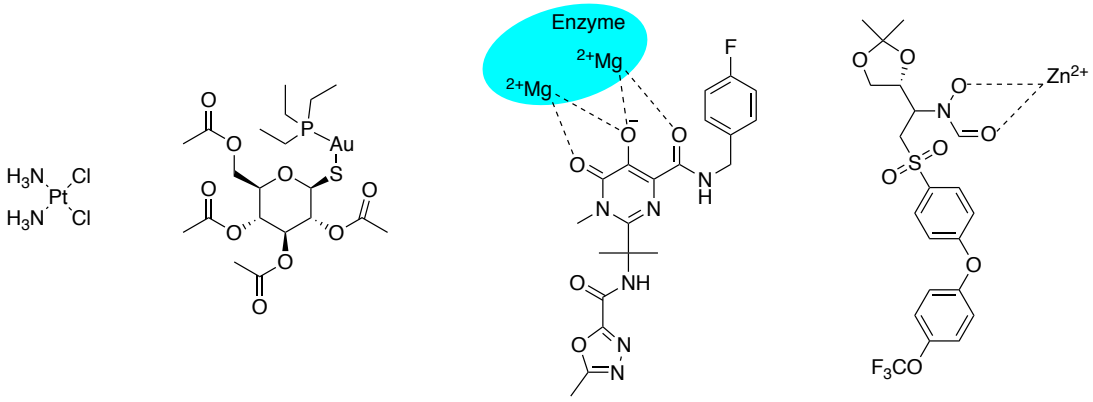


Figure 2.2 Example of metallodrugs and metalloenzyme inhibitors.

In this context, computational methods such as docking, and quantum mechanical/molecular mechanics (QM/MM) and classical molecular dynamics (MD) simulations have played a part in the development of both metallodrugs and metalloenzyme inhibitors. Molecular dynamics and QM/MM studies go beyond what is possible with classical rigid docking studies by allowing for the dynamics and

plasticity of an active site to be explored, allowing for the calculation of thermodynamic and kinetic data along with physical pathways of binding/unbinding to be obtained, and thus, providing a better overall understanding of the mechanisms involved in target/ligand interactions.

Both force field based, and QM/MM studies can be used to estimate ligand-target compatibility of libraries of drug like compounds in virtual screening studies. An example, the successful identification of broad-spectrum thiol inhibitors of metallo-β-lactamase *via* chelation to one or more zinc centres was achieved [49] (Figure 2.3).

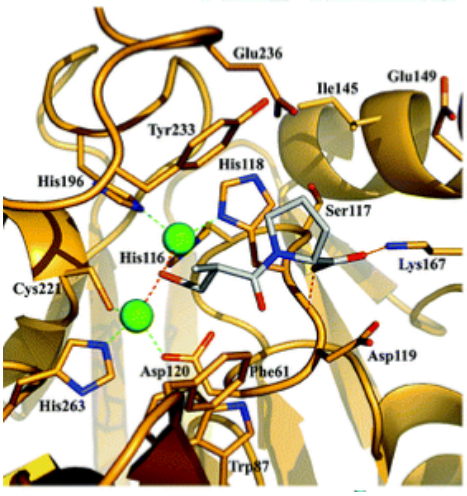


Figure 2.3 Interactions of D-captopril chelator to two zinc ions (green spheres) of metallo-β-lactamase BlaB (pdb ID 1M2X)[49].

Some issues arise from such calculations in that force field-based techniques fail to take into account the electron density of the metals in either the enzymes catalytic site, in the case of metalloenzyme inhibitors or the metal centre with organometallic and coordination compounds. The use of QM based methods can overcome some of the limitations by allowing step by step changes in the partial charges, and variable protonation states to be calculated via *ab initio* methods rather than the fixed charges used in forcefields, allowing for more accurate electrostatic potentials to be implemented. However, as previously stated, QM methods come with an increase in computational power and a reduction in system size, and so they are not suited to high-throughput virtual screening and are more appropriate for lead generation and optimisation.

Therefore, the integration of different computational approaches is often necessary to provide a realistic and informative picture of the system under investigation. An example of this strategy includes the recent work by Graziani el al[50], whereby a multi-level theoretical approach was used to determine the binding mode of Au(III) compound $[Au(bipy)Cl_2]^+$ (bipy = bipyridine, Aubipy, Figure. 2.4) with the human water/glycerol membrane channel AQP3[50]. The compound was previously shown to be a selective and potent inhibitor of this protein isoform[51]. Initially, the study used DFT to determine the free-energy of Aubipy aquation and therefore, the most probable gold-based species likely to react with the protein in physiological conditions. Classical molecular dynamics simulations were then performed on three systems: **A,** where Aubipy was not present, **B**, where Aubipy was placed to non-covalently interact with the protein extracellular pocket and **C**, where Aubipy was covalently bound to the thiolate sidechain of Cys40[50] (the putative Au(III) binding site). To determine the effects of Aubipy, the water flux of each system was measured and compared along with channel radius to determine if Aubipy induced conformational changes within the pore. The results show the pore size of systems **B** and **C** were reduced when compared to the unbound protein (system**)**, accounting for the inhibition of glycerol permeation by Aubipy. Water flux was found to be retarded more in system B (non-covalently bonded) than in system C (covalently bonded) with figures of 0.151 x 10^{-14} cm^3s^{-1} and 1.81 x 10^{-14} cm^3s^{-1} respectively, both notably lower than the value for system A of 4.51 x 10^{-14} cm^3s^{-1}.

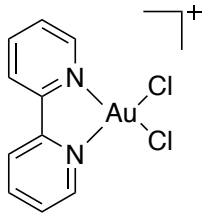


Figure 2.4 Structure of the Au(III) compound $[Au(bipy)Cl_2]^+$ (Aubipy), inhibitor of human AQP3[50].

Finally, the geometry of Aubipy in the non-bonded and bonded (systems **B** and **C**) was calculated using QM/MM (Figure. 2.5). Only the residues and water molecules within a c.a. 4 Å of the metal centre were included in the calculations, which equated to the compound, 61 amino acid residues and 8-11 water molecules. Geometries were compared to the classical MD simulations and found to be in good agreement with bond length distances typically within less than 2% and angles within 5%.

By using this multi-level approach, a better understanding of how Aubipy inhibits AQP3 water flux could be ascertained at a molecular level, with each theory level focusing on different yet complementary aspects of the system.

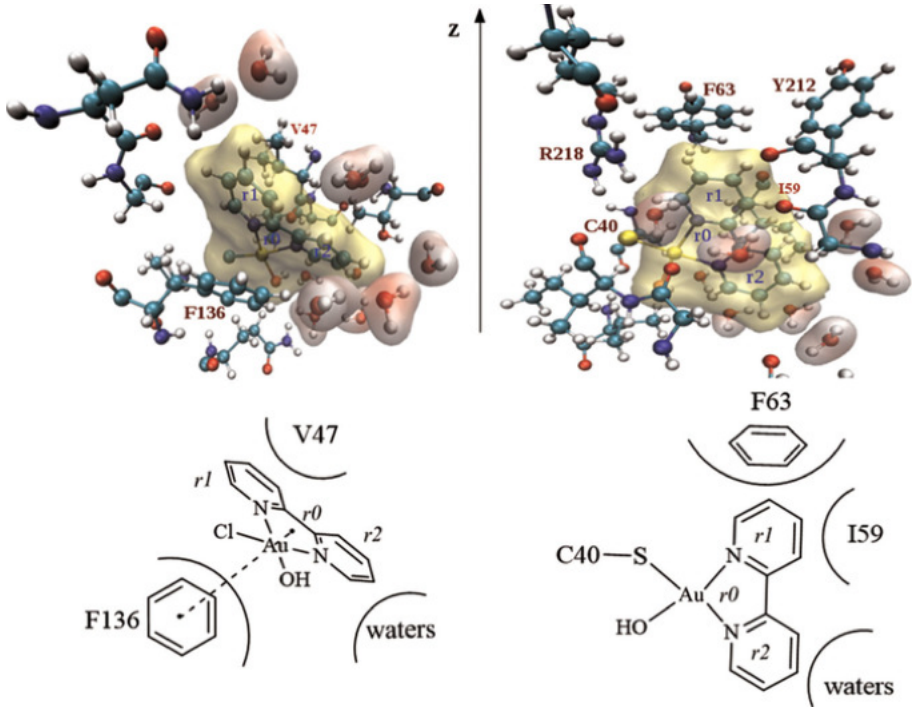


Figure 2.5 Position of Aubipy (shown with a yellow surface) in **a)** the non-covalently bound (system **B**), and **b)** the covalently bound (system **C**) models. 2D schematic is shown below with indication of the stacking interactions[50].

2.3 References

1 J. R. Perilla, B. C. Goh, C. K. Cassidy, B. Liu, R. C. Bernardi, T. Rudack, H. Yu, Z. Wu and K. Schulten, *Curr. Opin. Struct. Biol.*, 2015, **31**, 64–74.

2 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindah, *SoftwareX*, 2015, **1**–**2**, 19–25.

3 D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668–1688.

4 K. Bowers, E. Chow, H. Xu, R. Dror, M. Eastwood, B. Gregersen, J. Klepeis, I. Kolossváry, M. Moraes, F. Sacerdoti, J. Salmon, Y. Shan and D. Shaw, ACM, New York, NY, USA, 2006, p. 746.

5 H. Kitano, *Nature*, 2002, **420**, 206–210.

6 M. Levy, *Proc. Natl. Acad. Sci.*, 1979, **76**, 6062–6065.

7 R. Friedman, K. Boye and K. Flatmark, *Biochim. Biophys. Acta - Rev. Cancer*, 2013, **1836**, 1–14.

8 E. H. Lee, J. Hsin, M. Sotomayor, G. Comellas and K. Schulten, *Structure*, 2009, **17**, 1295–1306.

9 J. Kästner, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2011, **1**, 932–942.

10 A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci.*, 2002, **99**, 12562 LP – 12566.

11 C. Abrams and G. Bussi, *Entropy*, 2014, **16**, 163–199.

12 J. S. Hub and B. L. de Groot, *Proc. Natl. Acad. Sci.*, 2008, **105**, 1198–1203.

13 F. Moraca, J. Amato, F. Ortuso, A. Artese, B. Pagano, E. Novellino, S. Alcaro, M. Parrinello and V. Limongelli, *Proc. Natl. Acad. Sci.*, 2017, **114**, E2136–E2145.

14 D. M. Nikolaev, A. A. Shtyrov, M. S. Panov, A. Jamal, O. B. Chakchir, V. A. Kochemirovsky, M. Olivucci and M. N. Ryazantsev, *ACS Omega*, 2018, **3**, 7555–7566.

15 S. Kelm, J. Shi and C. M. Deane, *Bioinformatics*, 2010, **26**, 2833–2840.

16 D. Xu, J. Zhang, A. Roy and Y. Zhang, *Proteins Struct. Funct. Bioinforma.*, 2011, **79**, 147–160.

17 A. Leaver-Fay, M. Tyka, S. M. Lewis, O. F. Lange, J. Thompson, R. Jacak, K. W. Kaufman, P. D. Renfrew, C. A. Smith, W. Sheffler, I. W. Davis, S. Cooper,

A. Treuille, D. J. Mandell, F. Richter, Y.-E. A. Ban, S. J. Fleishman, J. E. Corn, D. E. Kim, S. Lyskov, M. Berrondo, S. Mentzer, Z. Popović, J. J. Havranek, J. Karanicolas, R. Das, J. Meiler, T. Kortemme, J. J. Gray, B. Kuhlman, D. Baker and P. Bradley, in *Computer Methods, Part C*, eds. M. L. Johnson and L. B. T.-M. in E. Brand, Academic Press, 2011, vol. 487, pp. 545–574.

18 F. Jensen, *Introduction to Computational Chemistry*, John Wiley & Sons Ltd, Chichester, 2nd edn., 2007.

19 A. H. C. Horn, *Essentials of Computational Chemistry, Theories and Models*, 2003, vol. 43.

20 J.E. Jones, 1924, **4**.

21 R. A. Buckingham, *Proc. R. Soc. A Math. Phys. Eng. Sci.*, 1938, **168**, 264–283.

22 P. M. Morse, *Phys. Rev.*, 1929, **34**, 57–64.

23 Mark E. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation*, Oxford University Press Inc, New York, NY, USA, 2010.

24 J. W. Gibbs, *Elementary Principles in Statisitical Mechanics*, Charles Scribner's Sons, New York, NY, USA, 1902.

25 I. R. McDonald, *Mol. Phys.*, 1972, **23**, 41–58.

26 C. D. Christ, A. E. Mark and W. F. van Gunsteren, *J. Comput. Chem.*, 2010, **31**, 1569–1582.

27 D. Gioia, M. Bertazzo, M. Recanatini, M. Masetti and A. Cavalli, *Molecules*, 2017, **22**, 1–21.

28 G. M. Torrie and J. P. Valleau, *Non-Physical Sampling Distributions in Monte-Carlo Free-Energy Estimation - Umbrella Sampling*, 1977, vol. 23.

29 A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.*, 1988, **61**, 2635–2638.

30 S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 1011–1021.

31 A. Laio and F. L. Gervasio, *Reports Prog. Phys.*, 2008, **71**.

32 A. Barducci, G. Bussi and M. Parrinello, *Phys. Rev. Lett.*, 2008, **100**, 1–4.

33 V. Limongelli, M. Bonomi and M. Parrinello, *Proc. Natl. Acad. Sci.*, 2013, **110**, 6358–6363.

34 S. C. Basak, *Drug Dev. Res.*, 2011, **72**, 225–233.

35 G. Sliwoski, S. Kothiwale, J. Meiler and E. W. Lowe Jr, *Pharmacol. Rev.*, 2014, **66**, 334–395.

36 C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Adv. Drug Deliv.*

Rev., 2012, **64**, 4–17.

37 L. Riccardi, V. Genna and M. De Vivo, *Nat. Rev. Chem.*, 2018, **2**, 100–112.

38 R. D. Malmstrom and S. J. Watowich, *J. Chem. Inf. Model.*, 2011, **51**, 1648–1655.

39 B. Rosenberg, L. Van Camp and T. Krigas, *Nature*, 1965, **205**, 698–699.

40 R. Canetta, M. Rozencweig and S. K. Carter, *Cancer Treat. Rev.*, 1985, **12**, 125–136.

41 D. Townsend, *xPharm Compr. Pharmacol. Ref.*, 2011, **3**, 1–4.

42 C. Rocha, M. Silva, A. Quinet, J. Cabral-Neto and C. Menck, *Clinics*, 2018, **73**, 1–10.

43 C. K. Mirabelli, R. K. Johnson, D. T. Hill, L. F. Faucette, G. R. Girard, G. Y. Kuo, C. M. Sung and S. T. Crooke, *J. Med. Chem.*, 1986, **29**, 218.

44 A. de Almeida, B. L. Oliveira, J. D. G. Correia, G. Soveral and A. Casini, *Coord. Chem. Rev.*, 2013, **257**, 2689–2704.

45 A. Di Pizio, M. Agamennone and M. Aschi, *PLoS One*, 2012, **7**.

46 P. Bencsik, K. Kupai, A. Görbe, É. Kenyeres, Z. V. Varga, J. Pálóczi, R. Gáspár, L. Kovács, L. Weber, F. Takács, I. Hajdú, G. Fabó, S. Cseh, L. Barna, T. Csont, C. Csonka, G. Dormán and P. Ferdinandy, *Front. Pharmacol.*, 2018, **9**, 1–14.

47 C. Wada, *Curr. Top. Med. Chem.*, 2005, **4**, 1255–1267.

48 Z. Hajimahdi and A. Zarghi, *Iran. J. Pharm. Res.*, 2016, **15**, 595–628.

49 B. M. R. Liénard, G. Garau, L. Horsfall, A. I. Karsisiotis, C. Damblon, P. Lassaux, C. Papamicael, G. C. K. Roberts, M. Galleni, O. Dideberg, J. M. Frère and C. J. Schofield, *Org. Biomol. Chem.*, 2008, **6**, 2282–2294.

50 V. Graziani, A. Marrone, N. Re, C. Coletti, J. A. Platts and A. Casini, *Chem. - A Eur. J.*, 2017, **23**, 13802–13813.

51 A. P. Martins, A. Marrone, A. Ciancetta, A. G. Cobo, M. Echevarría, T. F. Moura, N. Re, A. Casini and G. Soveral, *PLoS One*, 2012, **7**.

Chapter 2 - Aims

This project will cover two topics, the first on the mechanism of aquaporin permeation and effects on inhibition of metal compounds (Chapters 3 to 6). The second topic covers the development of G-quadruplex stabilisers and their potential as therapeutics (Chapters 7 and 8). A combination of both established steered molecular dynamics and state of the art metadynamics methods will be used to elucidate the free energies and mechanisms of action for the systems.

The first part of this project is split into two aims;

The first aim consists in elucidating the mechanism of aquaporin permeation of different substrates (water and glycerol) through the use of molecular dynamics approaches. Aquaporin homology models will be used of human AQP3 and AQP7. With AQP3, the permeation by H_2O and glycerol will be studied to determine the exact molecular mechanism of permeation, helping the design of highly specific aquaporin inhibitors. With detailed information on the aquaporin permeation, it will be possible to design new metal-based inhibitors which would inhibit only one aquaporin isoform.

With AQP7, the effect of pH on water and glycerol permeation will be studied via both experimental and in silico methods. Changes in pH from neutral to acidic have been found to induce a gating mechanism, molecular dynamic methods will be used to ascertain the possible mechanism of such gating.

The second aim of the project is to shed light into the detailed mechanisms of AQPs inhibition by metal compounds. Thus, the most promising metal-based compound (e.g. Au(III)) will be bound to the aquaporin channel to evaluate their inhibitory mode of action, using the same techniques described above. A main challenge of this phase of the project is the identification of parameters for metal complexes allowing to mimic and define the coordination bonds in silico.

Overall, both project's aims converge to help the design of potent and selective AQPs inhibitors that could be used as therapeutic agents (e.g. inhibiting the activity of cancer related AQPs), but also as chemical probes, to evaluate the role each protein isoform in physiology and patho-physiology.

The second part of this project is to help develop highly selective G4 stabilisers using a combination of both in silico and experimental techniques. We hope to use state of the art computational methods, such as metadynamics, to calculate the free energy of binding of organometallic ligands with a number of both promotor and telomeric G-quadruplex structures, such as hTelo and C-KIT1, allowing the further development of potent, highly specific G4 stabilisers. This will also require the development of force field parameters for the organometallic ligands.

Chapter 3 - Aquaporins and molecular dynamics

3.1 Aquaporin structure and function

The controlled movement of water both between and within cells is of great physiological importance. A highly specific family of transmembrane spanning proteins called aquaporins controls this process[1]. The human family of aquaporins consists of thirteen isoforms (AQP0 – AQP12), organized into two distinct groups: the orthodox aquaporins (AQP0, AQP1, AQP2, AQP4, AQP5, AQP6 and AQP8), which are permeated mainly by water; and the aquaglyceroporins (AQP3, AQP7, AQP9, AQP10 and AQP11), which are also permeated by glycerol and other small molecules, such as urea and H_2O_2[1].

Aquaporins are tetrameric protein complexes constructed from four discrete monomers, with van der Waals forces maintaining the 3-dimensional structure[2]. Each monomer consists of six helices (H1-H6) joined together by five loops (A-E), as shown in Figure 3.1. Two smaller helices (HB and HE), located in loops B and E (Figure 3.1 A)[3], are close to each other at the centre of the channel to form one of two selectivity filters, asparagine-alanine-proline motif (NPA motif), found in all aquaporins. The second selectivity filter consists of a conserved arginine residue and two (aquaglyceroporins) or three (aquaporins) other amino acid residues named the aromatic/arginine (ar/R SF) (Figure 3.1 B), which are always aromatic and may include a cysteine residue in some isoforms[4].

The ar/R SF constricts the pore size to c.a. 1-2 Å[5], creating a steric barrier and preventing large molecules passing through the pore (Figure 3.1 B), while the NPA motif created by helices HB and HE, in conjunction with the half hydrophobic/hydrophilic internal pore surface, creates an electrostatic barrier, preventing the passage of charged species through the pore[6].

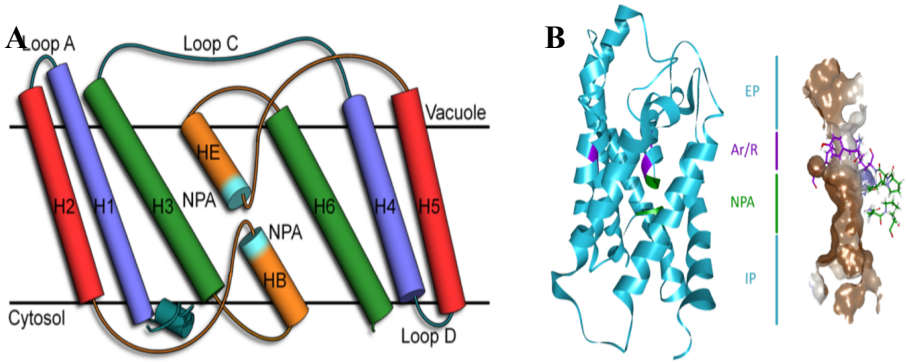


Figure 3.1 Aquaporin monomer structure. A. Schematic representation of a generic AQP secondary structure and the relative position of each of the elements in the lipid bilayer[3]. B. AQP3 monomer in ribbon representation (left) and surface of the channel (right). Residues Phe63, Tyr212 and Arg218 (purple) constitute the ar/R S/F with Asn83 and Asn215 (green) form the NPA S/F.

3.2 Aquaporins and cancer

The facilitation of water transport is the key aquaporin function. Therefore, AQPs are abundantly expressed in tissue types involved in maintaining fluid balance such as the kidneys[7] and airways[7,8] along with the central nervous system[9]. Due to their abundance, aquaporins allow cells to react efficiently to changes in osmotic gradient created by the active transport of ions and other neutral solutes[1]. A reduction in the ability to transport water can have a detrimental effect on these tissues, as it has been shown in mice with reduced ability to concentrate urine (AQP1)[10], defective secretion of saliva (AQP5)[11] and reduced skin elasticity (AQP3)[12].

Glycerol is a key molecule in physiology and has a role in skin hydration, ATP generation and lipid synthesis[1]. For example, maintaining glycerol levels within the stratum corneum helps preventing water evaporation by acting as a humectant[12]. Glycerol can be combined with long chain fatty acids to form triglycerides, for storage in adipose tissue, while the triglyceride (TAG) is phosphorylated and incorporated into the cell membrane during cell proliferation[13,14]. Additionally, glycerol is also a precursor to Acetyl-CoA, used in the tricarboxylic acid (TCA) cycle to produce ATP[14,15] (Figure 3.2 A). Thus, aquaglyceroporin expression is particularly high in cell types such as adipocytes and liver due to their role in glycerol metabolism.

The overexpression of orthodox aquaporins in tumour cells has been linked to cell invasion and metastasis[1]. Whilst AQP1, AQP4 and AQP5 are known to be involved in cell migration[14], overexpression of these isoforms in tumour cell lines has been shown to increase their ability to invade adjacent tissues, both *in vitro* and *in vivo*[16]. One hypothesis for this observation is that increased levels of aquaporins allow for rapid volume changes and rapid water flow, in and out of the cells, thus, allowing for rapid cell shape changes resulting in increased cell motility[16].
Recently, immunostaining of suspected malignant thyroid nodules for AQP4 upregulation has been patented as a method of distinguishing between benign and malignant oncocytic lesions, due to the known overexpression of this isoform in certain cancerous tissues[17].

One aquaporin isoform of particular interest is aquaglyceroporin-3 (AQP3), which has been found to be overexpressed in a number of tumour cell types including skin cancer and breast cancer[18]. AQP3 is known to transport both glycerol and H_2O_2 along with water, even though its water permeability is lower than the one of other aquaporin isoforms. It is therefore hypothesised that glycerol may play a role in tumour growth and proliferation[14]. As with healthy cells, there are two processes for which cancer cells require glycerol uptake; lipid synthesis and the production of ATP.
Glycerol is initially converted to triacylglycerol (TAG) and stored in adipose tissue[19]. When required, glycerol can be released from the adipose tissue and transported into cells to undergo phosphorylation to glycerol-3-phosphate (G-3-P), a precursor capable of being converted to both glucose *via* gluconeogenesis and free fatty acid (FFA) *via* lipogenesis (Figure 3.2). Through the tricarboxylic acid cycle (TCA) glucose can be converted into adenosine triphosphate (ATP) the chemical energy source for cellular processes. In the case of lipogenesis, the FFA units can convert into phospholipids, the building blocks of cell membranes. These processes are required by both healthy cells (Figure 3.2A) and tumour cells (Figure 3.2B), with the transport of glycerol between cells being facilitated by a number of aquaglyceroporin isoforms such as AQP7 in the case of adipocytes[19] and AQP9 in liver cells[20]. Due to the increased rate of cancer cell proliferation, these processes are accelerated and therefore, require the upregulation of AQPs to aid necessary solute permeation.

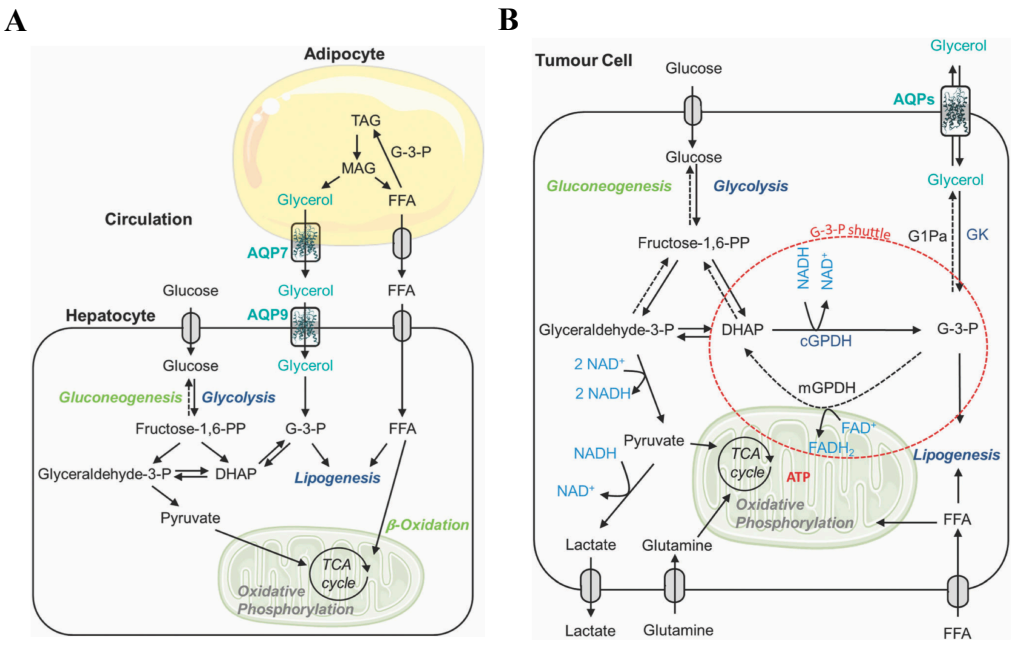


Figure 3.2 AQP3 glycerol transport function and metabolism. **A** Glycerol metabolism in liver and adipose tissue **B** Glycerol metabolism in cell proliferation and cancer development (Abbreviations: ATP = adenine triphosphate, DHAP = dehydroxy- acetone phosphate, FAD^+ = flavine adenine dinucleotide, $FADH_2$ = hydroquinone form of flavine adenine dinucleotide, FFA = free fatty acid, G-3-P = glycerol 3-phosphate, cGPDH and mGDPH = cytosolic and mitochondrial, respectively, glycerol-3-phosphate dehydrogenase, G1Pa = glucose 1-phosphate-adenyltransferase, MAG = monoacylglycerol, NAD^+ = oxidized nicotinamide adenine nucleotide, NADH = reduced nicotinamide adenine nucleotide, TAG = triacylglycerol, TCA = tricarboxylic acid)[14].

Hydrogen peroxide, also an AQP3 permeant, is a reactive oxygen species (ROS), mostly produced in mitochondria. It can be involved in cellular damage, but it is also used as a signalling molecule for different cellular processes[21]. Extracellular H_2O_2, produced by NADPH oxidase-2 (Nox2), is believed to be taken up *via* AQP3, and in cancer cells it has been shown to induce a signalling cascade that leads to increased cell migration and metastasis formation[18] (Figure 3.3).

Whilst ROS are an important part in cell homeostasis, overproduction can lead to DNA damage and oxidation of lipids and amino acids and eventually apoptosis[22]. The protein kinase AKT induces cell replication and metastasis. This process is controlled by phosphatase and tensin homolog (PTEN), which, can be deactivated *via* the oxidation of specific cystine residues by H_2O_2. This allows the uncontrolled proliferation of cells[23].

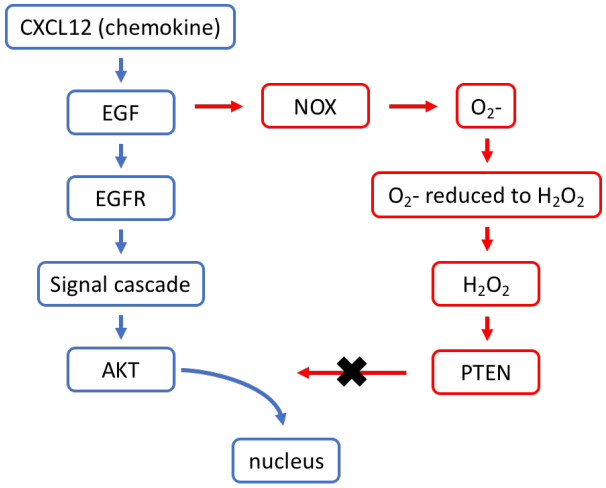


Figure 3.3 Possible pathway for H_2O_2 induced secondary signaling pathway in cancer cells. (Abbreviations: AKT = protein kinase, EGF = epidermal growth factor, EGFR = epidermal growth factor receptor, NOX = nicotinamide adenine dinucleotide phosphate oxidases, PTEN = phosphatase and tensin homolog).

Overall, both glycerol and H_2O_2 appear to be important players in cancer development. Therefore, by selectively inhibiting their aquaglyceroporin permeation, tumour proliferation and migration could be reduced. Thus, the development of selective aquaporin inhibitors is of great interest for therapeutic purposes, as well as to understand the roles of AQPs in physiological and pathophysiological processes. In order to develop highly selective and effective inhibitors, the mechanism of small molecule permeation through the pore must be well understood. The use of *in silico* techniques is, therefore, very important to elucidate the molecular mechanism and to complement experimental evidences obtained by *in vitro* and *in vivo* studies. In the next chapter we will provide an overview of the results obtained on metal compounds as AQPs inhibitors.

3.3 AQP3 inhibition by gold compounds

Inhibition of AQPs' water permeation by metal-based compounds has been investigated for several years. Mercurial compounds such as $HgCl_2$, along with zinc and nickel salts, $AgNO_3$ and silver sulfadiazine have shown inhibitory properties[24]. Unfortunately, due to a combination of lack of selectivity and high toxicity, these compounds are all unsuitable for therapeutic applications. At a molecular level, some

information on the binding mechanisms of these compounds to AQPs comes from X-ray structural analysis,[25] as well as from site-directed mutagenesis studies[26]. In the case of mercury, it appears that metal binding to Cys residues is responsible for its inhibition effects. The affinity of mercury for sulfur donors can be explained by the hard and soft (Lewis) acid and base theory (HSAB theory)[27], whereby mercury acts as the acid and sulfur acts as the base. Whilst mercury has been shown to be an effective AQP3 inhibitor, its promiscuity and therefore, toxicity, prevents it from being developed into a highly selective inhibitor.

The idea of using HSAB theory of metal ions to selectively target certain available residues within a protein structure, such as the sulfur of Cys40 in AQP3, but with greater control of the metal compound's reactivity, has proven viable. Recently, our research group discovered gold coordination complexes as selective aquaglyceroporins inhibitors[24]. Specifically, the Au(III) coordination complex Auphen [Au(1,10-phenanthroline) Cl_2]Cl (Figure 3.4.A) was found to cause highly selective and potent inhibition for human AQP3[28] and AQP7[29] *vs* the orthodox channel AQP1. Au(III) ions are relatively 'soft' Lewis acids and, similarly to mercury, can react with sulfur donor. Therefore, it was hypothesized that Auphen could bind to Cys residues in the protein channel, such as Cys40 in human AQP3, inducing inhibition of substrate permeation by steric hindrance. The Cys40 residue was first identified using homology modelling approaches and showed to be accessible for gold binding at the extracellular side of the pore (Figure 3.4. B). To further ascertain the accessibility and affinity of Auphen and its derivatives for Cys40, QM/MM calculations were performed to determine the most stable non-covalent binding modes of the compound inside the extracellular pocket[30]. The calculations showed that Auphen, along with similar Au(III) complexes with bidentate N-donor ligands, to have the best binding affinity with a ΔG value of -48 kJ mol^{-1}. In detail, Auphen was shown to interact, *via* π-π stacking, with residues Tyr212 and Phe63 of the ar/R selectivity filter and indicated the mechanism of how Au(III) compounds inhibit AQP3. Later on, this residue was validated as a binding site for Auphen using mutagenesis studies[31].
This study also showed the link between tumour progression and glycerol uptake and the antiproliferative effect of Auphen on A431 skin cancer cells, overexpressing the AQP3 isoform.

Finally, in 2019 our group, using a combined classical MD and QM/MM approach, has reported on the importance of non-coordinative adducts with the protein surface in modulating the AQP3 inhibition properties of a series of Au(III) coordination compounds[32]. The study concluded that these interactions, prior to the covalent protein/metal complex adduct formation, should be further considered in the future design of isoform-selective AQP inhibitors.

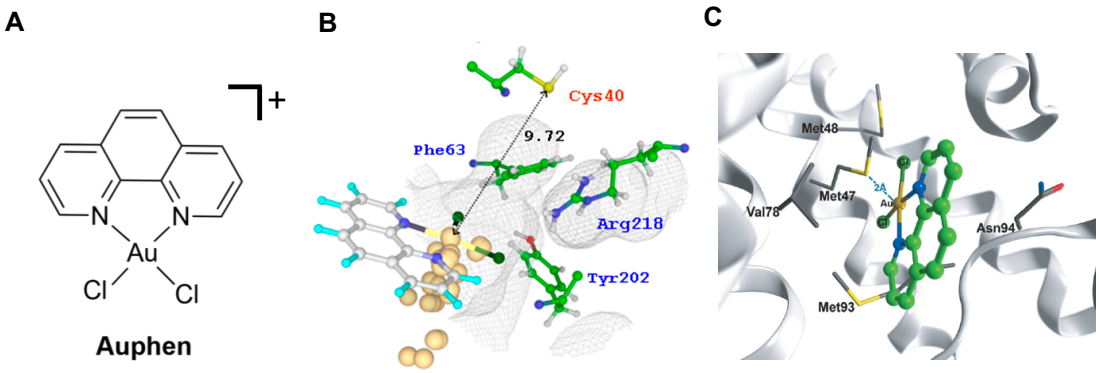


Figure 3.4 A. Au (III) coordination complex Auphen. **B.** Auphen docking with Cys40 of AQP3 and its position relative to the residues of the ar/R SF[24]. **C**. Auphen docked within the NPA region of AQP7, showing interactions between Au and Met47[29].

Further studies showed that Auphen can also inhibit glycerol permeation *via* AQP7[29]. This isoform, like AQP3, has been found to transport both water and glycerol along with other small uncharged solutes. AQP7 is found in abundance in the plasma membrane of adipocytes and is the primary glycerol transporter in adipose tissue[19]. Moreover, AQP7 can also be found in thyroid[33] and uterus[34] tissue and has been found to be overexpressed in malignant ovarian cancer cells when compared to heathy ovarian tissue[35].

Unlike AQP3, AQP7 does not have an exposed cystine residue on the extracellular side of the channel, therefore a different mechanism of inhibition was proposed. In this case, using a homology modelling of human AQP7 and non-covalent docking studies, three methionine residues were identified, below the NPA region on the intracellular side of the channel, as the possible gold binding sites (Figure 3.4. C)[31]. Two of the methionine's, namely Met47 and Met93, were identified as the likely

binding sites for Auphen, resulting in channel blockage. As these residues are found on the cytoplasmic side of the protein, Auphen must enter the cell before it can bind to the methionine residues. However, this mechanistic hypothesis needs further validation.

3.4 Molecular dynamics simulations of aquaporins

Molecular dynamics (MD) has been widely used to elucidate the potential energies of the ar/R and NPA selectivity filters concerning substrate permeation, mainly focusing on water and glycerol [36] along with the free energy calculations for a variety of substrates using umbrella sampling[37].

The most studied aquaporins using *in silico* techniques are human AQP1 and bacterial glycerol facilitator (GlpF), both with their structures already solved by X-ray crystallography[38]. These simulations provided the first insights into the mechanism of water permeation, along with the structural differences between the orthodox aquaporins and aquaglyceroporins that exclude or allow the passage of certain molecules, such as H_2O, glycerol and urea. These studies revealed the highly conserved selectivity filters ar/R and NPA and the roles each plays in aquaporins selectivity, as well as calculating the free energy of permeation for H_2O, glycerol NH_3 and urea[37]. One study using umbrella sampling to calculate the potential of mean force (PMF) of conductance for a number of small solutes *via* AQP1 and GlpF, showed that although they differ in size, both H_2O and glycerol were found to have equivalent ΔG values of 13.5 kJ/mol[37]. In this case the ΔG value for water was calculated based on water density and not by umbrella sampling as was the case for glycerol.

As crystal structure data is available, a number of molecular dynamics studies have been carried out on GlpF[36,37,39,40], hAQP1[36,37], and hAQP4[41,42]. In this context, de Groot and Grubmüller[36] identified the ar/R and NPA selectivity filters and the water dipole flip (Figure 3.5). This now established motif can be seen in all the available studies on aquaporins, with the interactions within the NPA region helping to maintain

the single file water transport mechanism as well as prevent the passage of charged species.

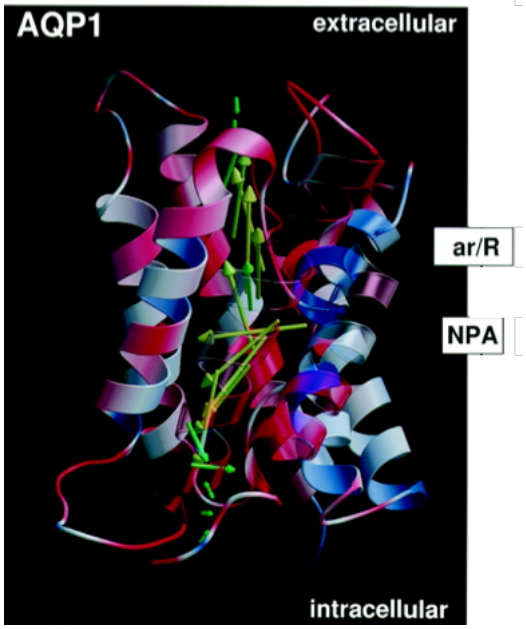


Figure 3.5 Water dipole moment orientation within AQP1 pore as calculated by de Groot and Grubmüller[36]. Arrows' head indicates positive end of dipole. (Adapted from de Groot and Grubmüller (2001)[36].

Historically, the lack of structural information on human aquaglyceroporins has fostered the use of *in silico* techniques, such as homology modelling, MD and DFT to ascertain the possible binding modes of potential inhibitor/protein adducts as described above. However, in 2018 Gotfryd et al resolved the structure of hAQP10 *via* X-ray diffraction[43], the first time a human aquaglyceroporin has been determined. This structure will certainly lead to more accurate modelling in future aquaglyceroporin studies.

An earlier study on glycerol permeation *via* the bacterial glycerol facilitator (GlpF) protein by Jensen et al[39] used Jarzynski's identity to calculate the irreversible work done by a pull code to encourage glycerol to pass though the GlpF pores to calculate the PMF. Whilst this method and umbrella sampling provide two-dimensional energy profiles of the pathway of solutes through the aquaporin pores, they fail to provide any physiologically relevant trajectory information on the events. Each method uses the combination of a number of calculations, whereby the solute is restrained, to provide

the free energy plots. As the interactions of solutes with the protein and the transition between bulk water and entering the pore are highly complex, trajectory information is essential to understand the mechanism of permeation. Although no dynamic information is available from these data, they do provide highly relevant information on the nature of the pore residue/solute interactions with regards to the two selectivity filters. Both show an energy increase and the ar/R SF and NPA region within the pore[39].

Further work by Jensen *et al* in two papers in 2001[39] and 2002,[40] studied first the mechanism of spontaneous conductance of glycerol *via* GlpF, and secondly the energetics involved in the formation of hydrogen bonds between glycerol, water and GlpF pore residues during conductance. The results showed how orientation of glycerol affects the energy of translocation through the pore, and thus an optimal orientation is more energetically favourable (Fig 3.6)[39]. A third key area within the pore was also identified featuring a highly conserved histidine (His66 in the case of GlpF) residue below the NPA motif, towards the cytoplasmic pore entrance. Whilst not thought to be involved in solute selectivity, this area is involved in hydrogen bond network and therefore plays a role in the free energy landscape of solute conductance.

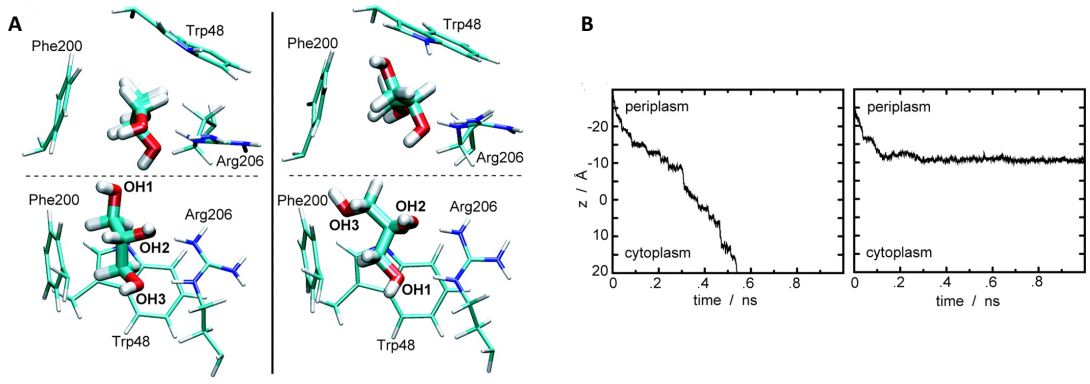


Figure 3.6 A, The favourable (left) and unfavourable (right) orientations of glycerol as they pass through the ar/R SF of GlpF. **B**, Position within the pore (along the z-axis vs time for the favourable (left) and unfavourable (right) glycerol orientations, showing how in the unfavourable orientation glycerol fails pass through the pore.(Adapted from Figure 3, Jensen el al[40]).

In 2008, Hub and de Groot showed how energy barriers of solute uptake differ between orthodox aquaporins (AQP1) and aquaglyceroporins (GlpF)[37]. This work also

provided evidence on the energy expenditure difference of solutes diffusing through a lipid bilayer and traversing an aquaporin pore. Steered molecular dynamics (SMD) in the form of umbrella sampling with the trajectory of each solute initiated by a pull code was used to determine the energy barriers of conductance. The obtained results show that it is more efficient for small apolar molecules such as O_2 and CO_2 to diffuse through a lipid bilayer, whilst with more polar and hydrophilic molecules such as ammonia, water and glycerol, the energy barrier is far greater and thus discourages diffusion[37]. The free energy barrier for these molecules are greatly reduced when passing through the aquaporin pores (Table 3.1). Furthermore, the main finding of the paper was the role that water molecules-pore interactions play in pore selectivity and free energy of solute uptake, rather than the interactions between solutes and the pore itself. The study analysed the hydrogen bond interactions of O_2 during permeation of hAQP1 and GlpF. O_2 was found in reduce water-protein interactions within the ar/R region of hAQP1 during permeation and data showed both a reduction in H-bonding and thus, a reduction in energy of c.a.60kJ/mol. This was not the case for O_2 permeation of GlpF which showed little to no reduction in water-protein interaction in the ar/R region during conductance. The difference in the results is thought to be due to the increased diameter and hydrophobic nature of the GlpF pore in comparison to hAQP1, thus, reducing the selectivity of GlpF for small polar solutes.

Table3.1 Summary of free energy barriers ΔG_{max} in kJ/mol, (values taken from Table 1 Hub and de Groot (2008)[37]. (* values calculated from water density)

Solute	hAQP1	GlpF	hAQP1-HA/RV	POPE	POPC
O_2	27	12.5	8.5	6	4
CO_2	22	13.5	5	4	1.5
NH_3	18	12.5	9	19	15
H_2O	14*	13.5*	13*	31	27
Glycerol	(24)	13.5	13.5	30	27
Urea	32.5	29	16	32	34

The pioneering work carried out in these studies has allowed a far greater understanding of the role played by each of the selectivity filters during conductance events as well as the importance of the single file water chain to the solute uptake. With the water mechanism as essential to the selectivity and movement of solutes through the pore, it is necessary to understand how solutes transition from bulk water to the single file water chain in the various aquaporins. Unfortunately, none of the studies above can answer this question due to the nature of the molecular dynamics

techniques used, as the classical MD does not allow for an unbiased conductance trajectory to be calculated at a physiologically relevant timescale.

Recently, molecular dynamics simulations provided evidence on the mechanism of inhibition of Hg^{2+} of both water and glycerol permeation in human AQP3[44]. This study showed that Hg^{2+} induces conformational changes within the aquaporin pore by coordination to the residues within the ar/R SF and a number of water molecules present in the pore. Hg^{2+} shows an affinity for the sulfur of residue Cys40, located close to the ar/R SF, forming an octahedral coordination to Cys40 in each of the four monomers along with the oxygen of Ile146 and between two and four water molecules present in the pore (Figure 3.4).[44] By inducing protein conformational changes, and not by sterically hindering the AQP3 channel as originally postulated, mercury prevents both water and glycerol flux in three of the four pores of the AQP3 tetramer.

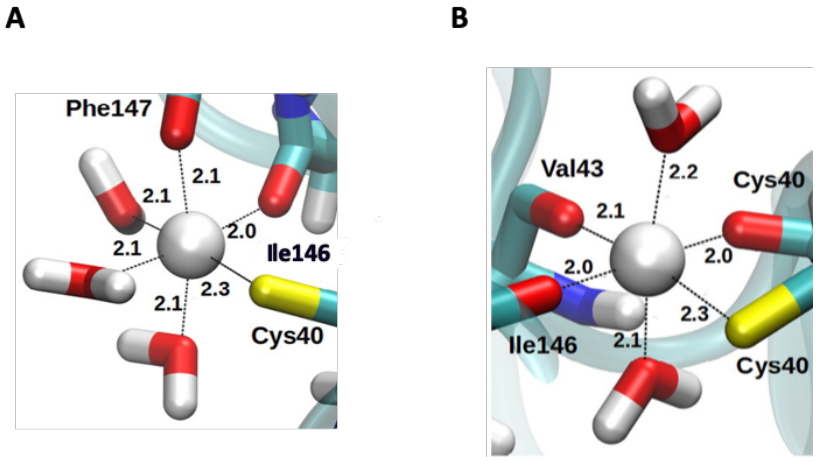


Figure 3.4. Hg^{2+} octahedral coordination within the AQP3 pores which is responsible for inhibition of glycerol permeation: In all cases, metal ion coordinates to Cys40 and other residues in the protein's pore as well as to water molecules. Adapted from Spinello et al 2016[44].

Overall, based on these promising results, by using state of the art techniques, we hope to begin to answer the question of how solutes such as glycerol permeate through AQP3 and what the determinants of AQPs selectivity for a certain substrate are. Moreover, we aim at elucidating the mechanisms of AQPs inhibition by metal complexes by MD methods, to achieve structure-activity relationships and improve the design of selective inhibitors.

3.5 References

1 A. S. Verkman, M. O. Anderson and M. C. Papadopoulos, *Nat. Rev. Drug Discov.*, 2014, 13, 259–277.

2 R. M. Cordeiro, *Biochim. Biophys. Acta - Gen. Subj.*, 2015, **1850**, 1786–1794.

3 A. Kirscht, S. S. Kaptan, G. P. Bienert, F. Chaumont, P. Nissen, B. L. de Groot, P. Kjellbom, P. Gourdon and U. Johanson, *PLoS Biol.*, 2016, **14**, 1–19.

4 A. Kirscht, Y. Sonntag, P. Kjellbom and U. Johanson, *BMC Struct. Biol.*, 2018, **18**.

5 L. Yu, R. A. Rodriguez, L. L. Chen, L. Y. Chen, G. Perry, S. F. McHardy and C. K. Yeh, *Protein Sci.*, 2016, **25**, 433–441.

6 A. Pelagalli, C. Squillacioti, N. Mirabella and R. Meli, *Int. J. Mol. Sci.*, 2016, **17**.

7 R. E. Day, P. Kitchen, D. S. Owen, C. Bland, L. Marshall, A. C. Conner, R. M. Bill and M. T. Conner, *Biochim. Biophys. Acta - Gen. Subj.*, 2014, **1840**, 1492–1506.

8 C. Bai, M. A. Matthay, A. S. Verkman and A. S. Verkman, 1999, **m**, 555–561.

9 R. Maugeri, G. Schiera, C. M. di Liegro, A. Fricano, D. G. Iacopino and I. Di Liegro, *Int. J. Mol. Sci.*, 2016, **17**, 1–18.

10 T. Ma, B. Yang, A. Gillespie, E. J. Carlson, C. J. Epstein and A. S. Verkman, *J. Biol. Chem.*, 1998, **273**, 4296–4299.

11 T. Ma, A. Gillespie, E. J. Carlson, Y. Song, A. S. Verkman and C. J. Epstein, *J. Biol. Chem.*, 2002, **274**, 20071–20074.

12 M. Tonghui, M. Hara, R. Sougrat, J. M. Verbavatz and A. S. Verkman, *J. Biol. Chem.*, 2002, **277**, 17147–17153.

13 C. GM, *The Cell: A Molecular Approach*, Sinauer Associates, Sunderland (MA), 2nd editio., 2000.

14 B. Aikman, A. De Almeida, S. M. Meier-Menches and A. Casini, *Metallomics*, 2018, **10**, 696–712.

15 R. A. Harvey and D. R. Ferrier, *Biochemistry*, Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia, 5th editio., 2011.

16 J. Hu and A. S. Verkman, *FASEB J.*, 2006, **20**, 1892–1894.

17 WO2013116537, 2013.

18 H. Satooka and M. Hara-Chikuma, *Mol. Cell. Biol.*, 2016, **36**, 1206–1218.

19 N. Maeda, *Mol. Aspects Med.*, 2012, **33**, 665–675.

20 C. Lindskog, A. Asplund, A. Catrina, S. Nielsen and M. Rützler, *J. Histochem. Cytochem.*, 2016, **64**, 287–300.

21 G. P. Bienert and F. Chaumont, *Biochim. Biophys. Acta - Gen. Subj.*, 2014, 1840, 1596–1604.

22 G. P. Bienert, A. L. B. Møller, K. A. Kristiansen, A. Schulz, I. M. Møller, J. K. Schjoerring and T. P. Jahn, *J. Biol. Chem.*, 2007, **282**, 1183–1192.

23 C. Polytarchou, M. Hatziapostolou and E. Papadimitriou, *J. Biol. Chem.*, 2005, **280**, 40428–40435.

24 A. P. Martins, A. Marrone, A. Ciancetta, A. G. Cobo, M. Echevarría, T. F. Moura, N. Re, A. Casini and G. Soveral, *PLoS One*, 2012, **7**.

25 D. F. Savage and R. M. Stroud, *J. Mol. Biol.*, 2007, **368**, 607–617.

26 M. Zelenina, A. A. Bondar, S. Zelenin and A. Aperia, *J. Biol. Chem.*, 2003, **278**, 30037–30043.

27 R. G. Pearson, *J. Am. Chem. Soc.*, 1963, **85**, 3533–3539.

28 A. Serna, A. Galán-Cobo, C. Rodrigues, I. Sánchez-Gomar, J. J. Toledo-Aral, T. F. Moura, A. Casini, G. Soveral and M. Echevarría, *J. Cell. Physiol.*, 2014, 229, 1787–1801.

29 A. Madeira, A. de Almeida, C. de Graaf, M. Camps, A. Zorzano, T. F. Moura, A. Casini and G. Soveral, *Chembiochem.*, 2014, 15, 1487–1494.

30 A. P. Martins, A. Ciancetta, A. de Almeida, A. Marrone, N. Re, G. Soveral and A. Casini, *ChemMedChem.*, 2013, 8, 1086–1092.

31 A. Serna, A. Galán-Cobo, C. Rodrigues, I. Sánchez-Gomar, J. J. Toledo-Aral, T. F. Moura, A. Casini, G. Soveral and M. Echevarría, *J. Cell. Physiol.*, 2014, **229**, 1787–1801.

32 M. N. Wenzel, A. F. Mósca, V. Graziani, B. Aikman, S. R. Thomas, A. de Almeida, J. A. Platts, N. Re, C. Coletti, A. Marrone, G. Soveral and A. Casini, *Inorg. Chem.*, 2019, **58**, 2140–2148.

33 L. Lacroix, V. Lazar, S. Michiels, H. Ripoche, P. Dessen, M. Talbot, B. Caillou, J. P. Levillain, M. Schlumberger and J. M. Bidart, *Am. J. Pathol.*, 2005, **167**, 223–231.

34 U. Alabalık, G. Türkcü, A. N. Keleş, İ. İbiloğlu, A. Özler, Z. Urakçı and H. Büyükbayram, *Biotechnol. Biotechnol. Equip.*, 2017, **31**, 148–155.

35 J. H. Yang, C. X. Yan, X. J. Chen and Y. S. Zhu, *J. Int. Med. Res.*, 2011, **39**,

702–711.

36 B. L. de Groot and H. Grubmuller, *Science (80-.).*, 2001, **294**, 2353–2357.

37 J. S. Hub and B. L. de Groot, *Proc. Natl. Acad. Sci.*, 2008, **105**, 1198–1203.

38 D. Fu, A. Libson, L. J. W. Miercke, C. Weitzman, P. Nollert, J. Krucinski and R. M. Stroud, *Science (80-.).*, 2000, **290**, 481–486.

39 M. Jensen, E. Tajkhorshid and K. Schulten, *Structure*, 2001, **9**, 1083–1093.

40 M. O. Jensen, S. Park, E. Tajkhorshid and K. Schulten, *Proc. Natl. Acad. Sci.*, 2002, **99**, 6731–6736.

41 N. J. English and J. A. Garate, *J. Chem. Phys.*, 2016, **145**.

42 M. Assentoft, S. Kaptan, H.-P. Schneider, J. W. Deitmer, B. L. de Groot and N. MacAulay, *J. Biol. Chem.*, 2016, **291**, 19184–19195.

43 K. Gotfryd, A. F. Mósca, J. W. Missel, S. F. Truelsen, K. Wang, M. Spulber, S. Krabbe, C. Hélix-Nielsen, U. Laforenza, G. Soveral, P. A. Pedersen and P. Gourdon, *Nat. Commun.*, 2018, **9**, 4749.

44 A. Spinello, A. De Almeida, A. Casini and G. Barone, *J. Inorg. Biochem.*, 2016, **160**, 78–84.

Chapter 4 - The mechanism of aquaporin inhibition by gold compounds elucidated by biophysical and computational methods

Based on the paper:

Andreia de Almeida, Andreia F. Mosca, Darren Wragg, Margot Wenzel, Paul Kavanagh, Giampaolo Barone, Stefano Leoni, Graça Soveral and Angela Casini. Chemical Communications, 2017, 53, 3830–3833.
DOI: 10.1039/c7cc00318h

Au (III) compounds synthesised and tested by Dr Margot Wenzel and Dr Andreia F. Mósca respectively.
AQP3 homology model constructed by Dr Andrea de Almeida.
DFT vibrational frequency calculations carried out my Dr Giampaolo Barone.

Abstract: The inhibition of glycerol and water conductance *via* human aquaglyceroporin-3 (AQP3) by Au(III) complexes has been studied by stopped-flow spectroscopy and, for the first time, its mechanism has been described using molecular dynamics (MD) and density functional theory (DFT). The obtained MD results showed the most effective gold-based inhibitor, bound to Cys40 in AQP3, is able to induce a protein conformational change that closes the pore, thus, preventing glycerol and water permeation. Moreover, a good correlation between the affinity of the Au(III) complex for binding to sulfur donors in Cys residues and AQP3 inhibition effects was highlighted; while no influence of the different oxidative character of the complexes was observed.

4.1 Introduction

Water movement through cells is a prerequisite for all life forms. The discovery and characterization of an abundant protein of the erythrocyte membrane, termed Aquaporin-1 (AQP1), 25 years ago, represented a paradigm shift in the understanding of molecular and membrane water transport. Since then, several studies have shown that AQP1 is a member of a universal family of water and solute-permeable membrane proteins – Aquaporins (AQPs) – which have now been shown to be ubiquitous in all domains of life.[1]

To date, 13 aquaporin isoforms have been identified in mammals (AQP0-12) which are expressed in a wide range of tissues, organized as tetramers embedded in membranes, and permeating water (orthodox aquaporins), glycerol (aquaglyceroporins), and other small solutes.[2]

Aquaporins have been shown to play important roles in physiology and pathophysiology, e.g. maintaining cellular homeostasis and being essential in cellular metabolism. Thus, they have been identified as potential targets for drug development.[3–5] One specific type, aquaglyceroporins, regulate the glycerol content in the epidermis and other tissues and appear to be involved in skin hydration, cell proliferation, carcinogenesis and fat metabolism.[1,6]

To validate the various roles of AQPs in health and disease, and to develop AQP-targeted therapies, the use of highly selective inhibitors provides great potential. However, so far, no reported small-molecule AQP inhibitors possess sufficient isoform selectivity to be candidates for clinical development.[7] The first reported potent and selective inhibitor of human aquaglyceroporin-3 (AQP3) from our group was the water-soluble Au(III) compound, $[Au(phen)Cl_2]Cl$ (phen = 1,10-phenanthroline) (Auphen, Figure 1).[8] Auphen inhibited glycerol transport in human red blood cells (hRBC) with an IC_{50} ca. 0.8μM, while having no inhibitory effect on AQP1 water conductance. *In silico* approaches, including homology modeling, were used to investigate the non-covalent binding of Auphen with human AQP3 at an atomistic level and suggested that the compound could bind to Cys40, with the thiol group being a likely candidate for coordinating to Au(III) ions.[8] The involvement of this residue in the inhibition mechanism was also subsequently confirmed by site-directed mutagenesis studies.[9]

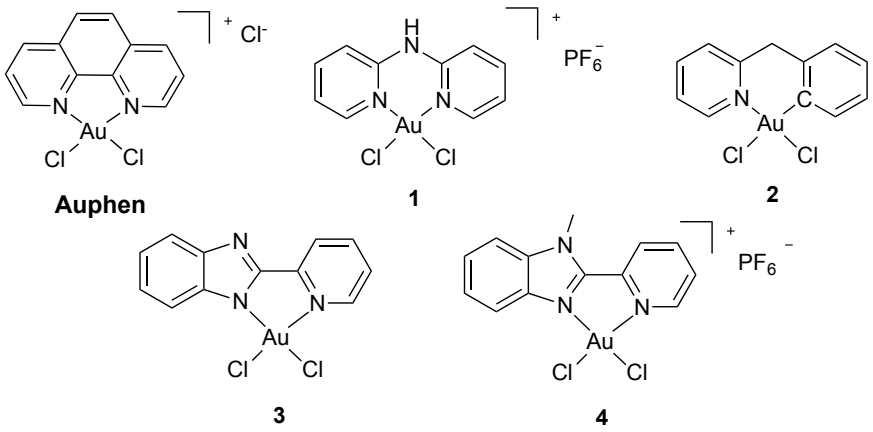


Figure 4.1 Gold(III) complexes tested as human AQP3 inhibitors.

Further studies on Au(III) compounds with different N^N ligand scaffolds allowed the establishment of preliminary structure–activity relationships (SAR).[10] Notably, Quantum Mechanics/Molecular Mechanics (QM/MM) calculations provided evidence that the ligand moiety may play a major role in the compound's selectivity towards a particular AQP isoform,[10] favouring the initial non-covalent interactions between the inhibitor and the extracellular binding pocket of AQP3. Furthermore, Molecular Dynamics (MD) simulations on the adducts of Hg^{2+} ions (benchmark inhibitors of all

AQPs[11,12]) with AQP3 showed that pore closure may be due to indirect protein conformational changes upon metal binding, rather than direct steric blockage of the channel by the inhibitor.[13] Following up on these results, we describe here the hAQP3 inhibition properties of four Au(III) complexes (Figure 4.1) assessed in hRBC by stopped-flow spectroscopy. The series of compounds included three coordination complexes with a dipyridin-2-ylamine (DipyAm) ligand $[Au(Dipyam)Cl_2]PF_6$ (**1**),[14] and with (pyridyl)benzimidazole type ligands – $[Au(PbIm)Cl_2]$ (**3**) (PbIm = 2-(pyridin-2-yl)-benzimidazole)[15] and $[Au(PbImMe)Cl_2]PF_6$ (**4**) (PbImMe = 1-methyl-2-(pyridin-2-yl)-benzimidazole), respectively. Also, for the first time, an organometallic Au(III) compound with a C^N cyclometalated 2-benzylpyridine (py^b-H) ligand (**2**)[16] was tested as AQP inhibitor.

The mechanism of AQP3 inhibition was studied using MD simulations for the most potent inhibitor of the series, allowing the observation of protein conformational changes which lead to pore closure upon gold binding to Cys40. The identification of structure–activity relationships that may link the electronic/structural properties of Au(III) compounds to their biological effects was also carried out. Thus, density functional theory (DFT) calculations were performed to further substantiate and interpret the observed biological effects.

4.2 Results and discussion

Compounds **1**–**3** were synthesized using previously reported procedures[14,17–19]. The new Au(III) complex **4** was obtained in 82% yield by the reaction between 1-methyl-2-(pyridin-2-yl)-benzimidazole[20] in MeCN with an equimolar aqueous solution of $NaAuCl_4$ and an excess of PF_6 at room temperature (r.t.) for 3 h, and was then characterized by various methods. The gold complexes were tested for their AQP1 and AQP3 inhibition properties in hRBC *via* stopped-flow spectroscopy.[21]

As previously reported for Auphen and related compounds,[8,10] some of the new complexes act as inhibitors of glycerol permeation *via* AQP3, whilst not affecting water permeation *via* AQP1. The coordination complex **1** shows moderate inhibition

of water and glycerol permeability (IC_{50} > 20μM), which may be due to its poor physiological stability.[14] Of note, the neutral organometallic compound [Au(py^b-H)Cl_2] **2** showed limited activity (IC_{50} > 50μM). It was hypothesized that the lack of a positive charge would disfavour the non-covalent interaction of the complex with the extracellular surface of AQP3, thus preventing its approach to the pore prior binding to Cys40. In line with this observation, while the neutral complex **3** proved a weak inhibitor of glycerol transport (IC_{50} > 50 M), the novel cationic compound [Au(PbImMe)Cl_2]PF_6 **4** is a very potent AQP3 inhibitor (IC_{50} = 0.6 ± 0.1μM), even more effective than Auphen, and ca. 3 orders of magnitude more potent than complex **3**.

The mechanism of hAQP3 inhibition by complex **4** was analysed using classical molecular dynamics (MD). Due to the lack of an available crystal structure, the tetrameric protein structure of AQP3 was prepared using homology modelling, following an approach previously described by our group[13] and reported in the methodology section.[23] The compound was initially parameterised using DFT and QM/MM to generate Au(III) parameters for the applied Amber force field, and then directly bound to the thiolate of Cys40, in the form [Au(PbImMe)Cl]$^{2+}$. Geometry optimisation was performed on this fragment (Figure 4.2), which was subsequently embedded into monomer A of the AQP3 model. The charge of Au was set to +3 (Au(III)).

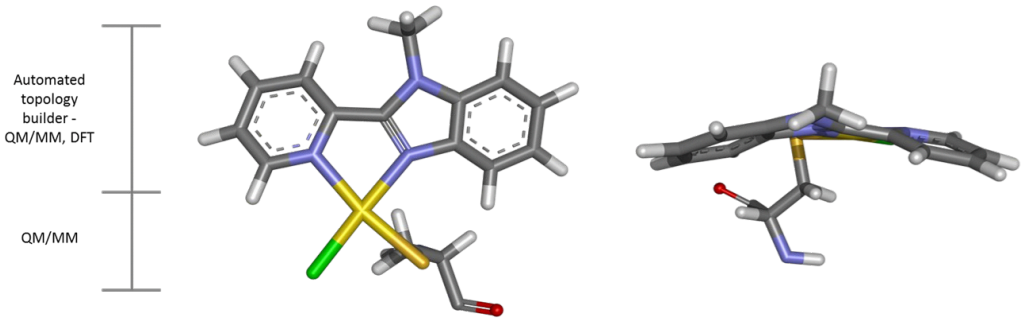


Figure 4.2 Geometry of modified Cys40 with complex **4** after energy minimisation.

Five independent MD simulations (0.5 ns) were conducted to determine the effect of compound **4** on water and glycerol permeation using both: (i) AQP3 or (ii) gold-bound AQP3 (AQP3-Au) models. Figure 4.3 shows the pore size comparison of the structures for monomer A obtained from two representative simulations of each model. Binding

of complex **4** to Cys40 induces shrinkage of the pore, thus, impeding both glycerol and water permeability.

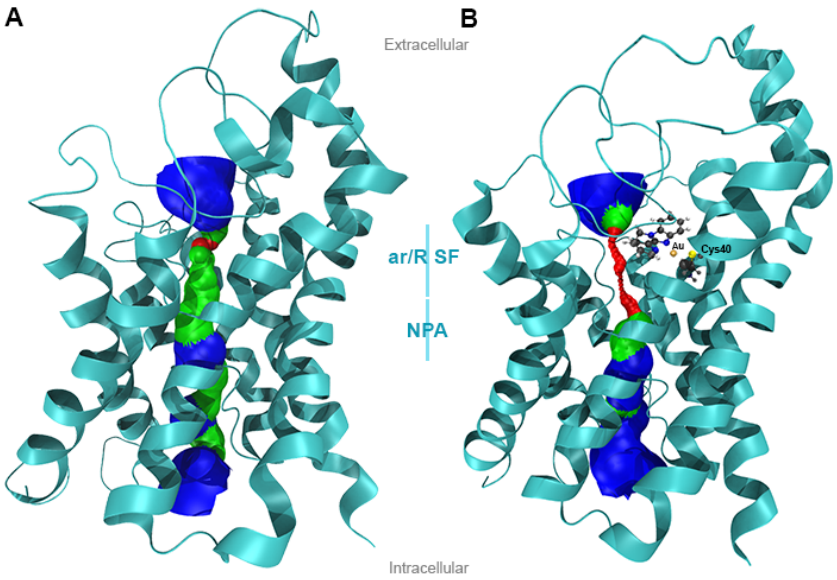


Figure 4.3 (A) Human AQP3 monomer A and (B) AQP3 with modified Cys40 (AQP3-4), showing the effect on pore size (based on VDW radii): red = smaller than single H_2O, green = single H_2O, blue = larger than single H_2O. Complex 4 and Cys40 are shown in ball and stick representation, with atoms coloured by atom type. Generated with HOLE[24] and VMD.[25]

From one simulation of each model, 30 snapshots were taken, and the pore size of each monomer was measured using the HOLE[24] software package as detailed in Figure 4.4. In order to validate the observed trend, five snapshots were taken (100, 300, 450, 600 and 800 frames) from each of the five independent simulations for each model, and the pore size was measured. The average pore size of each monomer was calculated (Figure 4.5) and allowed the ruling out of spontaneous pore geometry fluctuations during the simulation. Surprisingly, the average of the pore size analysis shows that binding of complex **4** to monomer A also constricts monomer D, although not sufficiently to prevent solute conductance, whilst monomers B and C are unaffected (Figure 4.5).

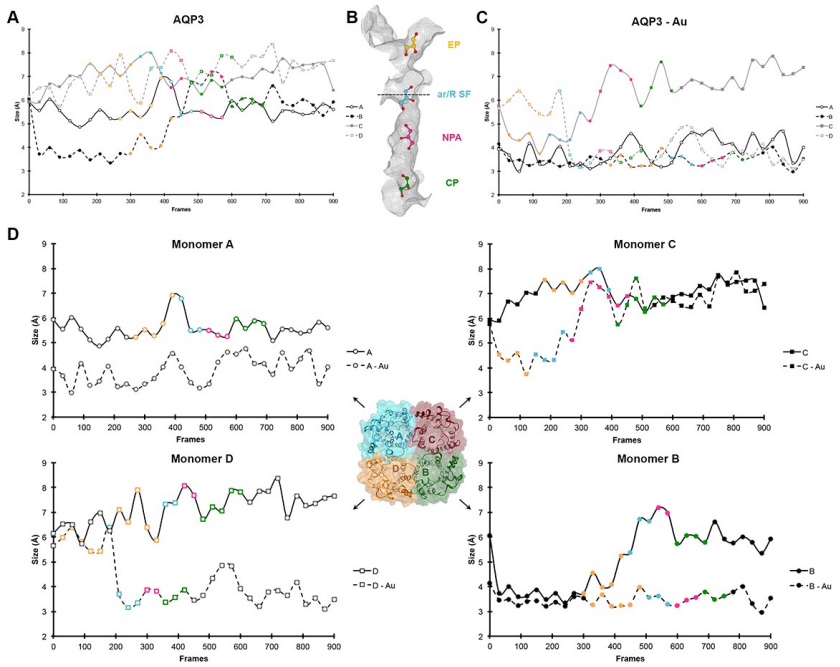


Figure 4.4 Pore size taken at the narrowest point of the channel (ar/R SF), from a representative simulation (0.5 ns) of AQP3 permeation by glycerol in the absence **(A)** or presence **(C)** of the gold complex. Each monomer is represented by a trace, and the coloured markers indicate the glycerol passage from extracellular to cytoplasmic side. **(B)** Pockets for glycerol passage, indicated in yellow (extracellular pocket), blue (ar/R SF), pink (NPA motif) and green (cytoplasmic pocket). The narrowest point of the channel is indicated by a dashed line (approximate localization). In panel D the pore sizes of each monomer are shown, with (dashed line) and without (full line) gold complex. The same colour code of the point markers was used.

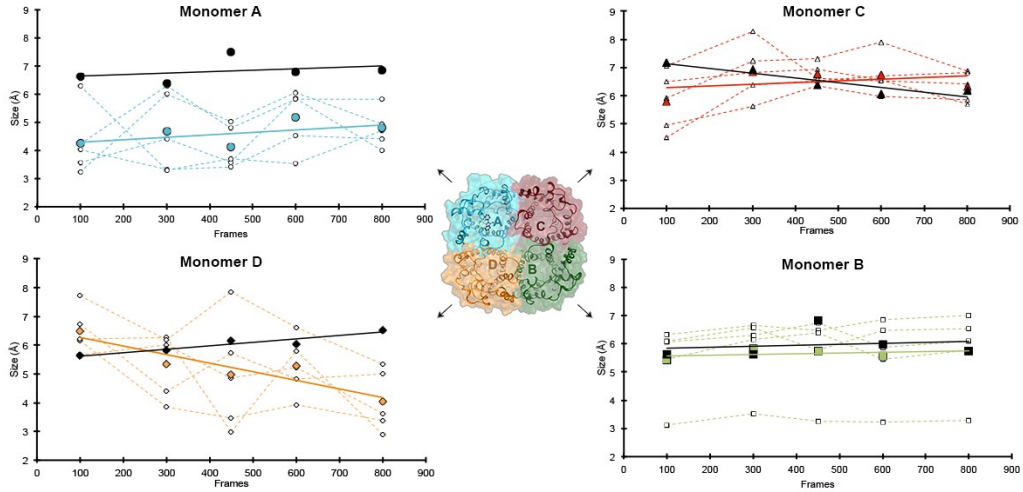


Figure 4.5 Pore size taken at the narrowest point of the channel (ar/R SF) from 10 independent simulations (0.5 ns) of AQP3 (5 without and 5 with gold complex) for each monomer. In black markers, the pore sizes of each monomer are show as an average of 5 independent simulations and with a linear trend line (shown in black). In coloured dashed lines with white markers, the results for each of the 5 individual simulations of AQP3 with the gold complex are shown. In coloured markers is represented the average size for each point, for AQP3 with gold complex **4**, with an added linear trend line, shown in the respective colour.

Root mean square displacement (RMSD) shows that the overall protein conformation is conserved upon gold binding (Figure 4.6).

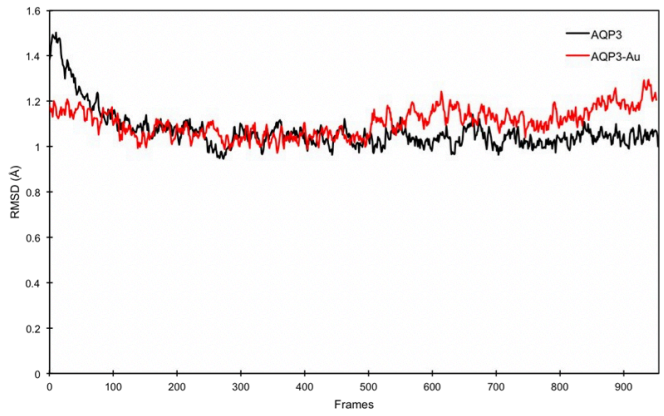


Figure 4.6 RMSD (Å) of the protein backbone atoms of AQP3 (black) and AQP3-Au (red).

Binding of the complex within the pore induced rearrangement in the side chains of the aromatic/arginine selectivity filter residues (ar/R SF) (Figure 4.7). Compound **4** does not appear to be positioned within the channel in a way that would prevent glycerol or water to flux. However, upon Au(III) binding to Cys40, located close to the ar/R SF, compound **4** prevents the Arg218 residue from forming an H-bond with the protein backbone, as usually observed in AQP3, pushing the side-chain outwards into the channel (Figure 4.7).

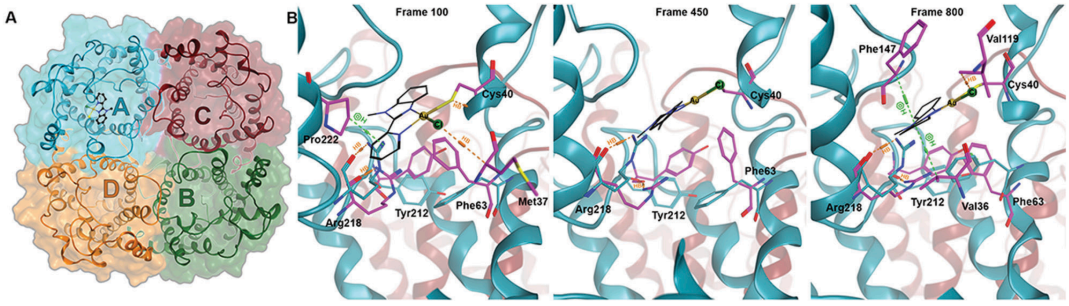


Figure 4.7 A, Tetrameric view of human AQP3 bound to complex **4**. **B**, Structure of the ar/R SF of AQP3 (blue) and upon binding of the gold complex (pink). The gold complex **4** is shown in black with thin sticks, gold in yellow-gold colour and chloride in green, both in ball and stick representation. H-Bonds are shown using orange dashed lines (HB), while H-arene interactions are shown using green dashed lines. Figures in panel B were generated with MOE.[26]

These effects suggest that inhibition of AQP3's water and glycerol permeability is due mainly to conformational changes within the protein induced by binding of the gold complex to Cys40, rather than by the complex's steric hindrance, agreeing with the

previous MD study on the binding of Hg^{2+} to AQP3.[13] Similar effects were observed when a longer MD simulation was run (8 ns, Figure 4.8).

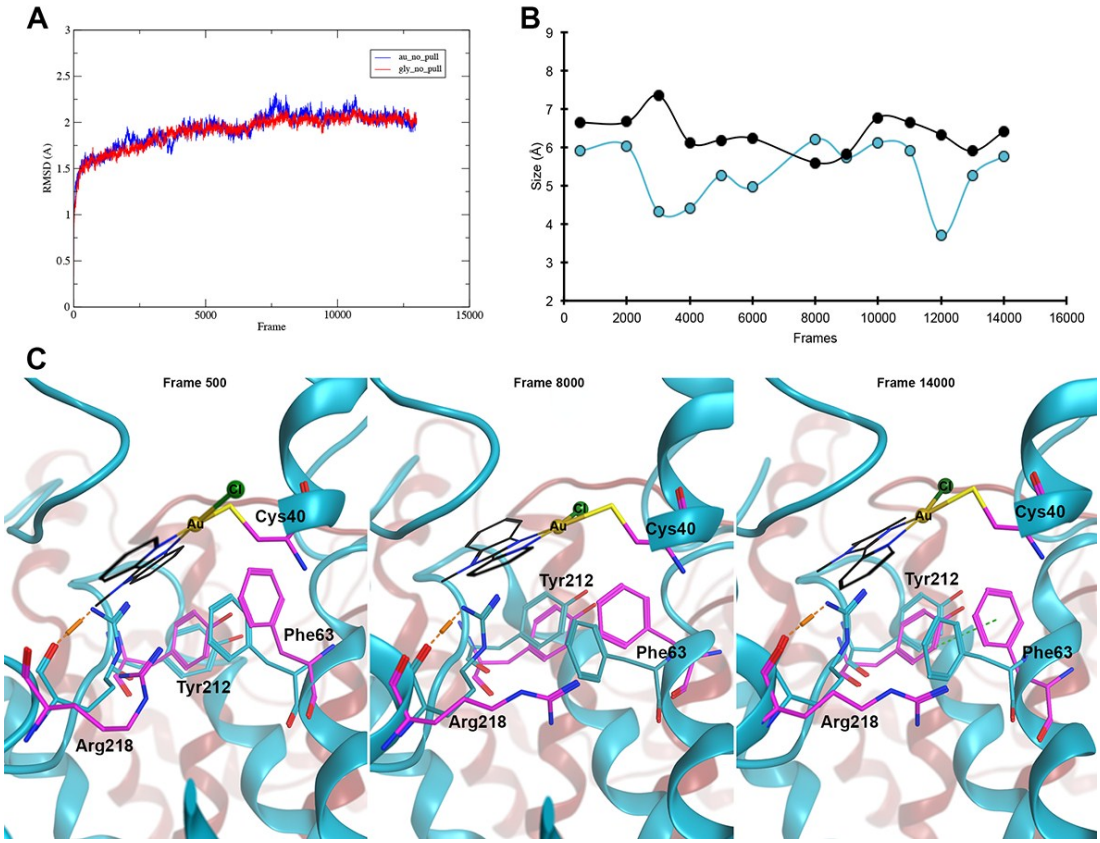


Figure 4.8. MD simulation of AQP3 permeation by glycerol in the presence and absence of the gold complex without pull (8 ns). During the time of the simulation, due to the absence of pull, none of the glycerol molecules crosses the AQP3 channels. **A)** RMSD of the backbone of AQP3's simulation with (blue trace) and without (red trace) gold complex. **B)** Pore size of monomer A measured at the narrowest point of the channel. The simulation with gold complex is shown in blue markers and line, while the simulation without the complex is depicted in black. **C)** Structure of the ar/R SF of AQP3 (blue) and upon binding of the gold complex (pink). The gold complex **4** is shown in black with thin sticks, gold in yellow-gold colour and chloride in green, both in ball and stick representation. H-bonds are shown in orange dashed lines, while H-arene interactions are shown in green dashed lines. The figures in panel C were generated with MOE.[26]

The observed structural changes upon gold binding appear to increase the hydrophobicity of the pore entrance of monomer A in comparison with monomers B to D and the unbound tetrameric model, due to increased exposure of hydrophilic side chains (Figure 4.9). Overall, the symmetry of the tetramer is disrupted in the AQP3-Au model, as can be seen in Figure 4.9, due to the exposure of hydrophilic residues (Arg50 and Asp125). These relatively small changes greatly affect the approach of glycerol molecules to the channel entrance.

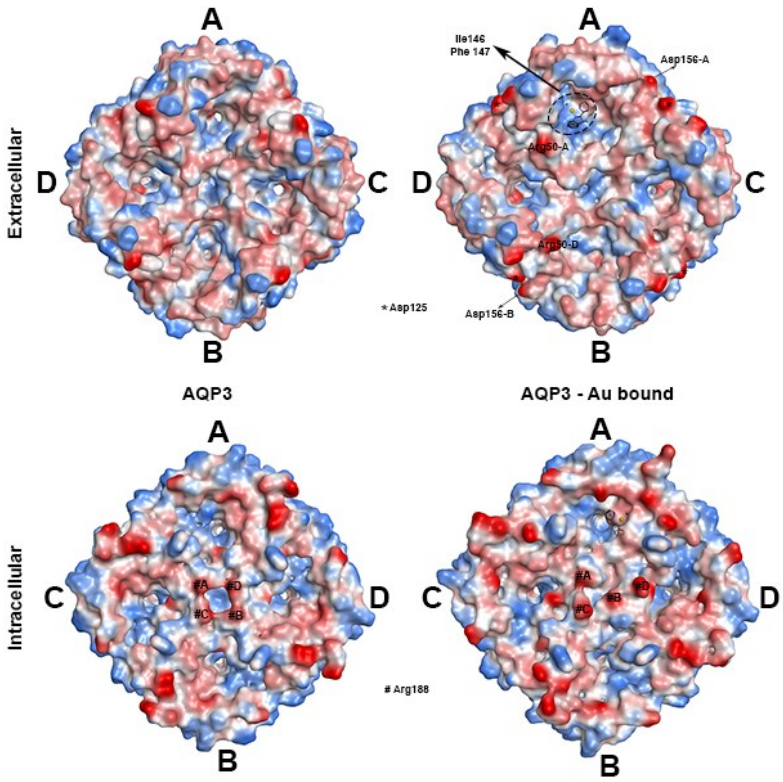


Figure 4.9 Top intra and extracellular view of AQP3 (left) and AQP3-Au bound (right) tetramers hydrophobic surfaces at the start of the simulation. The gold complex **4** is shown over the surface only for indication of its relative position within the tetramer. This does not represent the compound's location across the z-axis. Blue = hydrophobic, red = hydrophilic. Generated with MOE.[26]

To further investigate why complexes **3** and **4** have different AQP3 inhibitory effects and assuming Cys40 as the gold binding site as in the case of Auphen, DFT calculations were performed on the adducts between a cysteinate ligand and compounds **2**–**4**, as well as Auphen, after the substitution of one of the two chlorido ligands (Figure 4.10). The standard Gibbs free energy values of adduct formation are reported in Table 4.1 and show a larger formation energy is observed for the positively charged Au(III) complexes with respect to neutral complex. These results support the hypothesis that the cationic compounds, $[Au(phen)Cl_2]^+$ and $[Au(PbImMe)Cl_2]^+$ **4**, can more easily form adducts with cysteinato residues than the neutral complexes $[Au(py^b)Cl_2]$ **2** and $[Au(PbIm)Cl_2]$ **3**. Therefore, the corresponding AQP3 inhibition effects agree with this trend.

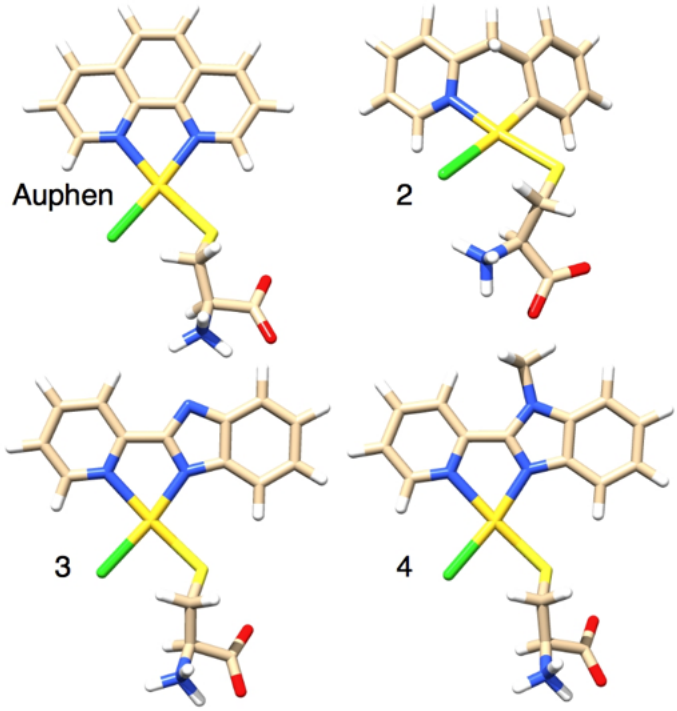


Figure 4.10 Structure of the cysteinate complexes [Au(ligand)ClCys]$^{n+}$ obtained for Auphen, 2, 3 and 4, whose geometry was optimized by DFT calculations; n = 1, 0, 0, 1, for the adducts of Auphen, 2, 3 and 4, respectively.

Table 4.1 Standard Gibbs free energy of formation of [Au(ligand)ClCys]$^{n+}$ adducts in H_2O solution (in kJ/mol, in terms of the SCF energy, ΔE, and of standard Gibbs free energy, ΔG°) calculated through DFT and by Eq. 0 (see Methodology section).

Compound	ΔE	ΔG°
[Au(phen)ClCys]$^+$	-53.7	-45.0
2 [Au(py^b)ClCys]	-10.7	-0.2
3 [Au(PbIm)ClCys]	-27.8	-18.9
4 [Au(PbImMe)ClCys]$^+$	-53.0	-46.6

4.3 Conclusions

Due to the broad range of functions of AQPs in physiology and in disease, the development of small-molecules selective inhibitors is valuable, as these could be used as either chemical probes to detect their function, or as innovative therapeutic agents in a variety of diseases. In this work, a new series of Au(III) complexes has been studied for their human AQP3 inhibition properties, and the cationic complex **4** featuring bidentate N-donor ligands was identified as the most effective inhibitor of

glycerol conductance. Interestingly, the neutral complex **3**, with a similar ligand system, was found to be inactive in comparison. DFT studies showed a good correlation between the compounds' calculated affinity for binding to cysteinate residues and their AQP3 inhibitory activity.

Of note, MD studies conducted for the first time on the Au(III) complex adduct with AQP3 have enabled the observation of protein conformational changes, upon metal binding to Cys40, which are mainly responsible for the observed inhibition of glycerol and water conductance. This finding has important implications for the next generation of AQP3 inhibitors design, in that other amino acid residues could be a potential target, if their coordination to Au(III) ions leads to similar protein conformational changes which result in channel closure. Interestingly, binding of the compound in one monomer appeared to affect substrate permeability in the adjacent monomers and to also alter the overall extracellular hydrophobic/hydrophilic nature of the tetramer surface, which, in turn, affects the approach of the substrates to the pore entrance.

4.4 Methodology Section

Homology modelling and molecular dynamics:

The 3D structure of hAQP3 was obtained by homology modelling using the Molecular Operating Environment (MOE 2013.08).[26] The choice of a template structure was based on the sequence identity between hAQP3 and the sequence of the AQPs with available resolved structures. The isoform showing the highest sequence similarity with hAQP3 is the bacterial isoform GlpF, with 34.68% of sequence identity, which was then chosen as a template structure to generate a homology model of hAQP3. The template was selected among the structures with the best resolution (2.70 Å) without any substrate (pdb 1LDI).[28] The tetrameric form was assembled and the structure was prepared and protonated at pH 7, with the Amber12EHT force field, as described in Spinello et al[13], using the Molecular Operating Environment (MOE 2013.08).5. 50 intermediate models of AQP3 were generated and averaged to obtain the final homology model. The model obtained was checked for reliable rotamers involving the

side chains in the regions of ar/R SF and NPA, by comparison with the available crystal structures of all the other human and microbial AQP isoforms (pdb codes 1H6I, 3GD8, 3D9S, 1RC2, 1LD1 and 3C02). The structure was protonated at pH 7 and an energy minimization refinement was performed, with fixed Cα atoms.

Afterwards, the effect of binding of Au(III) complex 4 on AQP3 glycerol permeability was investigated using classical molecular dynamics (MD). The molecular system consisted of the homology model of the hAQP3 tetramer within a double layer of 166 palmitoyl-oleyl-phosphatidyl-coline (POPC) lipid, prepared using the CHARMM-GUI online server[29,30], using Amber99SB-ILDN in combination with the Slipids (Stockholm lipids) force field for lipids.[23,31] Four glycerol molecules were placed into the system, one above each pore entrance, approximately 30 to 35 Å (to residue TRY212 at the pore entrance). The system was solvated with 35379 water molecules and used a modified amber99SB-ILDN force field, with the parameters for glycerol and 1-methyl-2-(pyridine-2-yl)-benzimidazole generated by the Automated Topology Builder and Repository (ATB,version 2.2) website using the B3LYP/6-31G* basis set[32], using a combination of semi-empirical QM and DFT. Complex **4** was parameterised using DFT and QM/MM in order to obtain the parameters for the Au(III) ions in the force field. Afterwards, the complex was bound to the thiolate of a cysteine residue, to further integrate in the protein complex. The complex's geometry was further optimized (see Figure 4.2) and then incorporated into monomer A of AQP3.

All simulations were run using the GROMACS 5.1.2 simulation software.[33] Particle-mesh Ewald method was used for calculating electrostatic interactions. The Verlet cut-off scheme with a cut-off distance of 1.4 nm was used for short-range repulsive and attractive interactions and Lincs was used to constrain all bond lengths. Nose-Hoover temperature coupling was used to maintain the temperature of the system (τ = 0.5 ps) at 310 K. The Parrinello-Rahman algorithm was used to maintain the pressure of the system at 1 bar with a coupling constant of τ = 1.0 ps. Simulations were equilibrated for 100 ps before production. The four individual glycerol molecules were defined in the index and coupled in the pull code (e.g. gly_1 to chain A). A total of 10 MD simulations (5 with the Au(III) complex present and 5 without) were run for 0.5 ns using the direction COM pull procedure, in each case applying a separate yet equal harmonic restraint force to each solute molecule of 600 kJ mol^{-1} nm^2 with a rate of

0.02 nm ns-1 along the z-axis. Two 4000000 step runs, or 8 ns were run using the same two model systems and parameters. For these runs the pull code was omitted, therefore removing any biasing of the system.

Pore size measurements: Each pore radius was calculated using the Hole 2.0 program[24], which determines the internal surface based on atomic van der Waals radii. A snapshot at the beginning of each simulation was taken and coordinates for the center of each pore (monomers A to D), at the ar/R SF, were used to generate the pore radius along the z-axis.

In order to assess if the gold complex significantly affects the pore size by inducing protein conformational changes, five simulations, with and without gold complex, were performed. From these, 5 snapshots were taken (100, 300, 450, 600 and 800 frames) and the pore size was measured in each. For consistency, the narrowest part of the pore was considered to be the distance between the side chains of Tyr212 and Arg218 and, for each frame, the distance between the closest N atom of the side chain of Arg218 and C of the aromatic ring of Tyr212. The same methodology was used to assess the pore size of monomer A for the long (8 ns) simulation.

DFT calculations:

DFT calculations were performed on the structures of cysteine, HCl, Auphen, **2**, **3**, **4** and on the adducts obtained by substituting one chlorido with a cysteinato ligand (see Figure 4.10), by using the M06-L[34] functional, the Lanl2dz[35] basis set for Au, S and Cl atoms and the 6-31G(d,p)[36,37] basis set for C, N, O and H atoms. Solvent effects were evaluated by full geometry optimization within the implicit water solvent, reproduced by the polarizable continuum model (PCM).[38] Vibrational frequency calculations, within the harmonic approximation, were performed to confirm that each optimized geometry corresponded to a minimum in the potential energy surface. Moreover, vibrational frequency calculations allowed us to estimate the standard Gibbs free energy values, at 298.15 K, of each energy minimum structure, both in vacuum and in solution. All calculations were performed by the Gaussian 09 program package.[39] The energy values of formation of the gold complex-Cys adducts, reported in Table 4.1 were obtained by the following hypothetical reaction: $[Au(Ligand)Cl_2]^{n+}$ + HCys → $[Au(Ligand)ClCys]^{n+}$ + HCl, where n is the charge of the metal complex,

and calculated by the Eqn. 0 below, where E can be either the self-consistent field (SCF) energy or the standard Gibbs free energy in solution:

$$\Delta E = E[Au(Ligand)ClCys] + E[HCl] - E[Au(Ligand)Cl_2] - E[HCys] \quad \text{(Eq. 0)}$$

4.5 References

1 G. Soveral, S. Nielsen and A. Casini, *Aquaporins in Health and Disease: New Molecular Targets for Drug Discovery*, CRC Press, Taylor & Francis Group, 2017.

2 G. Benga, *Mol. Aspects Med.*, 2012, **33**, 514–517.

3 A. De Almeida, G. Soveral and A. Casini, *Med.Chem.Commun*, 2014, **5**, 1444–1453.

4 A. S. Verkman, M. O. Anderson and M. C. Papadopoulos, *Nat. Rev. Drug Discov.*, 2014, **13**, 259–77.

5 E. Beitz, A. Golldack, M. Rothert and J. Von Bülow, *Pharmacol. Ther.*, 2015, **155**, 22–35.

6 A. S. Verkman, *Annu. Rev. Med.*, 2012, **63**, 303–316.

7 G. Soveral and A. Casini, *Expert Opin. Ther. Pat.*, 2017, **27**, 49–62.

8 A. P. Martins, A. Marrone, A. Ciancetta, A. G. Cobo, M. Echevarría, T. F. Moura, N. Re, A. Casini and G. Soveral, *PLoS One*, 2012, **7**.

9 A. Serna, A. Galán-Cobo, C. Rodrigues, I. Sánchez-Gomar, J. J. Toledo-Aral, T. F. Moura, A. Casini, G. Soveral and M. Echevarría, *J. Cell. Physiol.*, 2014, **229**, 1787–1801.

10 A. P. Martins, A. Ciancetta, A. deAlmeida, A. Marrone, N. Re, G. Soveral and A. Casini, *ChemMedChem*, 2013, **8**, 1086–1092.

11 R. I. Macey and R. E. L. Farmer, *Biochim. Biophys. Acta*, 1970, **211**, 4–6.

12 G. M. Preston, T. P. Carroll, W. B. Guggino and P. Agre, *Science (80-.).*, 1992, **256**, 26–28.

13 A. Spinello, A. De Almeida, A. Casini and G. Barone, *J. Inorg. Biochem.*, 2016, **160**, 78–84.

14 A. Casini, M. C. Diawara, R. Scopelliti, S. M. Zakeeruddin, M. Grätzel and P. J. Dyson, *Dalt. Trans.*, 2010, **39**, 2239–2245.

15 M. Serratrice, M. A. Cinellu, L. Maiore, M. Pilo, A. Zucca, C. Gabbiani, A. Guerri, I. Landini, S. Nobili, E. Mini and L. Messori, *Inorg. Chem.*, 2012, **51**, 3161–3171.

16 M. A. Cinellu, A. Zucca, S. Stoccoro, G. Minghetti, M. Manassero, M. Sansoni and V. I. Chcme, 1996, 4217–4225.

17 M. A. Cinellu, A. Zucca, S. Stoccoro and G. Minghetti, *J. Chem. Soc. Dalt.*

Trans., 1996, 4217–4225.

18 L. S. Hollis and S. J. Lippard, *J. Am. Chem. Soc.*, 1983, **105**, 4293–4299.

19 B. Bertrand, S. Spreckelmeyer, E. Bodio, F. Cocco, M. Picquet, P. Richard, P. Le Gendre, C. Orvig, M. A. Cinellu and A. Casini, *Dalt. Trans.*, 2015, **44**, 11911–11918.

20 W.-K. Huang, C.-W. Cheng, S.-M. Chang, Y.-P. Lee and E. W.-G. Diau, *Chem. Commun.*, 2010, **46**, 8992.

21 A. De Almeida, A. F. Mó, D. Wragg, M. Wenzel, P. Kavanagh, G. Barone, S. Leoni, G. Soveral and A. Casini, *Chem. Commun.*, 2017, 3830–3833.

22 D. F. Savage and R. M. Stroud, *J. Mol. Biol.*, 2007, **368**, 607–617.

23 J. P. M. Jämbeck and A. P. Lyubartsev, *J. Chem. Theory Comput.*, 2013, **9**, 774–784.

24 O. S. Smart, J. G. Neduvelil, X. Wang, B. A. Wallace and M. S. P. Sansom, *J. Mol. Graph.*, 1996, **14**, 354–360.

25 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33–38.

26 C. C. G. ULC, 2018.

27 A. de Almeida, B. L. Oliveira, J. D. G. Correia, G. Soveral and A. Casini, *Coord. Chem. Rev.*, 2013, **257**, 2689–2704.

28 D. Fu, A. Libson, L. J. W. Miercke, C. Weitzman, P. Nollert, J. Krucinski and R. M. Stroud, *Science (80-.).*, 2000, **290**, 481–486.

29 S. Jo, T. Kim, V. G. Iyer and W. Im, *J. Comput. Chem.*, 2008, 29, 1859–1865.

30 E. L. Wu, X. Cheng, S. Jo, H. Rui, K. C. Song, E. M. Dávila-Contreras, Y. Qi, J. Lee, V. Monje-Galvan, R. M. Venable, J. B. Klauda and W. Im, *J. Comput. Chem.*, 2014, **35**, 1997–2004.

31 J. P. M. Jämbeck and A. P. Lyubartsev, *J. Chem. Theory Comput.*, 2012, **8**, 2938–2948.

32 A. K. Malde, L. Zuo, M. Breeze, M. Stroet, D. Poger, P. C. Nair, C. Oostenbrink and A. E. Mark, *J. Chem. Theory Comput.*, 2011, **7**, 4026–4037.

33 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindah, *SoftwareX*, 2015, **1**–**2**, 19–25.

34 Y. Zhao and D. G. Truhlar, *J. Chem. Phys.*, 2006, **125**.

35 P. J. Hay and W. R. Wadt, *J. Chem. Phys.*, 1985, **82**, 270–283.

36 P. C. Hariharan and J. A. Pople, *Theor. Chim. Acta*, 1973, **28**, 213–222.

37 M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J.

DeFrees and J. A. Pople, *J. Chem. Phys.*, 1982, **77**, 3654–3665.

38 J. Tomasi, B. Mennucci and R. Cammi, *Chem. Rev.*, 2005, **105**, 2999–3093.

39 and D. J. F. G. W. T. M. J. Frisch, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A, 2013.

Chapter 5 - Molecular basis of aquaporin-7 permeability regulation by pH

Based on the paper:

Andreia F. Mósca, Andreia de Almeida, Darren Wragg, Ana P. Martins, Farzana Sabir, Stefano Leoni, Teresa F. Moura, Catarina Prista, Angela Casini, Graça Soveral. Cells, 2018, 7, 207.
DOI: 10.3390/cells7110207

Experimental analysis carried out by Dr Andreia F. Mósca.
AQP7 homology models constructed by Dr Andrea de Almeida.

Abstract: The aquaglyceroporin AQP7, a member of the aquaporin membrane channel family, facilitates the conductance of water and glycerol across cell membranes and is essential for lipid and energy homeostasis. Regulation of glycerol permeability *via* AQP7 is considered a promising therapeutic target for fat-related metabolic complications. Here, we used a combination of *in vitro* and *in silico* approaches to investigate its regulation by pH. We found that AQP7 moves from fully permeable at neutral pH to virtually closed at acidic pH and that Tyr135 and His165 facing the extracellular environment are crucial residues for channel permeability. Furthermore, rather than a reduction in the pore size, protonation of key residues changes AQP7's protein surface electrostatic charges which may result in a decrease glycerol's binding affinity and result in decreasing permeability. It was also found; a number of the pH-sensitive residues are located at the monomer-monomer interface and the decrease in permeability may be due to cooperativity between AQP7 monomers. Due to the importance of glycerol permeation *via* AQP7 in multiple pathophysiological conditions, the mechanism of hAQP7 pH-regulation could help in the design of highly selective modulators for targeting aquaglyceroporin-related disorders.

5.1 Introduction

Aquaporins (AQPs) are a family of small membrane proteins expressed in almost every organism and tissue type. They are involved in the bidirectional conductance of water and small uncharged solutes across cell membranes in response to concentration gradients or osmotic and hydrostatic pressure. The thirteen mammalian AQPs (AQP0-12) are grouped based on their primary structure and permeability. Aquaglyceroporins (AQP3, 7, 9, 10) facilitate the permeation of glycerol and other small solutes in addition to water[1]. Due to this selective permeability, aquaglyceroporins play an important role in glycerol acquisition and metabolism within skin, fat and liver[2–4] tissues and are thought to be linked to obesity and metabolic-related complications, such as metabolic syndrome[5,6].

As with many types of channels and transporters, AQPs can be subject to regulation, either by transcriptional/translational mechanisms[7], protein trafficking[8] , or by short-term channel regulation, known as gating. Gating is usually achieved by mechanisms directly affecting the channel's conformation after it is embedded in the plasma membrane, which then impacts on its permeability[9]. It has been shown that various eukaryotic water-selective aquaporins can be gated by phosphorylation[10,11], pH[11–13], divalent cations[14,15], or membrane surface tension[16–19]. In comparison, the regulation of aquaglyceroporins and glycerol permeability is less understood. With AQPs becoming an emerging protein target for drug development[20], increasing the understanding of human aquaglyceroporin regulation could help facilitate the identification and development of new modulators with potential therapeutic applications[21,22]. Even with the limited research into aquaglyceroporin regulation, a number of studies have shown gating of AQP3 by pH, copper and nickel[23–26].

AQPs have a highly conserved hourglass-shaped internal pore, resulting in a steric barrier for solutes. This steric constriction within the pore consists of three (for aquaglyceroporins) or four (for aquaporins) amino acid residues, one of which is a highly conserved arginine, and at least one other having an aromatic side chain. Thus, this constriction site is named aromatic/arginine selectivity filter (ar/R SF). A second solute exclusion barrier is found in the centre of the pore, formed where the ends of two semi-helices meet and consists of NPA (asparagine-proline-alanine) amino acids in each helix. The NPA motif is not always conserved across isoforms, the proline and alanine may be substituted, whereas the two asparagine residues are highly conserved. The two asparagines create a positive dipole moment within the pore and prevent the permeation of certain charged species [21].

This study investigated the pH regulation of human AQP7, the main aquaglyceroporin expressed in human adipocytes and highly relevant in maintaining efficient glycerol fluxes in healthy adipose tissue, as well as involved in obesity and fat-related metabolic complications[4–6,27].The experimental investigation involved expressing hAQP7 in an optimized yeast model developed in the group of Prof. Graça Soveral (University of Lisbon)[9]. Using this cellular model, the function of the hAQP7 channel, with respect to water and glycerol permeability, was investigated. Moreover, inhibition of glycerol permeation via AQP7 was also confirmed for Auphen -

[Au(phen)Cl_2]Cl (phen = 1,10-phenanthroline) - a gold(III) compound previously reported by our group as a potent inhibitor of human aquaglyceroporins AQP3 and AQP7 in their native expression systems[28,29].

Afterwards, the hAQP7 channel regulation by pH was also studied using the same yeast model. The experimental data showed that, while at pH 7.4, hAQP7 is fully functional, with maximal permeability at pH 6.5 and above, a marked reduction (more than 90% decrease) of P_f and P_{Gly} was observed at pH 5.0 (Figure 5.1). Interestingly, despite this clear reduction, P_f and P_{Gly} were still found to be significantly higher than in control cells. This indicates that although the permeability is substantially reduced, the pore does not completely closed at pH 5.0. Interestingly, the pH-dependence profile was found to be similar for both water and glycerol permeation, with pK_a values estimated as 5.88 ± 0.01 and 5.85 ± 0.01, respectively. Additionally, the estimated E_a at three distinct pH values (pH 5.0, 6.5, and 7.5) corroborate the proposed channel regulation upon extracellular acidification[30].

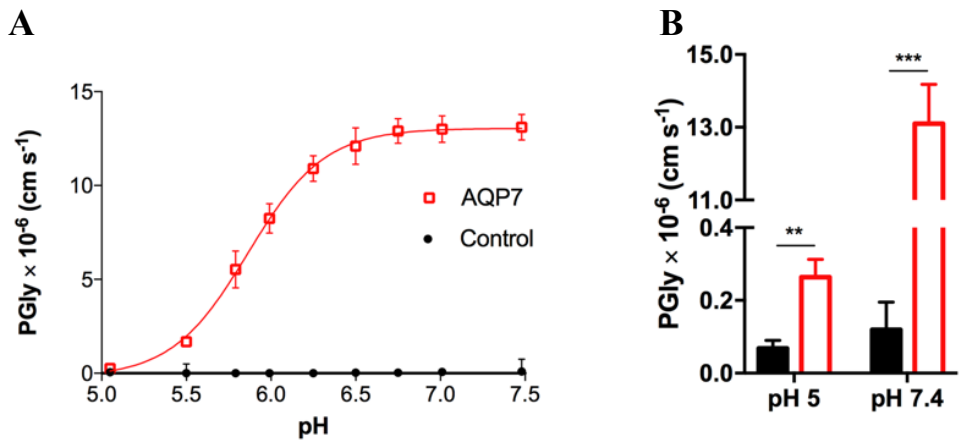


Figure 5.1 pH-dependence of hAQP7 permeability. (A) P_f and (B) P_{Gly} at pH 5 and pH 7.4. pH-dependence of hAQP7 was analysed by fitting the experimental data to a Hill equation from where the pKa values were estimated. Permeability assays were performed at 23 ºC. Data are mean ± SEM of four independent experiments. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Based on the experimental results we then used *in silico* techniques, namely molecular modelling and molecular dynamics (MD) approaches, to investigate the mechanisms involved in hAQP7-pH regulation at a molecular level and to identify the amino acid residues involved in the pH gating. This led to further experimental investigations using site-directed mutagenesis combined to help us understand how hAQP7 regulation can contribute to the roles of this isoform in physiology and

patho-physiology, helping the development of aquaglyceroporin-targeted therapies. The following chapter covers the computational aspects of this study along with a description of the mutagenesis results, which fostered additional computational work.

5.2 Results and discussion

5.2.1 WT hAQP7 molecular dynamics calculations

Based on the experimental findings on the pH gating of hAQP7, we initially used a molecular modelling approach, previously developed by our group[26], to gain mechanistic information. Therefore, a homology model of hAQP7's tetramer was obtained using MOE software (MOE 2012.10; CCG 2012)[31], and based on the best available structure of *E. coli*'s glycerol facilitator (GlpF, pdb code 1FX8)[32]. Homology modelling showed that while hAQP7 maintains the highly conserved aromatic/arginine selectivity filter (ar/R SF) (consisting of Phe74, Tyr223 and Arg229, it does not contain the NPA motif.

The refined homology model of hAQP7 was protonated using the PROPKA 3.1 package[33] at pH 5 and 7. Five independent MD simulations were run for 0.5 ns for each pH (pH 5 and 7) with four glycerol molecules, located in the extracellular side of the membrane, pulled along the Z-axis towards the intracellular space, mimicking imposed glycerol gradients. A total of 10 simulations were run to evaluate glycerol permeability. Afterwards, all simulations were analysed, and glycerol molecules monitored for successful permeation of hAQP7. To investigate the possible pore closure and conformational changes we suspected from the experimental results at low pH, the size and shape of the pore were analysed using HOLE[34], from snapshots taken at five time points in each MD simulation. Figure 5.2 shows a surface representation of channel A from both systems, next to the average channel size of all monomers throughout the simulations. The results showed no differences were observed in the size of the channels at the two pH values for both NPA and ar/R SF, implying that the decrease in glycerol and water permeability may not be due to protein conformational changes and pore closure. The pore size measurements do show there is an observable

difference in the extracellular pocket (EP) (0.5 Å broader), just above the ar/R SF in the case of the pH7 system. This small variation may be due to system fluctuations during the simulation (as previously observed for other aquaporin isoforms[10]), as it was not detected in all the individual calculations (Figure 5.3A). Yet, a broader extracellular binding pocket due to conformational changes at pH 7, could favour glycerol passage through the pore, leading to increased permeability.

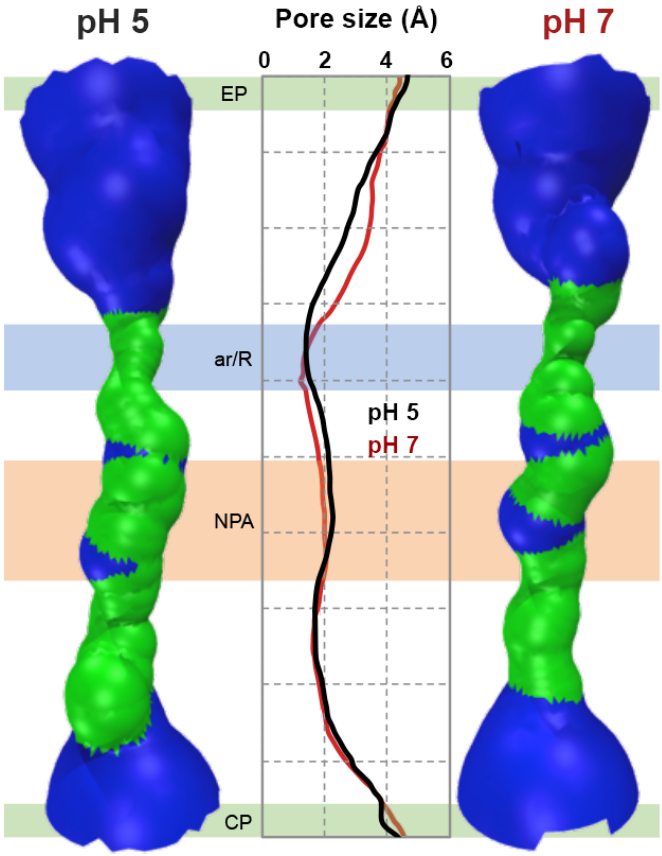


Figure 5.2 Pore size of hAQP7 monomer A at pH 5 and pH 7**.** Surface of a representative snapshot of monomer A (based on VDW radii): red = smaller than single H_2O, green = single H_2O, blue = larger than single H_2O. Pore size represented as an average of all monomers in 5 simulations at each pH value. EP – extracellular pocket, ar/R – aromatic/arginine selectivity filter, NPA – NPA motif, CP – cytoplasmic pocket. Figure generated by HOLE[34] and VMD[35].

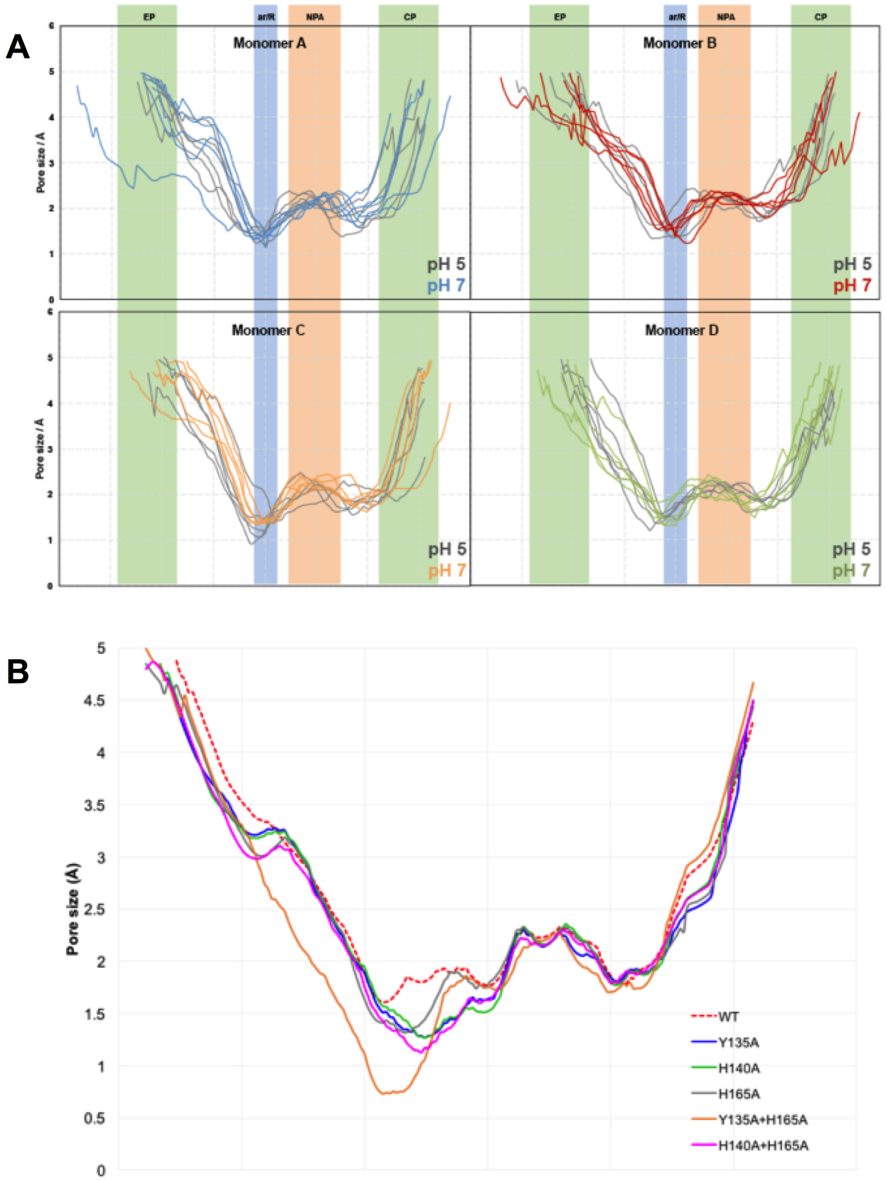


Figure 5.3 A, Pore size of the four AQP7 monomers at pH 5 and pH 7. Each line represents an average pore size for each simulation, by using the pore size taken from five snapshots of each simulation. EP – extracellular pocket, ar/R – aromatic/arginine selectivity filter, NPA – NPA motif, CP – cytoplasmic pocket. **B,** Pore size average of the four AQP7 monomers for each mutant and WT. Pore size obtained with HOLE for each homology model[34].

With conformational analysis of the pore indicating that the protein does not alter pore size when protonated at pH 5 (Figure 5.2), the next stage was to investigate if glycerol and water fluxes were affected within the simulations.

Figure 5.4A shows glycerol permeation as an average number of glycerol molecules able to cross the membrane *via* the monomer channels. Glycerol permeability is shown to be significantly higher at pH 7 than at pH 5 corroborating the obtained experimental P_{Gly} results (Figure 5.1).

To observe water permeation, longer simulations (20 ns) were calculated at both pH values, with the glycerol pull code removed. The passage of water molecules through the pores was evaluated using a python script to count successful permeation events, developed by us and based on previous work[36]. Only water molecules which completely crossed the AQP7 channel were counted, excluding any molecules crossing *via* the lipid bilayer. The molecular dynamics used represents an equilibrium simulation, therefore, it is only possible to estimate the diffusion constant (D_w), rather than the permeability coefficient. For this purpose, D_w of the single-file water molecules was estimated using the Einstein relation[37,38], as described by Horner et al[39] : $D_w = k_o z^2/2$, where z represents the average distance between two water molecules in the single-file region and k_o represents the transport rate. The transport rate was then calculated as the total number of water molecules crossing the channel in the Z-axis direction, calculated by the script, divided by the length of the simulation (in seconds). The final data is represented as the mean ±SEM. As shown in Figure 5.4B, D_w evaluated at pH 7 was significantly higher than at pH 5.

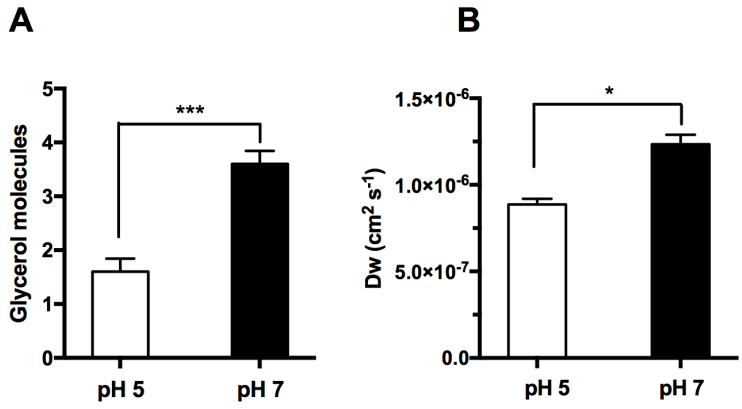


Figure 5.4 pH dependence of glycerol and water permeation across the hAQP7 channel (*in silico*). **A**, Number of glycerol molecules crossing the AQP7 tetramer in MD studies, represented as mean ± SEM (n = 5). **B,** Water diffusion coefficient calculated by MD simulations, represented as mean ± SEM (n = 3). n = number of simulations. * $P < 0.05$, *** $P < 0.001$.

To conclude, even though the MD and stopped-flow spectroscopy data may not be directly comparable, as the *in silico* studies are not performed under osmotic conditions, the *in silico* and *in vitro* data are in accordance with each other and demonstrate that glycerol and water permeation *via* hAQP7 are pH-regulated.

5.2.2 hAQP7 mutagenesis experimental studies

Previous research on human AQP3 using site-directed mutagenesis[24] has shown that only four residues (His53, Tyr124, Ser152 and His154) appear to be responsible for AQP3 pH gating. The relevance of these residues on hAQP3 was recently investigated using MD and it was proposed that gating occurs in the main due to protonation of His154 and its interaction with the neighbouring His129[26]. From this work, it was decided to investigate the mechanism of pH regulation based on sequence alignment of hAQP3 and hAQP7. A superposition of the two homology models (preparation of the hAQP3 homology model is described in reference[26]) showed that the corresponding residues in hAQP7 are Tyr64, Tyr135, His140, Pro163 and His165. The location of the key residues Tyr135, His140 and His165 in the hAQP7 tetramer are shown in Figure 5.5A and B.

Detailed analysis of the protonation sites for each model revealed only four residues are protonated at pH 5, which are not at pH 7 (Figure 5.6): these were, on extracellular side, His140 (in all monomers), intracellularly, Glu40 (monomers A, C and D), His92 (monomers B, C and D) and Glu202 (in all monomers). Based on the sequence/structural alignment and protonation studies of the homology models, the residues facing the extracellular environment were chosen for further site-directed mutagenesis studies. These residues were Tyr135, His140 and His165. As previously, shown in Figure 5.2, changes in pore size are only observed on the extracellular side at pH 7, thus suggesting extracellular amino acids may play a role in pH-dependency. As Pro163 does not have a protonatable side chain, and Tyr64 is not protonated at pH 5, these residues were not investigated further.

Thus, to determine the importance of His165, His140 and Tyr135 experimentally, the selected amino acids were substituted by alanine residues (Ala). Due to the proposed role of His165 in AQP3 pH gating[26], double mutants were also generated. Results showed no differences in membrane abundance were observed in cells incubated with different pH or osmolarity. Permeability of the AQP7 mutants was then further investigated under the same pH range as previously used for WT[30].

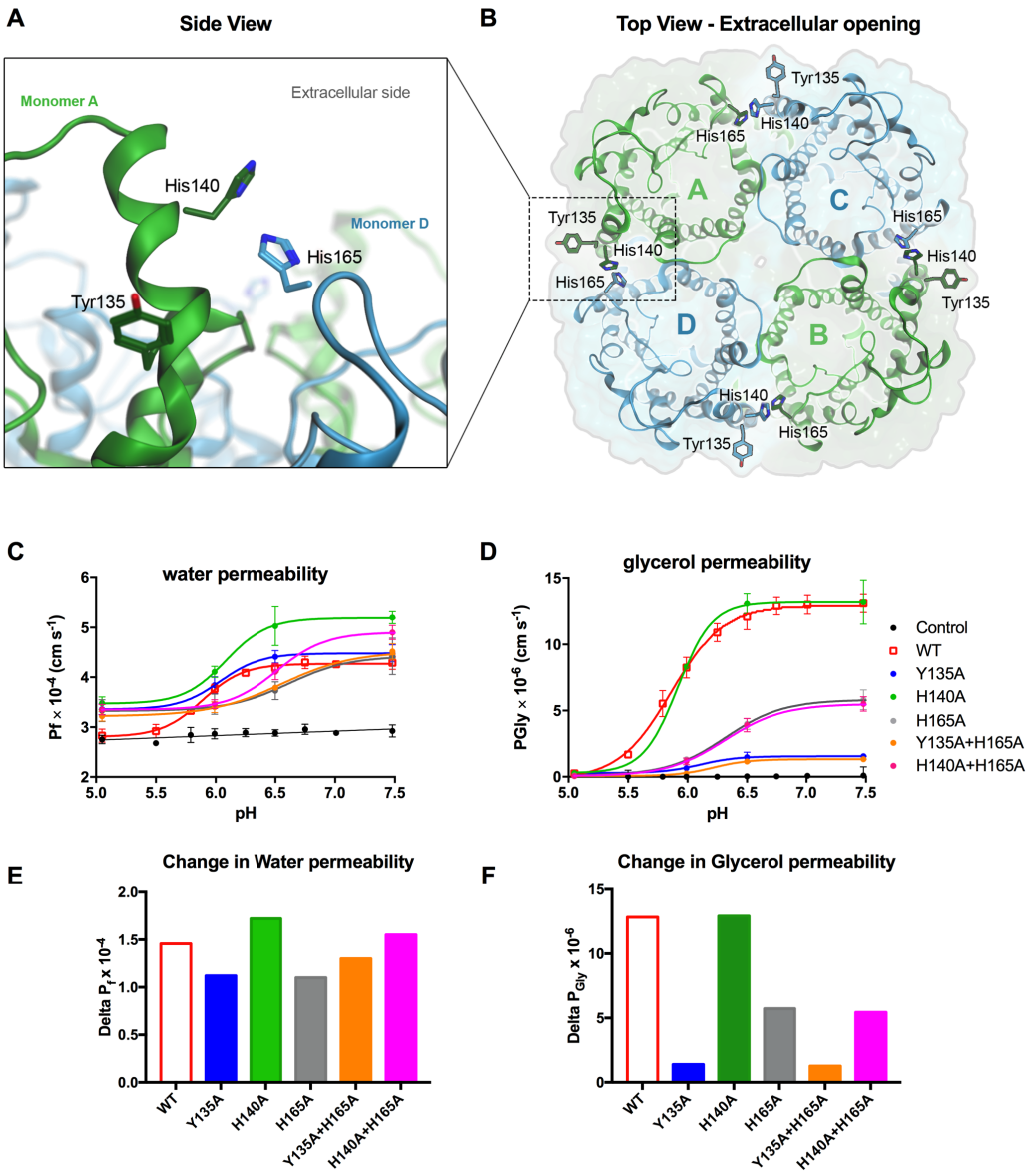


Figure 5.5 Putative residues involved in hAQP7 pH regulation. (A) side and (B) top views of the refined homology model of human AQP7 in its tetrameric assembly, in cartoon representation of the tertiary structure, with ribbon representation of each monomer. Amino acid residues are shown in stick representation and backbone and hydrogens are hidden for clarity. Residues are coloured according to the corresponding monomer, as shown in B. Figure generated with MOE[31]. (C) Water permeability (P_f) and (D) glycerol permeability (P_{Gly}) measured at pH 5 to 7.5, of control cells and cells expressing hAQP7-WT and -mutants. (E) Magnitude of the change in water (Delta P_f) and in (F) glycerol (Delta P_{Gly}) permeability (from pH 7.4 to pH 5) due to point mutations. All permeability assays were performed at 23 ºC. Data (mean ± SEM, n=4 for each data set) were fitted according to the Hill equation.

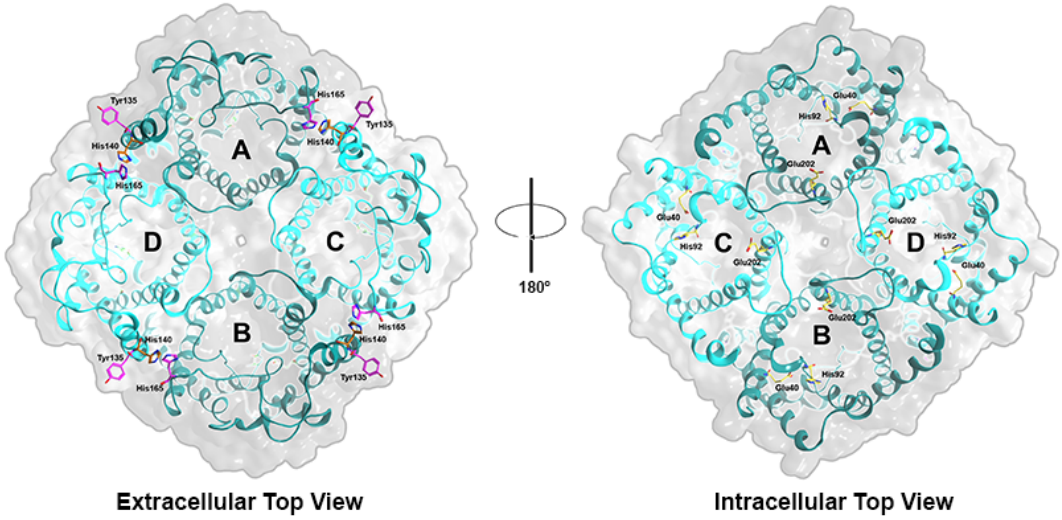


Figure 5.6 A, Top extracellular and **B**, intracellular views of the homology model of human AQP7 in its tetrameric assembly in cartoon representation of the tertiary structure, with ribbon representation (in blue) and surface representation of residues lining the channel (grey). Relevant amino acid residues are shown in stick representation. In yellow the residues that are protonated at pH 5 in the intracellular side are shown. His140 residues in each monomer are protonated at pH 5 and shown in orange, while His165 and Tyr135 are represented in magenta; all three residues were mutated for further studies. Figure generated with MOE.[31]

Data shows that the water permeability of all mutants remains at least equal, and in some cases even slightly higher than WT above pH 7, thus indicating that the hAQP7 mutants are functional water channels (Figure 5.5C and E). In regard to glycerol permeation; while the single H140A mutant displays similar behaviour to WT, all the other mutations show a notable reduction in glycerol permeability at pH 7.4 (Figure 5.5D and F). The mutants H165A and H165A+H140A reduced glycerol maximal permeability at pH 7.4 to 44% that of the WT, inducing a shift in the pK_a from 5.85 ± 0.01 to 6.30 ± 0.02. Interestingly, the Y135A and Y135A+H165A mutations rendered the protein relatively inactive, reducing the maximal P_{Gly} to 10% of that observed for WT (Figure 5.5D and F). These data suggest a key involvement of Tyr135 and His165 residues in the channel pore permeability, whilst His140 alone appears to have no contribution to pH-dependency.

5.2.3 hAQP7 mutagenesis homology model analysis

Following up on the *in vitro* results, showing the involvement of Tyr135 and His165 in hAQP7 permeability, a detailed analysis of the WT MD trajectories was then

performed, but a clear role for these residues in pH-gating was not evident. However, careful analysis of the H-bond network during permeation at pH 7 shows how glycerol preferentially forms hydrogen bonds with residues featuring protonatable side chains (e.g. Lys63, Tyr64). Protonation of amino acid residues near the extracellular pore entrance (such as His165) at pH 5 may alter the H-bond network, and therefore affect the residues responsible for glycerol passage. Based on these considerations, we proposed that the contribution of His165 and Tyr135 to AQP7 pore permeability is based on a hydrogen bond network which dynamically readapts to variations in external acidification. To ascertain if our hypothesis was correct, new homology models for all the mutants were prepared. The models were generated based on the mutated sequences, refined and protonated at both pH 5 and pH 7, as previously performed for the WT model. Each model was then analysed for pore size (Figure 5.3B).

The largest size difference was found for the Y135A+H165A double mutant (Figure 5.7A), which showed a decrease in pore size at the ar/R SF of approximately 1 Å in comparison to the WT. All the mutants showed a decrease in pore size in the ar/R SF region (Figure 5.3B), although not as marked (ca. 0.5 Å) as found in the double mutant. None of the mutants showed relevant changes in size around the NPA motif. A decrease in the size of the ar/R SF, the first selectivity filter encountered by solutes when permeating AQPs from the extracellular side, may contribute to the loss of permeability observed for the mutants. However, in the case of His140A, which also shows a slight decrease in ar/R SF, whilst having the same permeability of the WT. This implies pore size may not be the only factor affecting changes in AQP7 permeability. In fact, Figure 5.4 and Table 5.1 show the studied mutants also appear to alter the pK_a of water and glycerol conductance. The effect could be due to alterations in the electrostatic surface of the mutants and also induced by different pH conditions. The electric fields within protein pores have previously been identified as barriers to solute passage.[40] In order to investigate the effects of pH and mutations on the electrostatic surface of hAQP7, surfaces for each were generated using the Adaptive Poisson-Boltzmann Solver (APBS)[41] plugin in Chimera[42], and are shown in Figure 5.7B and 5.8.

The main observable difference we see is that the intracellular electrostatic surface of all the models appear to be much more negatively charged (shown in red colour) at pH 7 than at pH 5 (Figure 5.7B). At pH 7, the inner channel areas appear more positively charged in comparison to pH 5. Interestingly, the H140A mutant appears to be very similar to WT, both in surface electrostatic distribution at both pH values and in terms of protonation (Figure 5.8), which matches with the experimental findings.

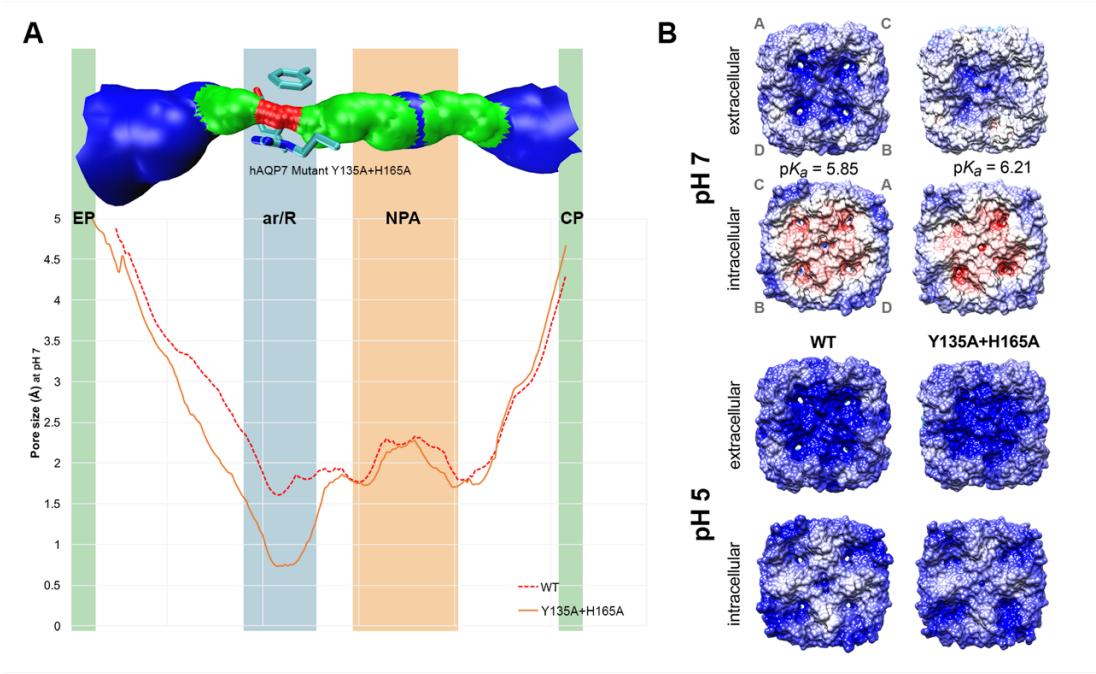


Figure 5.7 Pore size and protein surface electrostatic charges. **A) –** Pore size of human AQP7-WT and double mutant Y135A+H165A, analysed using HOLE[34]. Top channel surface represents the pore size of hAQP7 mutant Y135A+H165A (based on VDW radii): red = smaller than single H_2O, green = single H_2O, blue = larger than single H_2O. Figure generated by HOLE[34] and VMD[35]. **B)** Electrostatic surfaces of hAQP7-WT and mutant Y135A+H165A, at pH 5 and 7, with the following colour code: positively charged = blue, negatively charged = red, neutral = white. Top views of both intra and extracellular sides, with the pK_a of each protein indicated in the figure. Surfaces were generated using the Adaptive Poisson-Boltzmann Solver (APBS)[41] plugin in Chimera[42]. Monomers labelled A-D are shown in grey to indicate their position in each view.

The most marked changes were observed for the H165A and Y135A single and double mutants, indicating to the crucial role of these two residues in pH gating and glycerol permeability. Analysis of protonation at both pH 5 and 7 reveals the mutations do not induce protonation changes in the intracellular residues Glu40 and Glu202. At pH 5, the protonatable intracellular site His92 is deprotonated in the H165A and H140A+H165A mutants, with the H140A mutant partially deprotonated. Protonation changes of this residue do not appear to have any affect the permeability at pH 5 for

any of the mutants, implying that deprotonation of His92 is not involved in the permeation mechanism. Although none of the mutants shows protonation changes of intracellular residues at pH 7, we do see that the intracellular charge distribution is altered in Y135A, H165A and double mutants: the electrostatic surface of the channel appears to be more negative than WT. This observation supports the idea that changes to the extracellular residues may affect the overall electrostatic surface of the protein, including the intracellular surface. Therefore, changes in extracellular pH may also affect the intracellular charge distribution, possibly by altering the H-bond network of the protein and thus, influencing its permeability.

It has also been shown that the pK_a differences appear to be due to the alterations of electrostatic protein surface, induced by the mutations, which are able to alter the protein's affinity for substrates. Mutating these residues does not appear to render the protein pH-insensitive. One possible scenario is that the protein rearranges its H-bonding network based on the newly inserted residues and redistributes the electrostatic potential thus, changing its affinity for substrates, shifting the role of pH-sensitivity to nearby residues. The effect of electrostatics on the transport modulation is quite typical of biological channels and has been previously observed for different porins.[40]

Table 5.1 pH-dependency (pK_a values) of water and glycerol permeation of human AQP7 (wild-type and mutated). Data presented as mean ± SEM (n=4). *p < 0.05 vs wild-type.

	Water	*Glycerol*
	pK_a	pK_a
Wild-Type	5.89 ± 0.02	5.85 ± 0.01
Y135A	6.06 ± 0.01	6.04 ± 0.01
H140A	6.11 ± 0.02	5.91 ± 0.01
H165A	6.65 ± 0.01*	6.30 ± 0.02*
Y135A+H165A	6.58 ± 0.01*	6.21 ± 0.02*
H140A+H165A	6.53 ± 0.01*	6.29 ± 0.01*

Values were compared to AQP7 wild type. *, $P<0.05$

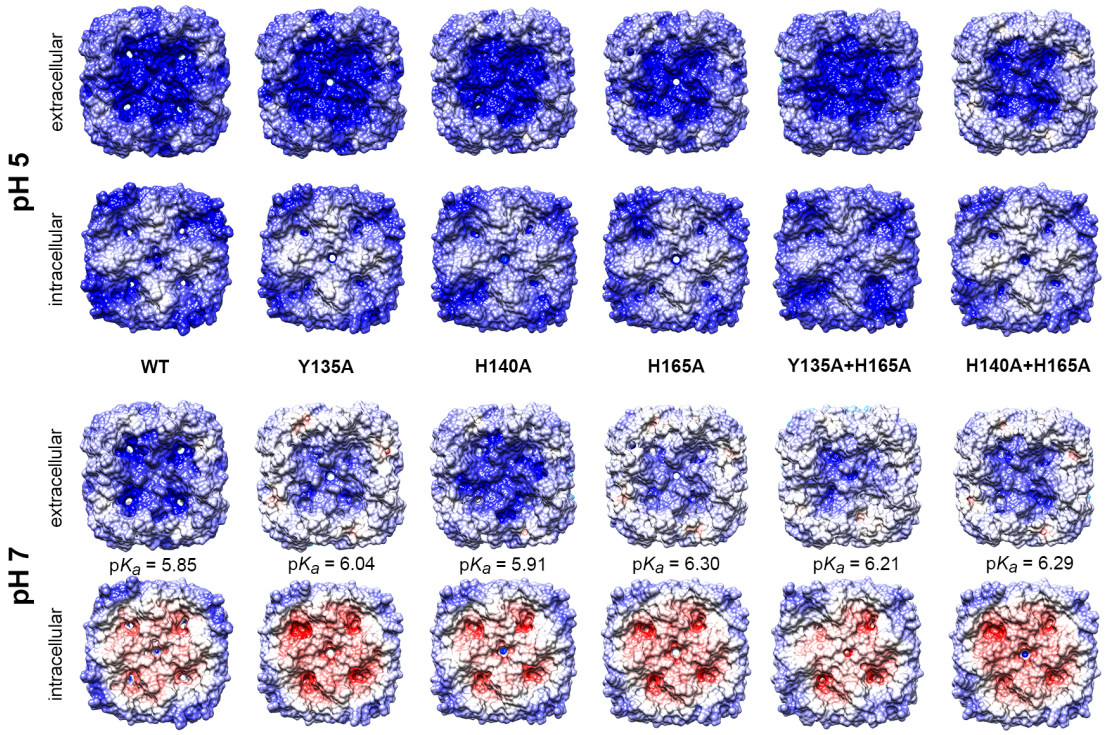


Figure 5.8 Electrostatic surfaces of hAQP7 WT and mutants, at pH 5 and 7. Protonation performed using PROPKA 3.1 package[33]. Surfaces were generated using the Adaptive Poisson-Boltzmann Solver (APBS)[41] plugin in Chimera[42].

5.3 Conclusions

Aquaglyceroporins play essential roles in energy metabolism and adipose tissue homeostasis which may have implications in obesity and metabolic-related complications, such as the metabolic syndrome. The main glycerol channel found to be expressed in human adipose tissues is AQP7, which has a fundamental role in glycerol release from adipocytes during lipolysis, in which triglycerides are hydrolysed to fatty acids. During fasting, intracellularly produced glycerol is transported by AQP7 into the bloodstream, where it is then taken up in the liver, *via* AQP9, where it can then be used for gluconeogenesis. Therefore, the fine-tuning of AQP7 efficiency for glycerol release from the adipose tissue and its subsequent uptake by the liver appears to be the rate-limiting step for the continuation of normal blood glucose levels. Previously AQP7 pH-sensitivity has been assessed, yet the molecular basis of pH-regulation has never been elucidated.

In this study we disclosed the mechanism of pH regulation of hAQP7 activity at a molecular level, showing how this channel can change from a fully active to almost inactive state when pH drops from 7.50 to 5.00. By using a combination of *in vitro* and *in silico* approaches we ascertained that residues Tyr135 and His165, which face into the extracellular environment, are central to channel permeability. We therefore, hypothesize that hAQP7's lack of glycerol permeability, at pH 5, may be due to changes in the overall H-bond network which affects the affinity of the substrates for the channel rather than a truly open/closed pore mechanism of gating. Finally, due to the positions of key pH sensitive residues within the monomer-monomer interface, it is possible that the change in permeability is due to monomer-monomer cooperativity, as observed for other aquaporins.[43,44]

5.4 Methodology Section

The 3D structure of hAQP7 was obtained by homology modelling using Molecular Operating Environment (MOE 2012.10) (CCG 2012)[31]. The choice of a template structure was based on the sequence identity between hAQP7 and the sequence of the AQPs with available resolved structures from human, bacteria and Plasmodium falciparum (UniProt 2013 codes O14520, C8TK05 and Q8WPZ6, respectively). The isoform that has the highest sequence similarity (67%) with hAQP7 is the bacterial isoform Glycerol Facilitator (GlpF), which was then chosen as a template structure. Three resolved structures for bGlpF, crystalized both with or without glycerol and solved by X-Ray diffraction, were accessed through the Protein Data Bank[45]. The template was selected which had the best resolution (2.2 Å, pdb 1FX8). The biological assembly, containing the tetramer structure, was prepared using MOE and protonated at pH 7 under forcefield Amber12EHT. Thus, the tetrameric form of the human AQP7 model was built: 50 intermediate models were generated and averaged to obtain the final homology model.

The obtained model was checked for more realistic rotamers of side chains in the regions of ar/R SF and NPA, by comparison with the available crystal structures of the other available AQP isoforms (pdb codes 1H6I, 36D8, 3D9S, 1RC2, 1LD1 and

3C02). The structure was protonated at pH 7 and an energy minimization refinement was performed, also under the Amber12EHT force field, during which the Cα atoms were fixed. After identification of the residues of interest for the mechanism of pH gating, the same energy minimization procedure was used to further refine them. The refined homology model of hAQP7 was protonated using the PROPKA 3.1 package[33] at pH 5.0 and 7.0. Electrostatic surfaces were generated using the Adaptive Poisson-Boltzmann Solver (APBS)[41] plugin in Chimera[42].

Two model systems, using the above-described homology model, were built using the PDB2PQR Server (version 2.0.0)[33], protonated at pH 7 and pH 5. Each molecular system consisting of the protonated tetrameric models of the AQP7 embedded in a double layer of 175 palmitoyl-oleoyl-phosphatidyl-coline (POPC) lipid. This was achieved using the charm-gui online server[46]. Four glycerol molecules were placed into the system, one above each pore entrance on the extracellular side, at an approximate distance of 30 to 35 Å. To evaluate glycerol uptake, the system was solvated with 33151 (pH 7) and 33098 (pH 5) TIP3P water molecules and used a modified amber99sb-ildn forcefield, with the parameters for glycerol generated by the Automated Topology Builder and Repository (ATb, version 2.2) website using the B3LYP/6-31G* basis set[47].

All simulations were run using the GROMACS 5.1.2 simulation software[48]. Particle-mesh Ewald method was used for calculating electrostatic interactions. The Verlet cut-off scheme with a cut-off distance of 1.4 nm was used for short-range repulsive and attractive interactions and Lincs was used to constrain all bond lengths. Nose-Hoover temperature coupling was used to maintain the temperature of the system (τ = 0.5 ps) at 310 K. The Parrinello-Rahman algorithm was used to maintain the pressure of the system at 1 bar with a coupling constant of τ = 1.0 ps. Simulations were equilibrated for 100 ps before production.

The four individual glycerol molecules were defined in the index and coupled in the pull code (e.g. gly_1 to chain_A). A total of 10 MD simulations (5 for the pH 7 system and 5 for the pH 5 system) were run using the direction COM pull procedure, in each case applying a separate yet equal harmonic restraint force to each glycerol molecule of 100 kJ mol-1nm^2 with a rate of 0.02 nm ns^{-1} along the z-axis in the direction of the

intracellular space. Simulations were run for 250 000 steps or 500 ps, with a timestep of 2 fs. Each pore radius was calculated using the Hole 2.0 program[34], which determines the internal surface based on atomic van der Waals radii. Snapshots were taken at 5 equal intervals throughout the trajectories, and coordinates for the centre of each pore (monomers A to D), at the ar/R SF, were used to generate the pore radius along the z-axis, for each simulation.

To test water permeation six MD simulations were performed, three for pH 7 and three for pH 5. Simulations were run for 10 000 000 steps or 20 000 ps (20 ns), with a time-step of 2 fs, using the same two model systems and parameters. In these runs, the pull code was omitted, therefore removing any biasing of the system. Water molecules were counted using a python script, based on a Tcl script for tracking water molecules within a simulation[36]. The script calculated each water molecule's position, based on the oxygen atom, at each step and tracked its progress over the simulation time, registering each completed passage through the tetramer as well as its direction of movement. The upper and lower limits on the tetramer height were taken from the pore radius data calculated using the HOLE 2.0 program[34]. Visualization of the trajectories and Hole 2.0 pore surfaces and sizes was performed using VMD[35].

Glycerol movement across hAQP7 was monitored by analysing its trajectories for the whole duration of the simulations. For each simulation, glycerol molecules able to completely cross the hAQP7 channels were accounted for and all others excluded. The final data is represented as the mean ± SEM of total glycerol molecules crossing the aquaporin channels.

The diffusion constant (D_w) data results for the molecular dynamics water flux calculations were expressed as mean ± SEM of *n* individual experiments. Statistical analysis between groups was performed by the unpaired Student's t-test using the Prism software (GraphPad Software Inc., San Diego, CA). *P* values < 0.05 were considered statistically significant.

5.5 References

1. Ishibashi, K., Hara, S. & Kondo, S. Aquaporin water channels in mammals. *Clin. Exp. Nephrol.* **13**, 107–117 (2009).
2. Hara-Chikuma, M. & Verkman, A. S. Physiological roles of glycerol-transporting aquaporins: The aquaglyceroporins. *Cell. Mol. Life Sci.* **63**, 1386–1392 (2006).
3. Rodríguez, A., Catalán, V., Gómez-Ambrosi, J. & Frühbeck, G. Aquaglyceroporins serve as metabolic gateways in adiposity and insulin resistance control. *Cell Cycle* **10**, 1548–1556 (2011).
4. Madeira, A., Moura, T. F. & Soveral, G. Aquaglyceroporins: Implications in adipose biology and obesity. *Cell. Mol. Life Sci.* **72**, 759–771 (2015).
5. da Silva, I. V., Rodrigues, J. S., Rebelo, I., Miranda, J. P. G. & Soveral, G. Revisiting the metabolic syndrome: the emerging role of aquaglyceroporins. *Cell. Mol. Life Sci.* **75**, 1973–1988 (2018).
6. da Silva, I. V. & Soveral, G. Aquaporins in Obesity. *Adv. Exp. Med. Biol.* **969**, 227–238 (2017).
7. Törnroth-Horsefield, S., Hedfalk, K., Fischer, G., Lindkvist-Petersson, K. & Neutze, R. Structural insights into eukaryotic aquaporin regulation. *FEBS Lett.* **584**, 2580–2588 (2010).
8. Kitchen, P. *et al.* Plasma membrane abundance of human aquaporin 5 is dynamically regulated by multiple pathways. *PLoS One* **10**, 1–17 (2015).
9. Madeira, A., Moura, T. F. & Soveral, G. Detecting Aquaporin Function and Regulation. *Front. Chem.* **4**, 3 (2016).
10. Janosi, L. & Ceccarelli, M. The Gating Mechanism of the Human Aquaporin 5 Revealed by Molecular Dynamics Simulations. *PLoS One* **8**, (2013).
11. Rodrigues, C. *et al.* Rat Aquaporin-5 Is pH-Gated Induced by Phosphorylation and Is Implicated in Oxidative Stress. *Int. J. Mol. Sci.* **17**, 2090 (2016).
12. Tournaire-Roux, C. *et al.* Cytosolic pH regulates root water transport during anoxic stress through gating of aquaporins. *Nature* **425**, 393 (2003).
13. Leitão, L., Prista, C., Moura, T. F., Loureiro-Dias, M. C. & Soveral, G. Grapevine aquaporins: Gating of a tonoplast intrinsic protein (TIP2;1) by cytosolic pH. *PLoS One* **7**, (2012).
14. Verdoucq, L., Grondin, A. & Maurel, C. Structure–function analysis of plant

aquaporin *At* PIP2;1 gating by divalent cations and protons. *Biochem. J.* **415**, 409–416 (2008).

15. Németh-Cahalan, K. L. & Hall, J. E. pH and calcium regulate the water permeability of aquaporin 0. *J. Biol. Chem.* **275**, 6777–6782 (2000).
16. Soveral, G., Madeira, A., Loureiro-Dias, M. C. & Moura, T. F. Membrane tension regulates water transport in yeast. *Biochim. Biophys. Acta - Biomembr.* **1778**, 2573–2579 (2008).
17. Fischer, G. *et al.* Crystal structure of a yeast aquaporin at 1.15 angstrom reveals a novel gating mechanism. *PLoS Biol.* **7**, e1000130–e1000130 (2009).
18. Ozu, M., Dorr, R. A., Gutiérrez, F., Politi, M. T. & Toriano, R. Human AQP1 is a constitutively open channel that closes by a membrane-tension-mediated mechanism. *Biophys. J.* **104**, 85–95 (2013).
19. Leitão, L., Prista, C., Loureiro-Dias, M. C., Moura, T. F. & Soveral, G. The grapevine tonoplast aquaporin TIP2;1 is a pressure gated water channel. *Biochem. Biophys. Res. Commun.* **450**, 289–294 (2014).
20. Verkman, A. S., Anderson, M. O. & Papadopoulos, M. C. Aquaporins: Important but elusive drug targets. *Nature Reviews Drug Discovery* **13**, 259–277 (2014).
21. Soveral, G., Nielsen, S. & Casini, A. *Aquaporins in Health and Disease: New Molecular Targets for Drug Discovery*. (CRC Press, Taylor & Francis Group, 2017).
22. Soveral, G. & Casini, A. Aquaporin modulators: a patent review (2010–2015). *Expert Opin. Ther. Pat.* **27**, 49–62 (2017).
23. Zeuthen, T. & Klaerke, D. A. Transport of water and glycerol in aquaporin 3 is gated by H+. *J. Biol. Chem.* **274**, 21631–21636 (1999).
24. Zelenina, M., Bondar, A. A., Zelenin, S. & Aperia, A. Nickel and extracellular acidification inhibit the water permeability of human aquaporin-3 in lung epithelial cells. *J. Biol. Chem.* **278**, 30037–30043 (2003).
25. Zelenina, M., Tritto, S., Bondar, A. A., Zelenin, S. & Aperia, A. Copper inhibits the water and glycerol permeability of aquaporin-3. *J. Biol. Chem.* **279**, 51939–51943 (2004).
26. De Almeida, A. *et al.* Exploring the gating mechanisms of aquaporin-3: New clues for the design of inhibitors? *Mol. Biosyst.* **12**, 1564–1573 (2016).
27. Rodríguez, A., Catalán, V., Gómez-Ambrosi, J. & Frühbeck, G. Role of

aquaporin-7 in the pathophysiological control of fat accumulation in mice. *FEBS Letters* **580**, 4771–4776 (2006).

28. Martins, A. P. *et al.* Targeting aquaporin function: Potent inhibition of aquaglyceroporin-3 by a gold-based compound. *PLoS One* **7**, (2012).
29. Madeira, A. *et al.* A Gold Coordination Compound as a Chemical Probe to Unravel Aquaporin-7 Function. *Chembiochem.* **15**, 1487–1494 (2014).
30. Mósca, A. *et al.* Molecular Basis of Aquaporin-7 Permeability Regulation by pH. *Cells* **7**, 207 (2018).
31. ULC, C. C. G. Molecular Operating Environment (MOE). (2018).
32. Fu, D. *et al.* Structure of a glycerol-conducting channel and the basis for its selectivity. *Science (80-.).* **290**, 481–486 (2000).
33. Søndergaard, C. R., Olsson, M. H. M., Rostkowski, M. & Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput.* **7**, 2284–2295 (2011).
34. Smart, O. S., Neduvelil, J. G., Wang, X., Wallace, B. A. & Sansom, M. S. P. HOLE: A program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.* **14**, 354–360 (1996).
35. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
36. Wambo, T. O., Rodriguez, R. A. & Chen, L. Y. Computing osmotic permeabilities of aquaporins AQP4, AQP5, and GlpF from near-equilibrium simulations. *Biochim. Biophys. Acta - Biomembr.* **1859**, 1310–1316 (2017).
37. Berezhkovskii, A. & Hummer, G. Single-File Transport of Water Molecules through a Carbon Nanotube. *Phys. Rev. Lett.* **89**, 64503 (2002).
38. Zhu, F., Tajkhorshid, E. & Schulten, K. Collective diffusion model for water permeation through microscopic channels. *Phys. Rev. Lett.* **93**, 1–4 (2004).
39. Horner, A. *et al.* The mobility of single-file water molecules is governed by the number of H-bonds they may form with channel-lining residues. *Sci. Adv.* **1**, 1–6 (2015).
40. Bajaj, H. *et al.* Molecular Basis of filtering carbapenems by porins from βlactam-resistant clinical strains of Escherichia coli. *J. Biol. Chem.* **291**, 2837–2847 (2016).
41. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A.

Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci.* **98**, 10037–10041 (2001).

42. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
43. Vajpai, M., Mukherjee, M. & Sankararamakrishnan, R. Cooperativity in Plant Plasma Membrane Intrinsic Proteins (PIPs): Mechanism of Increased Water Transport in Maize PIP1 Channels in Hetero-tetramers. *Sci. Rep.* **8**, 1–17 (2018).
44. Németh-Cahalan, K. L., Clemens, D. M. & Hall, J. E. Regulation of AQP0 water permeability is enhanced by cooperativity. *J. Gen. Physiol.* **141**, 287 LP – 295 (2013).
45. Tajkhorshid, E. *et al.* Control of the selectivity of the aquaporin water channel family by global orientational tuning. *Science (80-.).* **296**, 525–530 (2002).
46. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of computational chemistry.* **29**, 1859–1865 (2008).
47. Malde, A. K. *et al.* An Automated force field Topology Builder (ATB) and repository: Version 1.0. *J. Chem. Theory Comput.* **7**, 4026–4037 (2011).
48. Abraham, M. J. *et al.* Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
49. Zhu, F., Tajkhorshid, E. & Schulten, K. Collective Diffusion Model for Water Permeation through Microscopic Channels. *Phys. Rev. Lett.* **93**, 224501 (2004).

Chapter 6 - Unveiling the mechanisms of aquaglyceroporin-3 glycerol permeation by metadynamics.

Based on the paper:

Darren Wragg, Andreia de Almeida, Angela Casini and Stefano Leoni. Chemistry – A European Journal, 2019, 25, 8713–8718.

DOI: 10.1002/chem.201902121

AQP7 homology model constructed by Dr Andrea de Almeida.

Abstract: Aquaporins are a family of membrane spanning proteins responsible for the water flux required by all cells. Aquaporin-3 (AQP3), a member of the sub-family aquaglyceroporins, is capable of conducting a number of uncharged solutes glycerol, H_2O_2 and urea, in addition to water. AQP3 has been found to be expressed in a number of cancer cell lines including lung, skin and breast cancer; therefore, it has been postulated that AQP3, and its substrate glycerol, may play a role in tumour growth and metastasis. The mechanisms of water and glycerol conductance *via* AQP3 are of great interest for the development of novel anticancer therapeutics. In this chapter, water and glycerol permeation *via* human AQP3 are described using advanced metadynamics approaches, which allow for both the exploration of the free energies involved in pore permeation, as well as describing the mechanisms of water and glycerol conductance at an atomistic level of detail.

6.1 Introduction

Biological membranes exquisitely control the uptake and efflux of a number of substrates in cells. Water permeation through cell membranes is one of the most essential cellular processes for all organisms; yet, little was known about this process until the first description of membrane water channels by Peter Agre and co-workers, in 1992[1] and named *aquaporins* (AQPs). Since then, AQPs have been found throughout nature and shown to be involved in numerous physiological processes. Furthermore, AQPs have become relevant drug targets for the treatment of a number of diseases,[2,3] and new roles continue to emerge as the knowledgebase grows and more is known about their structure and selectivity towards certain substrates. Increased understanding of the underlying mechanisms of permeation *via* AQPs by water or other small molecules can contribute greatly to the understanding of the molecular mechanisms of certain diseases, especially those related to AQP over-expression and to the development of selective modulators capable of acting as either chemical probes or as possible therapeutics[2,3].

AQPs fall into two subfamilies: the *orthodox aquaporins* which conduct exclusively water, and the *aquaglyceroporins* which conduct glycerol efficiently and generally have lower water conductance.[4] The latter subfamily can also facilitate the transport

of a number of other uncharged solutes including urea and hydrogen peroxide (H_2O_2).[5,6,7] In humans, 13 AQP isoforms (AQP0-AQP12) have been characterised, among which aquaporin-3 (AQP3) is known to allow the transportation of glycerol[8], H_2O_2[5], ammonia[9] and urea[8,9]. AQP3 is present in a number of tissues, such as kidney[8], gut[10] and in the respiratory tract[11], and is also the most abundant aquaporin isoform in skin cells[12]. Additionally, it has been found to be expressed in various cancer types, including lung[11], skin[13] and breast cancer[14]. Therefore, it has been hypothesised that AQP3 and its substrates (glycerol and H_2O_2) may play a crucial role in tumour growth and proliferation[3].

Since the discovery of aquaporins, efforts have been made to characterize the selectivity mechanisms of the different AQP subfamilies and isoforms. Following the elucidation of the first three-dimensional structures of aquaporin-1 (AQP1)[15,16], and of the bacterial glycerol facilitator GlpF[17], a number of valuable computational studies have contributed important insights into the mechanistic dynamics and energetics of both water and glycerol conduction in AQPs[18–20]. These studies have confirmed and extended our knowledge of the sequence-based "hourglass" model[21], which states; an AQP monomer features a pore formed by six transmembrane helices (1-6), connected by five loops (A through E) and two semi-helices meeting at the centre of the structure. Close to the extracellular side, the aromatic/arginine selectivity filter (ar/R SF), the narrowest point of the pore, is responsible for size-exclusion of molecules[22]. Underneath, the pore centre is defined by two highly conserved asparagine-proline-alanine (NPA) motifs contained in the B and E loops and semi-helices responsible for exclusion of charged solutes through the formation of an electrostatic barrier.[23] In this respect, a number of studies have established the important role of this electrostatic field generated by the protein matrix in determining AQP selectivity, especially in the case of proton exclusion, although dehydration penalty is also assumed to play an important role[24–28].

Since the first crystal structure[15] and molecular dynamics (MD) simulations[18] were obtained, water transport through AQPs has been understood to be a single-file mechanism. MD studies have helped elucidate the role of the selectivity filters and their effects on the permeation of water and other small uncharged solutes[18,29]. Size and shape of the pore constrain water orientation and determine the internal water-

water interactions[30]. Further work has been undertaken to better understand the free energies involved in permeation events of a number of AQP isoforms[31], providing insight at an atomistic level into the role played by the selectivity filters in water transport. Recent studies on carbon nanotubes as bio-mimic channel systems, identified an optimal pore size threshold for single-file water transport of 0.8 nm, which significantly affects transport rate and molecular translocation direction[32]. Whilst not as complex as AQPs, carbon nanotubes show the efficiency of single file water transport over bulk water flux upon molecular confinement. Recent biophysical studies suggest that unitary water channel conductance, p_f , of AQPs depends exponentially on the number (N_H) of available hydrogen bond donors and acceptors within the pore[33]. However, different AQPs, despite having the same N_H, show markedly different p_f, which could be explained by the dehydration penalty that water molecules face upon entering the single-file region of the pore[33]. The single file water permeability p_f and the Gibbs activation energy barrier $G^{\ddagger}_t$ for facilitated water transport *via* AQPs are intrinsically linked, and factors other than hydrogen bonding may play a role, such as a positive surface charges at the channel mouth.

Minimal information is available concerning glycerol permeation of human AQPs. An early equilibrium MD study of glycerol-saturated bacterial glycerol facilitator (GlpF) by Schulten and co-workers proposed a mechanism for glycerol conductance[34]. In 2002, the same authors reported on steered-MD simulations of glycerol permeation through the same aquaporin isoform[35], which revealed glycerol-channel hydrogen bonding interactions along with the stereoselective nature of the channel. In 2008, Hub & de Groot studied AQP1 and GlpF selectivity for O_2, CO_2, NH_3, glycerol, urea, and water permeation using classical MD[31]. The main focus of the study was on the description of the key role played by the ar/R site, acting as a filter permeated only by small polar solutes[31].

In this context, the aim of the following work is to further investigate the mechanisms of glycerol permeation by the human AQP3 isoform at an atomistic level, which has, so far, never been addressed and to characterize the potential interplay between water and glycerol molecules during their conductance through the pore. Thus, the homology model of tetrameric hAQP3 was built as previously reported[36–38], based on the crystal structure of GlpF[17] (see methodology section for details).

6.2 Results and Discussion

Initially, the work undertaken Hub & de Groot[31] was used to establish a baseline for our studies on AQP3. Firstly, we calculated the potential of mean force (PMF) for the uptake of glycerol by steered-MD and umbrella sampling (see methodology, Figure 6.1 and Table 6.1).

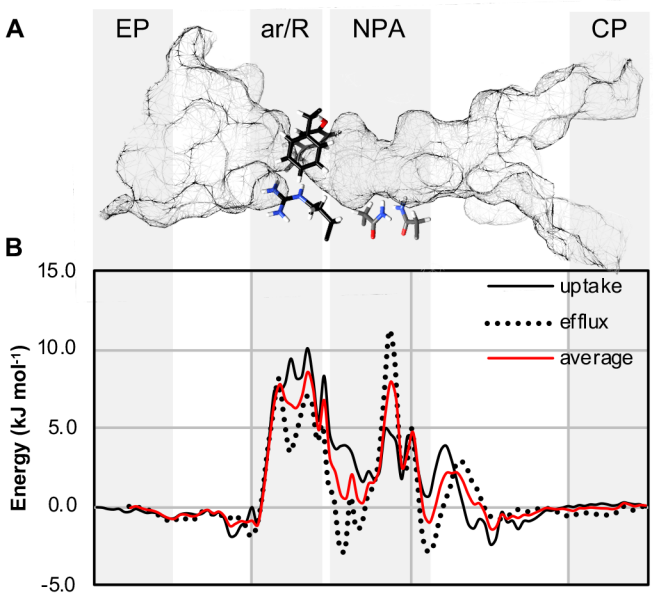


Figure 6.1. A, AQP3 internal pore solvent exclusion surface, indicating the position of the amino acid residues that constitute the ar/R SF and NPA, with amino acids shown in black (ar/R SF) or gray (NPA) (figure generated using Chimera)3. Extracellular (EP) and cytoplasmic (CP) pockets are also highlighted. **B**, Free-energy of glycerol uptake (solid line) and efflux (dashed line), calculated using umbrella sampling. Averaged curve shown in red.

Table 6.1. Average ΔG values for glycerol calculated by umbrella sampling.

	ΔG (kJ/mol^{-1})
	Glycerol
Uptake	16
Efflux	19

The results show the same trends in the energy profile of glycerol conductance through the AQP3 pore, as those reported by Hub & de Groot[31] for GlpF glycerol permeation, including similar free-energy ΔG values (16 kJ/mol^{-1} for AQP3 and 13.5 kJ/mol^{-1} for GlpF[31]). This small difference in ΔG values could be explained by the difference in

minimum pore size observed for AQP3 (2.6± 0.3 Å) and GlpF (3.175 ± 0.002 Å) (Figure 6.2).

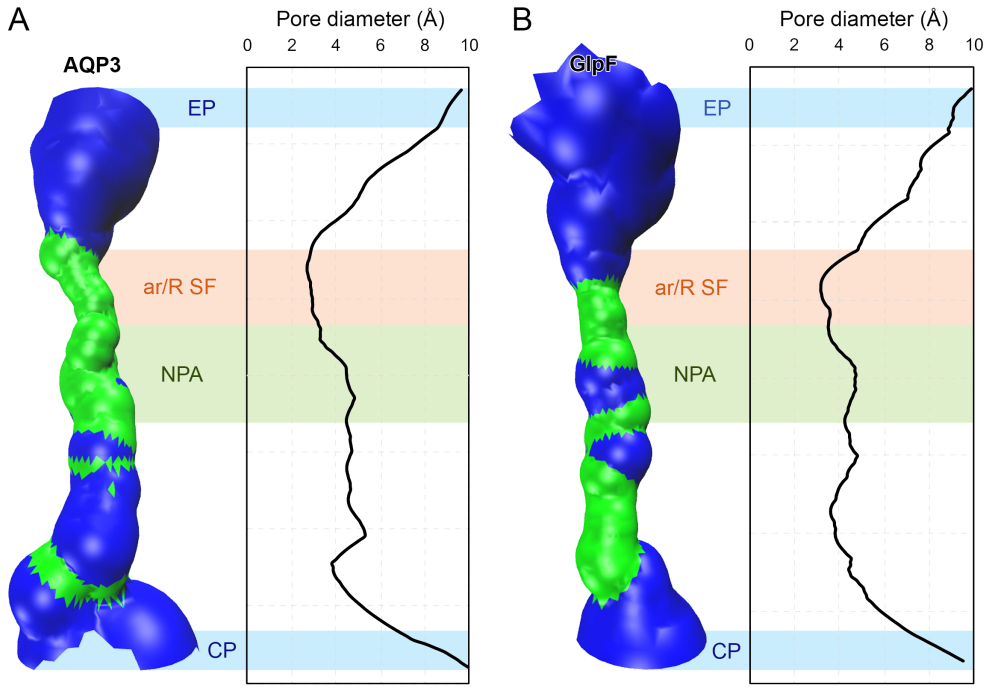


Figure 6.2: Pore size of hAQP3 and GlpF monomers. Representative surface of a monomer and respective pore diameter (Å) of (A) AQP3 and (B) GlpF. Pore surface based on van der Waals radii: Red = smaller than single H_2O, green = single of monomer A. Pore size represented as an average of all monomers. EP –extracellular pocket, ar/R SF – aromatic/arginine selectivity filter, NPA - NPA motif, CP – cytoplasmic pocket. Figure generated by HOLE[39] and VMD[40].

Whilst classical MD calculations provided the free-energy for glycerol permeation *via* AQP3 by collecting the relevant equilibrium configuration probability distributions, they provide no kinetic information of solute conductance, which is required for an unbiased mechanistic analysis over a physiologically significant timeframe. To achieve this, we used metadynamics to reconstruct the free-energy profile for the process of interest from multiple independent runs, each allowing for manifold glycerol and water permeation events *via* the AQP3 pores. Simulation lengths of 200 ns, and the inclusion of 20 glycerol molecules in a single simulation run, provided enough time for conformational changes to be observed within the tetramer during water and frequent glycerol permeation through the pores. Metadynamics accelerates rare event occurrence along selected reaction coordinates, so-called collective variables (CV)[41], and it has been recently successfully applied to calculate the free-

energy surface (FES) for the interactions of organometallic drugs with DNA secondary structures[42,43]. To monitor and encourage both water and glycerol molecules to explore the free energy landscape involved in pore permeation, a CV was defined as the distance between centre-of-mass (COM) of the glycerol molecules and a plane defined by four significant atoms inside the channels (see methodology section for details). Therefore, a large variation of the CV corresponded to successful translocation events, from a which 2D FES could be obtained (integrated[41])for each substrate under investigation. The FES highlights the energy expenditure during a permeation event, providing a highly detailed energy profile as the molecule crosses the pore (Figure 6.3), allowing the matching of local interactions to energy barriers involved in solute conductance. As there are no directional constraints on any of the solvent molecules and periodic boundary conditions (PBC) are applied in all directions, both uptake and efflux permeation can be observed independently.

We obtained free-energy profiles for water and glycerol, for both uptake and efflux pathways, through each AQP3 pore. Figure 6.3 shows, despite the ar/R being the narrowest section of the pore, our simulations show the electrostatic NPA motif provides the highest energy barrier for permeation for both water and glycerol. Whilst the ar/R SF constitutes a region of steric hindrance, requiring a molecule to be below a certain size and in a particular orientation to be able to pass through, it appears to be relatively low in energy demand when compared to the NPA region (Figure 6.3). The latter is the area of the pore where the highest ΔG values are reached for both water and glycerol substrates, namely ($\Delta G_{NPA\text{-}water} \approx 26$ kJ/mol^{-1} and $\Delta G_{NPA\text{-}glycerol} \approx 40$ kJ/mol^{-1}, respectively (Figure 6.3, Table 6.2).

Table 6.2. Free energies ΔG (kJ/mol^{-1}) for water and glycerol permeation through the NPA filter, for both uptake and efflux, calculated by metadynamics. Data shown as mean ± SEM. *n* = number of simulations. Data are calculated from the absolute ΔG of each successful permeation event and averaged.

	ΔG (kJ/mol^{-1})	
	Water	Glycerol
Uptake	26 ± 5 (*n* = 7)	40 ± 4 (*n* = 8)
Efflux	21 ± 5 (*n* = 14)	35 ± 10 (*n* = 10)

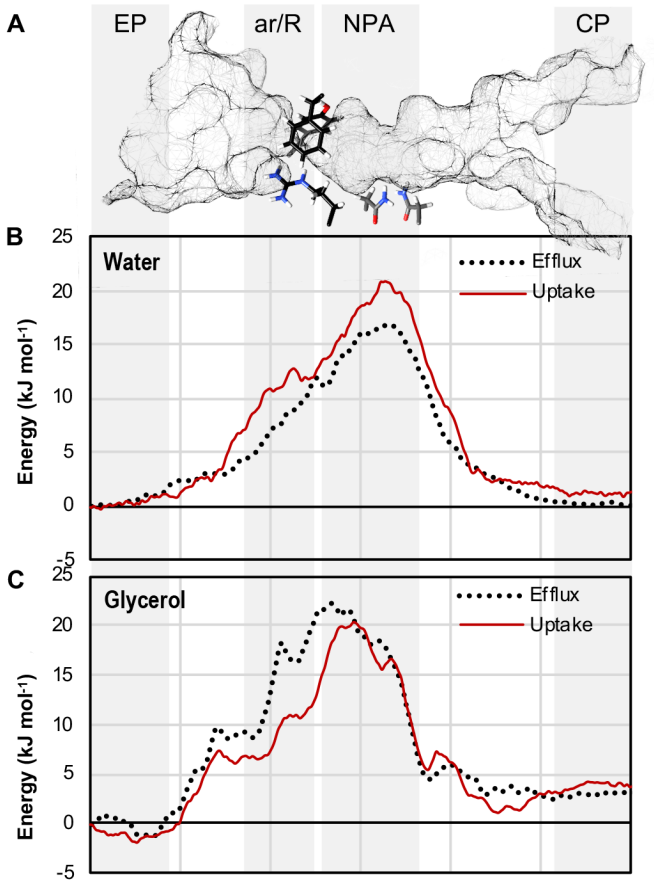


Figure 6.3. A, AQP3 internal pore solvent exclusion surface, indicating the position of the amino acid residues that constitute the ar/R SF and NPA, with amino acids shown in black (ar/R SF) or gray (NPA) (figure generated using Chimera)[44]. Extracellular (EP) and cytoplasmic (CP) pockets are also highlighted. Free energy of water (**B**) and glycerol (**C**) uptake (solid line) and efflux (dashed line). The data represents the averaged FES data from multiple successful permeation events calculated by metadynamics.

The metadynamics trajectories confirm the established flipping motif of water passing though the NPA region of the pore, as observed in other AQPs[15,36], and this helped validate the CV choice. In a total of 2.4 μs of combined simulation time, 30 water and 28 glycerol molecules (out of possible 100 water and 140 glycerol molecules selected in the simulations) were observed to successfully pass through one of the four pores in either direction. However, in the absence of an osmotic gradient, and without directional bias, the number of permeation events was imbalanced between uptake and efflux, and more significantly so in the case of water: efflux events were 80% more successful for water and 60% for glycerol molecules.

Free-energy profiles were obtained for water and glycerol, for both uptake and efflux pathways through the AQP3 pores. The mean ΔG energies for permeation were calculated for each solute molecule in each direction, for single crossing events only.

To remove any bias from multiple permeation events by the same molecule or where alternative paths were taken, the results for these molecules were excluded from the data. Table 6.2 summarizes the overall ΔG values obtained for water and glycerol permeation events of AQP3 (Figure 6.4).

The energy trend for water permeation indicates that the molecule has a lower free-energy profile than glycerol (Table 6.2). Moreover, the free-energies of glycerol permeation in both directions are highly variable, especially efflux (Figure 6.4). This higher variability suggests that the water permeation mechanism has a better-defined pathway than glycerol and is therefore more dependent on the inner-pore interactions during permeation.

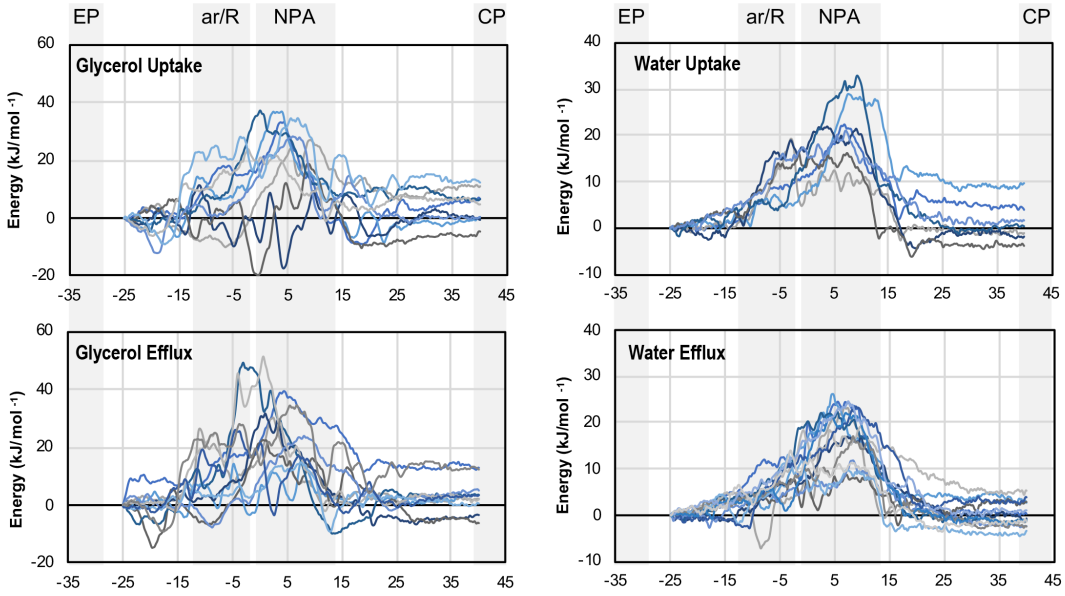


Figure 6.4. Metadynamics calculated free energy surface data overlay of the 20 successful glycerol (left) and 21 successful water (right) uptake and efflux events. ar/R SF, NPA, extracellular (EP) and cytoplasmic (CP) pockets are highlighted in grey. The ΔG of each individual FES was used to calculate the average ΔG value shown in Table 6.2.

In order to understand the molecular pathways, and to identify which amino acid interactions are common to both water and glycerol, the H-bond network for each successful permeation event calculated by metadynamics was analysed (Figures 6.5 and 6.6). The H-bond residence time (RT) for each molecule with different amino acids within the pores was evaluated to map crucial or preferential sites, in either direction of permeation (Figure 6.5).

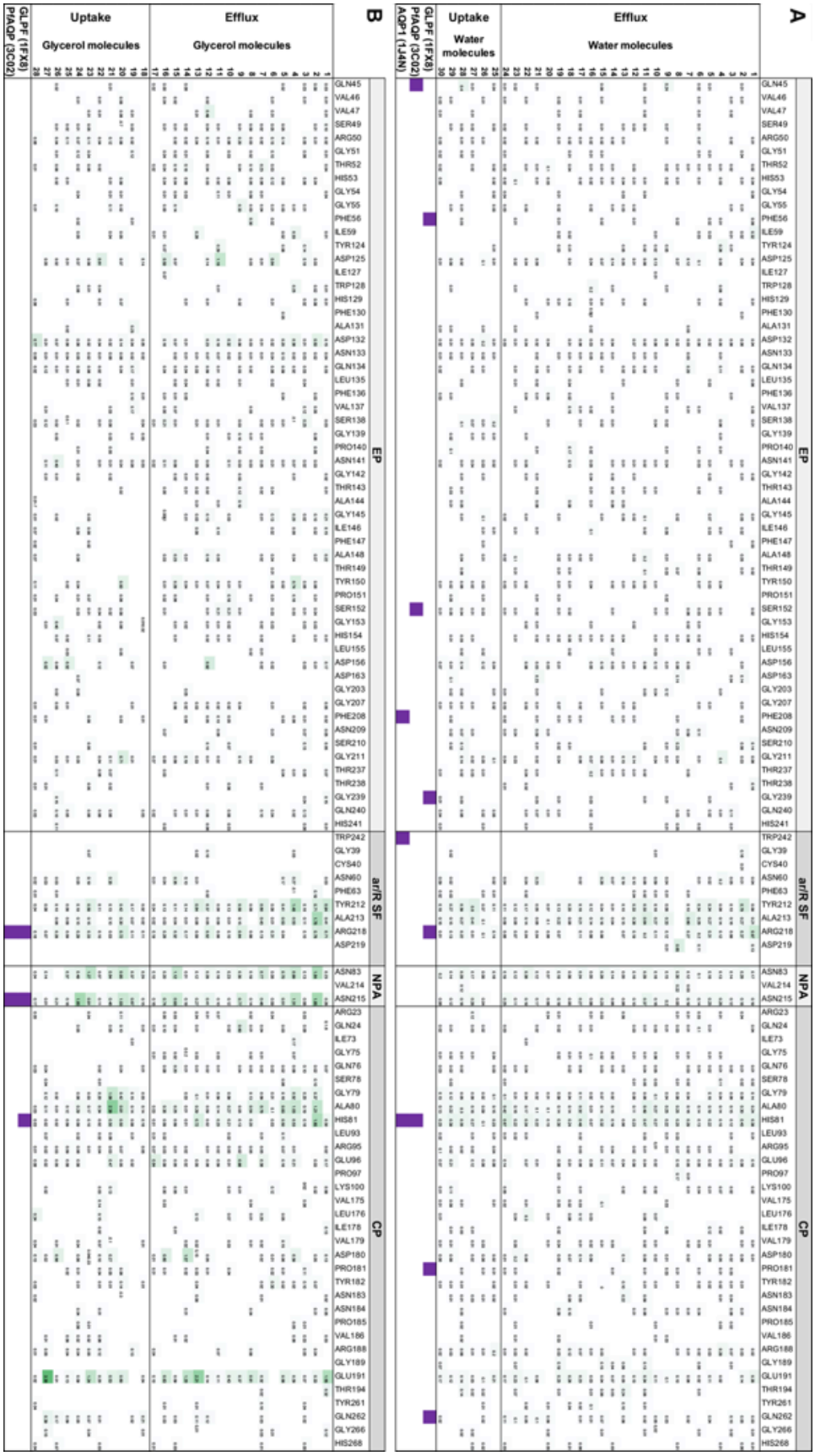


Figure 6.5. Hydrogen bond residence times for each successful permeation event of (A) water and (B) glycerol. Hydrogen bonding for GLPF (pdb 1FX8), PfAQP (pdb 3C02) and AQP1 (pdb 1J4N) with water or glycerol molecules is shown in purple. Colour coding: graded from white to dark green based on residence time in % of time of trajectory spent H-bonding (Simulation time was 200 ns).

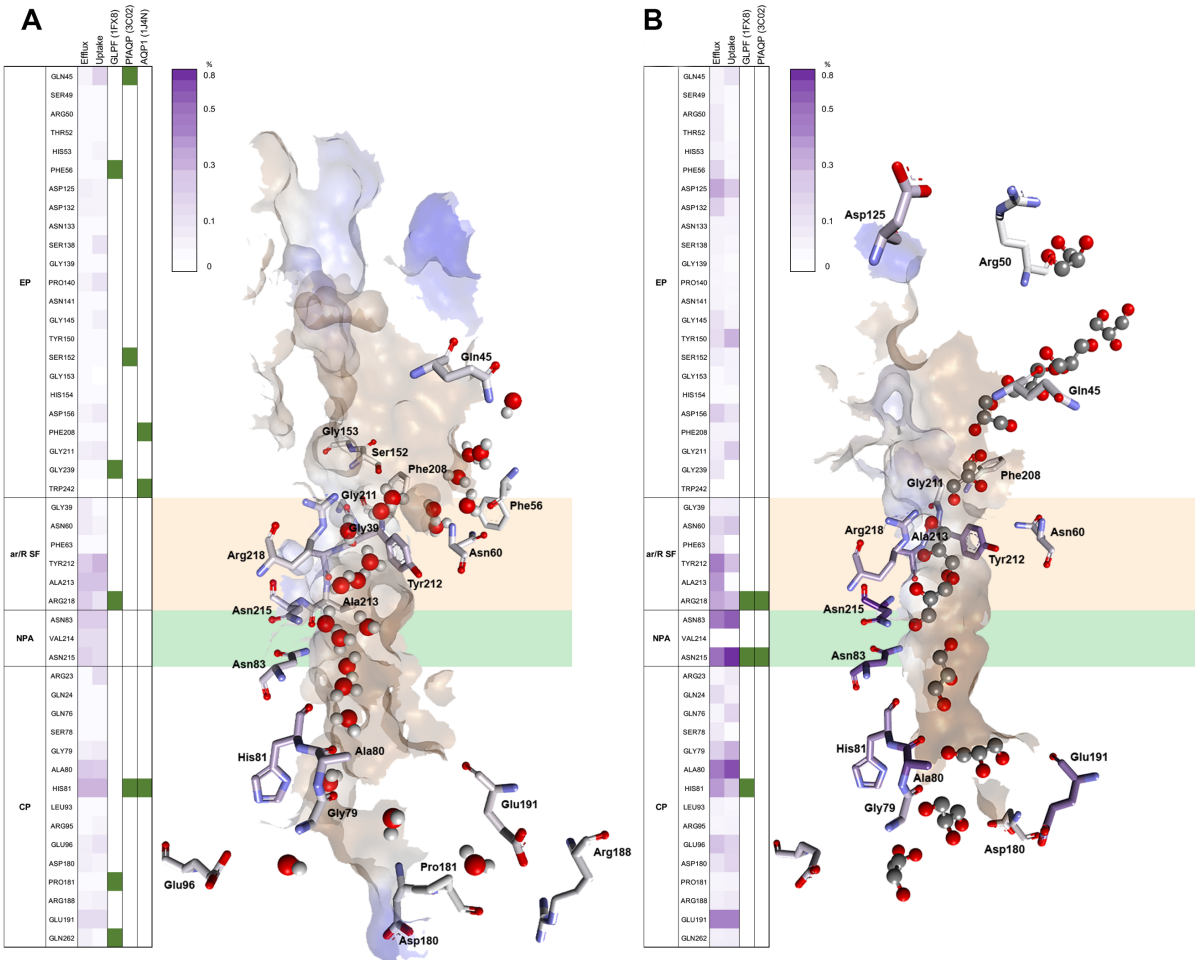


Figure 6.6. Water (**A**) and glycerol (**B**), permeation routes and H-bond patterns and RT (%) from metadynamics calculations. Average H-bond RT (%) of glycerol during uptake and efflux (on a scale of 0-0.8%) shown in a gradient purple colour, with the strongest colour indicating the highest RT. Multiple glycerol and water molecules' snapshots, taken from one representative simulation, are overlaid in one structure to create one single path. Amino acids that form crucial H-bonds are explicitly shown. Carbons are colour-mapped (white to purple) according to their corresponding RT (short to long). Green boxes represent the positions of water and glycerol molecules within the crystal structures of GlpF (pdb1FX8) and PfAQP (pdb 3C02). Water and glycerol molecules are shown in ball and stick representation, with atoms coloured by atom type. Pore colour representation based on hydrophobicity of the pore surface, blue = hydrophilic, brown = hydrophobic. Figure generated with Discovery Studio Visualiser[45].

For water permeation; residence times of water molecules ca. 1.4 ns are observed inside the ar/R SF, where waters form H-bonds mostly with Tyr212, Ala219 and Arg218. Additionally, residence times within the NPA are on average shorter (ca. 0.9 ns). In this region, water molecules interact with Asn83, Val214 and Asn215. In the extracellular pocket (EP), water appears not to bind to any specific residues and has an residence time of ca. 1.6, whereas, in the cytoplasmic pocket (CP), water hydrogen bonds mainly to three amino acids (Ala80, His81, Glu96) and the residence time can

increase up to ca. 2.2 ns (Figure 6.6A). Using the crystallographic structures of GlpF[17] (pdb 1FX8), PfAQP[46] (pdb 3CO2) and AQP1[47] (pdb 1J4N) to pinpoint the positions of water molecules, allowed us to identify possible key H-bonding interactions (Figure 6.6, green highlighted residues) and to match these to the corresponding amino acids in hAQP3. GlpF shows a H-bond with the Arg in the ar/R SF (Arg218 in AQP3), while PfAQP and AQP1 both show H-bonds to the highly conserved His located in the cytoplasmic pocket (His81 in hAQP3). Other H-bonding interactions are more specific to each particular isoform, due to the variation in the amino acid composition of the pores, therefore, there is no corresponding amino acid available for H-bonding in hAQP3.

In the case of glycerol permeation, its size and increased number of available hydrogen bonding groups (3 -OH) directly affects its permeation time, thus, increased H-bonding residence time was observed when compared to water (Figure 6.6). When looking at particular residues inside the pore, the longest residence time involving specific residues (ca. 2.3 ns) is observed for Asn83 and Asn215 within the NPA region, in both directions (Figure 6.6B). A longer residence time after forming H-bonds with these residues implies a higher free energy barrier in the NPA area, as shown in Figure 6.3. Glycerol molecules appear to spend less time in H-bonding with the residues in the ar/R SF (ca. 0.9-1.5 ns), especially during uptake. Compared to water, there are less H-bond interactions seen for glycerol in the structures of GlpF and PfAQP. However, it is possible to identify H-bonds with Arg218, Asn215 and, for GlpF, His81 in the cytoplasmic pocket, which are all also observed in our simulations.

Further analysis shows that the amino acids involved in both glycerol and water H-bonding are similar, in particular residues in the ar/R SF (Tyr212, Ala213, Arg218) and in the NPA (Asn83, Asn215), along with residues in the cytoplasmic pocket (His81, Glu191) (Figure 6.6). For both molecules, the residence time in each H-bond is lower for extracellular pocket than for cytoplasmic pocket (cumulative RT for water; EP = ca.1.6 ns, CP = ca. 2.2 and cumulative RT for glycerol; EP = ca.2.9 ns, CP = ca.4.7 ns), as can be observed in Figure 6.6. The fact that the extracellular pocket is funnel-shaped (see Figure 6.2) having highly flexible loops, narrowing as it approaches the ar/R SF, leads the substrate molecules to spend more time probing the

surface, and thus, spending less time overall interacting with each of the amino acid residues. On the other hand, the cytoplasmic pocket is more cylindrically shaped, and over the majority of its length showing a diameter of ca. 4.5 Å, starting from the NPA region, only broadening in diameter at the very end (Figure 6.2). Confinement below the NPA will lead to increased and more specific interactions of the molecules during permeation, causing both water and glycerol to spend a considerable amount of time in H-bonding with cytoplasmic pocket residues.

As observed in the X-ray structures of GlpF and PfAQP, in our simulations, glycerol molecules permeate the pore as part of a single-file water chain, i.e. there is no direct glycerol-glycerol contact, which would compete with or replace the water-based transport mechanism. This leads to a scenario in which glycerol switches H-bonding back and forth between water molecules within the pore and pore surface residues during permeation. Figure 6.7 illustrates how glycerol alternates H-bonding between the water chain and residues lining the pore surface during both uptake and efflux. When the H-bonding evolution is compared to the FES (Figures 6.3 and 6.4), a clear correlation can be seen between residence time and character of the H-bonding along with characteristic motifs in the free energy surfaces. The longer the glycerol spends in H-bonding to the protein and not interacting with water correlates to a higher the free energy peak. This result highlights the intrinsic nature of the role of a continuous flow of water molecules, into which glycerol is inserted as solute, in both its uptake and efflux conductance though the pore. The process explains the variations we observe in the FES (Figure 6.4) of glycerol permeation; as they are a result of the interplay between solvated glycerol and variable local water configurations and glycerol orientation. The latter comprises of molecular rotations and (in selected trajectories) competition among glycerol molecules in the vestibule regions, which are responsible for sensible deviations in the free energy profiles.

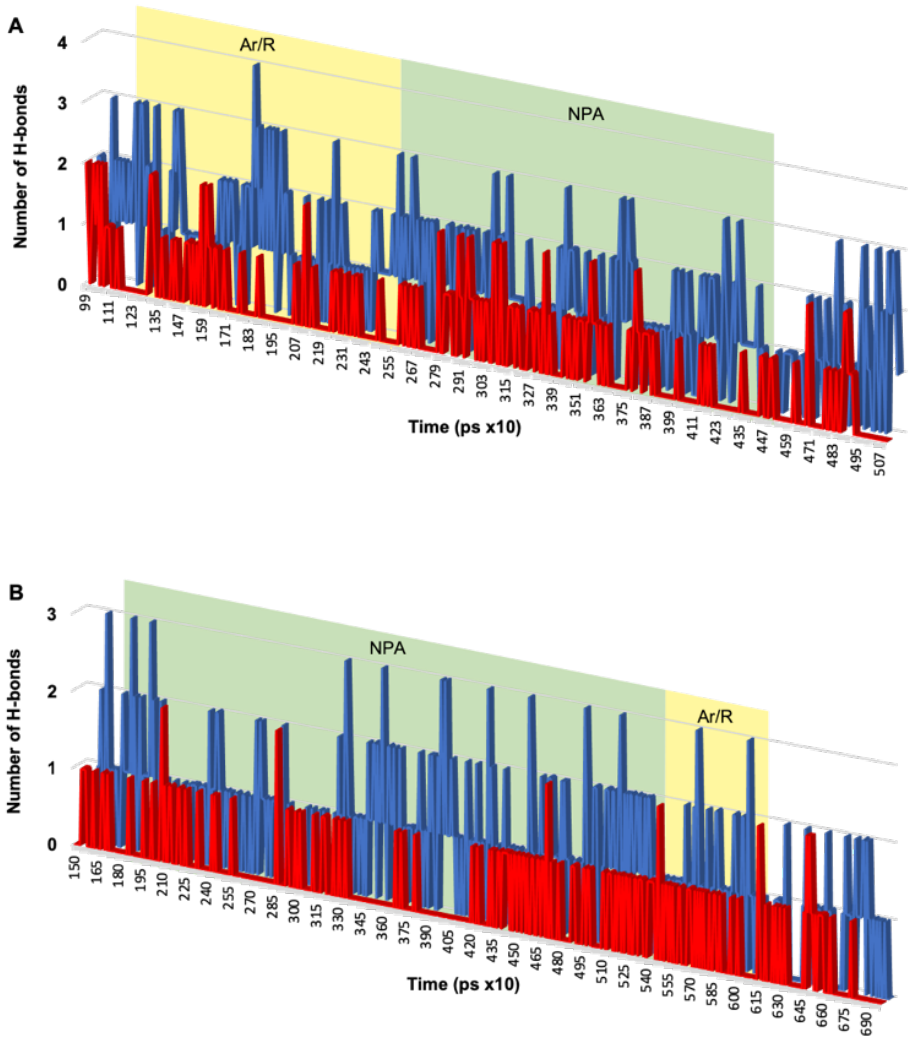


Figure 6.7. Hydrogen bond switching of glycerol with pore residues (red) and water (blue) during (**A**) uptake and (**B**) efflux permeation events

Analysis of the metadynamics simulations show, in the absence of osmotic pressure, water is able to traverse the pore in both directions *simultaneously*, rather than in a single direction at a time, by 'leap-frogging' over each other in the wider region of the pore below the ar/R SF. This means that efflux or uptake results from small perturbations in this base mechanism and does not require a complete inversion in the direction of water flux. Figure 6.8 shows water molecules inside the pore at the beginning of the simulation were free to move in either direction, without file disruptions (Figures 6.8A, B). To this end, water implements a "hopping mechanism", in which single water molecules are able to switch with the molecules in either side, in either direction, along the single file chain. We propose that this process offers, if

not an active regulatory mechanism, at least a structure capable of rapidly responding to environmental changes, as both transport directions are active simultaneously.

We observe that this bidirectional water flux is disrupted as a consequence of glycerol solute molecules entering the pore, inserting into the water chain in the direction of permeation (Figure 6.8C, D). Glycerol insertion prevents water crossing in the opposite direction due to its increased size. As soon as glycerol has traversed the pore back into bulk water, the leap-frog mechanism within the water chain re-establishes itself by filling the pore as the glycerol passes through (Figure 6.8E), i.e. without voids or latencies in the basal water transport mechanism.

As previously reported for glycerol conductance of GlpF[35], the small reduction in energy seen in the FES of glycerol at the extracellular and cytoplasmic vestibules (Figure 6.3) aids glycerol conductance *via* AQP3, increasing the probability of glycerol joining the single file water transport mechanism.

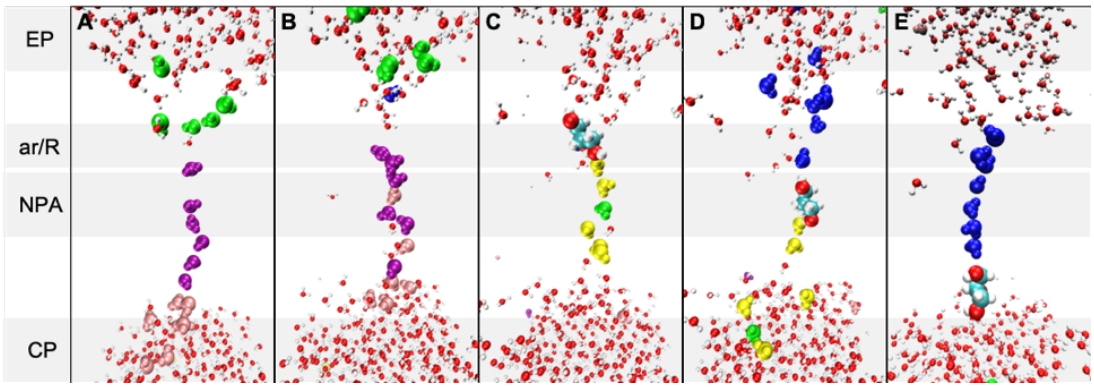


Figure 6.8. Sequence of snapshots showing how the passage of glycerol through the AQP3 pore disrupts the 'leapfrogging' water chain. (A) At the start, water molecules are coloured based on position: purple – located inside the pore, pink - located in the cytoplasmic pocket (CP), green – located in the extracellular pocket (EP). (B) Water molecules start moving from the CP into the channel, replacing the purple molecules. (C) A glycerol molecule enters the ar/R region from the EP, yellow – waters occupying the pore, originating from both EP and IP. (D) Glycerol passes through the NPA region, forcing the waters out of the channel into the CP area, with new water molecules (blue) following on and populating the channel area. (E) The chain of water molecules is re-established as glycerol leaves the pore and the pore is re-populated by single-file water molecules from EP.

6.3 Conclusion

The mechanisms of water and glycerol conductance *via* human AQP3 was investigated using the advanced molecular dynamics approach, metadynamics, where each substrate molecule is encouraged to explore the entirety of the simulation model and find the appropriate conformation to pass through the tetramer in either direction. Over a number of independent simulations, a full mechanistic picture and the underlying free energy could be reliably collected for both water and glycerol. Single-file water permeation *via* AQP3 appears always to be bi-directional at equilibrium conditions, whilst still maintaining a continuous transport flux when compared to bulk water. Moreover, glycerol permeation appears to critically depend on this single file water flux, with transport resulting from bond switches within a dynamic hydrogen bond scaffold created by the interplay of glycerol, water and pore amino acid residues. This reveals a novel scenario, in which solute molecules exploit an existing bidirectional water conduction mechanism in AQP3. Free energy results also suggest a binding affinity between key residues and solute molecules, on the exterior pore surface, facilitating solute transport. Further studies will be necessary to optimize the metadynamics approach to elucidates the transport mechanisms of glycerol transport in non-equilibrium conditions, i.e. mimicking a situation of osmotic imbalance.

6.4 Methodology section

Human AQP3 homology model

The 3D structure of hAQP3 was obtained by homology modelling using the Molecular Operating Environment (MOE 2013.08)[48]. The choice of a template structure was based on the sequence identity between hAQP3 and the sequence of the AQPs with available resolved structures. The isoform showing the highest sequence similarity with hAQP3 is the bacterial isoform GlpF, with 34.68% of sequence identity, which was then chosen as a template structure to generate a homology model of hAQP3. The template was selected among the structures with the best resolution (2.2 Å, pdb 1FX8)[49] The homology model was prepared and refined as described in[36–38].

In-silico simulations

Two different molecular dynamics free-energy methods were chosen to study the mechanism of permeation for glycerol and water: umbrella sampling (US), using a pull code to provide a trajectory along a given reaction coordinate, and metadynamics. The pull code uses a harmonic restraint attached to the centre of mass (COM) of the glycerol, on which a directional force is applied, encouraging the molecule to move in the desired direction through the aquaporin pores.

Each simulation box contained the AQP3 tetramer homology model inserted into a lipid bilayer, made up of 166 POPC lipids (carried out using the charm GUI online server[50,51]). Each system was solvated with TIP3P water molecules, with the number added being dependent on the system. For the US, ca. 52000 water molecules were added while for the metadynamics simulations the number of water molecules was ca. 33000. This was due to the difference in box size10x10x20 nm for US and 10x10x15 nm for metadynamics). A longer box is required for US due to the use of a pull code and the starting and end positions of the glycerol molecules. The metadynamics simulations do not have this requirement, resulting in a 25% shorter box edge, additionally allowing a significant reduction of the simulation calculation time. Periodic boundary conditions (PBC) were used for each system.

Four Na^+ ions were substituted for corresponding water molecules to neutralise the charges of each system. The amber99sb-ildn+slipid force field was used for all the simulations, with the parameters for glycerol generated by the Automated Topology Builder and Repository (ATb, version 2.2) website using the B3LYP/6-31G* basis set [11], using a combination of semi-empirical quantum mechanics (QM) and density functional theory (DFT).

All simulations were run using the GROMACS 2016.3 software[12]. Particle-mesh Ewald method was used for calculating long-range electrostatic interactions, the Verlet cut-off scheme with a cut-off distance of 0.7 nm was used for short range repulsive and attractive interactions and Lincs was used to constrain all bond lengths.

Nose-Hoover[13] temperature coupling was used to maintain the temperature of the system (τ = 0.5 ps) at 300 K. The Parrinello-Rahman[14] algorithm was used to maintain the pressure of the system at 1 bar with a coupling constant of τ = 1.0 ps. Simulations were equilibrated for 100 ps before production. Total energy of the system was monitored using gmx_energy program within Gromacs until total energy changes were less than 0.4% over 75 ps (figure 6.9).

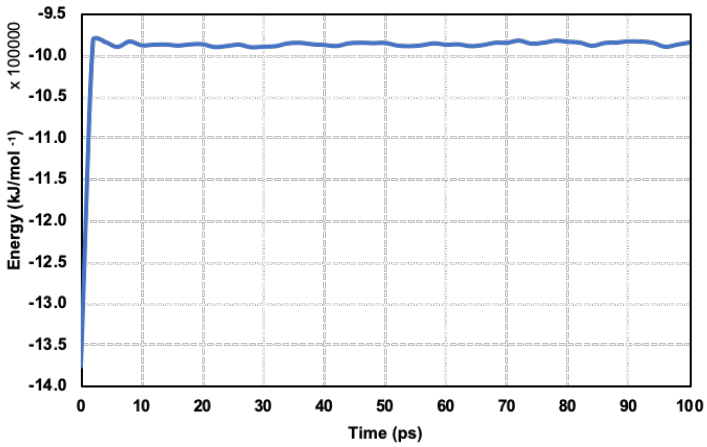


Figure 6.9. Total energy during 100 ps equilibration step for AQP3 with glycerol.

Umbrella sampling

Four individual glycerol molecules were defined in the index and coupled to a monomer in the pull code (e.g. gly_1 to chain_A). Each simulation was run for 0.5 ns, using the direction COM pull procedure, in each case applying a separate yet equal harmonic restraint force to each solute molecule of 600 kJ mol^{-1} nm^2 with a rate of 0.02 nm ns^{-1} along the z-axis. Simulations were carried out with the starting positions of the four glycerol molecules either on the extracellular or intracellular sides of the aquaporin tetramer, run separately in order to simulate both uptake and efflux.

The pull code provides a quick but not physiologically relevant pathway over a short time scale of 500 ps, which is too fast for sidechain movement and therefore provides a less accurate mechanism of permeation. Based on this initial trajectory, umbrella sampling (US) was run to map the free energy on a physiologically relevant time scale (500 windows, 250000 steps, 2 fs time scale: 62.5 ns simulation). Two different US runs were performed, based on different efflux and uptake COM pull reaction coordinates (RC).

From each trajectory, US was carried out on one glycerol molecule in each direction. Each trajectory was split into 500 sampling windows with each window undergoing a further 500 ps simulation, with the glycerol molecule restrained within each window of configurational space. The potential of mean force (PMF) for each was analysed using the weighted histogram analysis method (WHAM) (Figure 6.1).

Metadynamics simulations.

Well-tempered metadynamics simulations were run for 200,000,000 steps with a 1 fs time-step (200ns) using the Plumed plugin[15] for Gromacs. The distance collective variable (CV) was used between the COM of glycerol/water molecules and a plane formed by the backbone carbons of Asparagine 215 from each of the four monomers. This provided a reference plane in the centre of the system, equating to the top of the NPA motif. Each of the 20 glycerol/water molecules were given a separate plumed control within the same data file, allowing the free-energy for each molecule to be calculated within the same simulation. Gaussians were added every 2000 steps (2ps), giving a deposition rate of 0.025 kJ/mol.ps. Gaussian height was 50 TAU (TAU sets the height of the Gaussian equal to $k_B \Delta T$ *x* pace *x* timestep/TAU) and the Gaussian width was 0.025 Å. The bias factor was set to 12, ΔT was 3600 K.

Free energy surface data (Fig 10) were normalised by setting the baseline to 0 and excluding bulk water on either side of the pore to show more clearly the data for each permeation event (Fig 4). The individual FES plots for water and glycerol molecules are shown in Fig 4 and the average FES shown in Figure 6.3 are taken from the trajectories in which a water/glycerol molecule passed through a pore, in either direction, once during the simulation. Data for water/glycerol molecules re-crossing the pore or passing through the lipid membrane were not included. This resulted into 21 FES for water (7 for uptake and 14 for efflux) and 20 FES for glycerol (9 for uptake and 11 for efflux), which were used for calculating ΔG values.

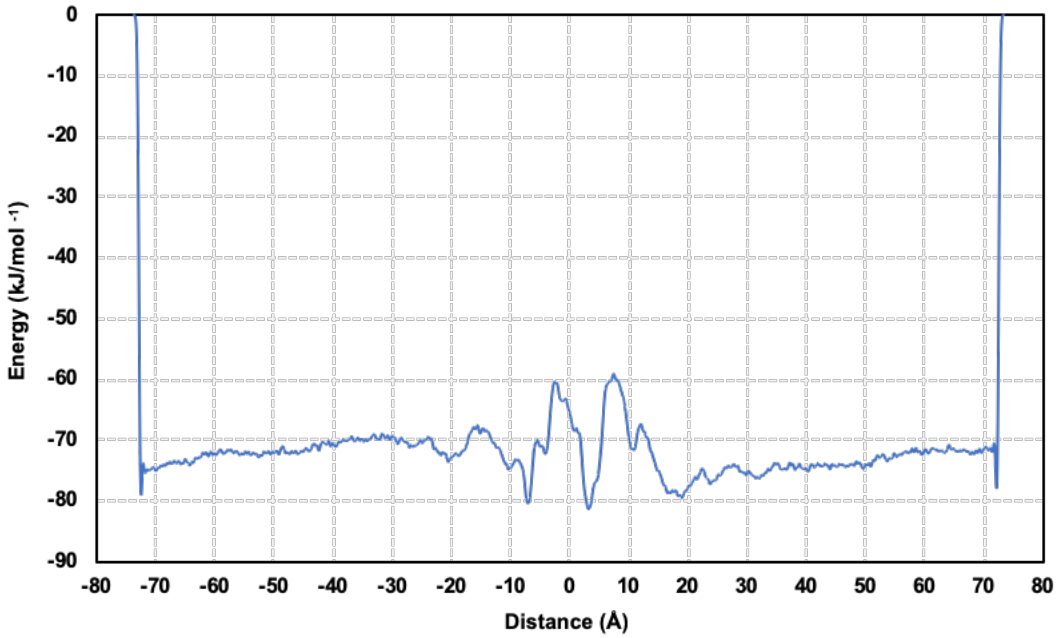


Figure 6.10. Free energy surface data for glycerol and uptake calculated *via* metadynamics.

6.5 References

1 G. M. Preston, T. P. Carroll, W. B. Guggino and P. Agre, *Science*, 1992, **256**, 385–387.

2 A. S. Verkman, M. O. Anderson and M. C. Papadopoulos, *Nat. Rev. Drug Discov.*, 2014, **13**, 259–77.

3 B. Aikman, A. De Almeida, S. M. Meier-Menches and A. Casini, *Metallomics*, 2018, **10**, 696–712.

4 G. Soveral, S. Nielsen and A. Casini, *Aquaporins in Health and Disease: New Molecular Targets for Drug Discovery*, CRC Press, Taylor & Francis Group, 2017.

5 G. P. Bienert and F. Chaumont, *Biochim. Biophys. Acta - Gen. Subj.*, 2014, 1840, 1596–1604.

6 A. Rojek, J. Praetorius, J. Frøkiaer, S. Nielsen and R. a Fenton, *Annu. Rev. Physiol.*, 2008, **70**, 301–327.

7 A. Madeira, S. Fernández-Veledo, M. Camps, A. Zorzano, T. F. Moura, V. Ceperuelo-Mallafré, J. Vendrell and G. Soveral, *Obesity*, 2014, **22**, 2010–2017.

8 K. Ishibashi, S. Sasaki, K. Fushimi, S. Uchida, M. Kuwahara, H. Saito, T. Furukawa, K. Nakajima, Y. Yamaguchi and T. Gojobori, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 6269–6273.

9 T. Litman, R. Sogaard and T. Zeuthen, *Handb. Exp. Pharmacol.*, 2009, 327–358.

10 A. Pelagalli, C. Squillacioti, N. Mirabella and R. Meli, *Int. J. Mol. Sci.*, 2016, **17**.

11 Y. L. Liu, T. Matsuzaki, T. Nakazawa, S. ichi Murata, N. Nakamura, T. Kondo, M. Iwashina, K. Mochizuki, T. Yamane, K. Takata and R. Katoh, *Hum. Pathol.*, 2007, **38**, 171–178.

12 M. Boury-Jamot, J. Daraspe, F. Bonté, E. Perrier, S. Schnebert, M. Dumas, J.-M. Verbavatz, F. Bonte, E. Perrier, S. Schnebert, M. Dumas, J.-M. Verbavatz, F. Bonté, E. Perrier, S. Schnebert, M. Dumas and J.-M. Verbavatz, in *Aquaporins*, Germany, 2009, pp. 205–217.

13 M. Hara-Chikuma and A. S. Verkman, *Mol. Cell. Biol.*, 2008, **28**, 326–332.

14 H. Satooka and M. Hara-Chikuma, *Mol. Cell. Biol.*, 2016, **36**, 1206–1218.

15 K. Murata, K. Mitsuoka, T. Hiral, T. Walz, P. Agre, J. B. Heymann, A. Engel

and Y. Fujiyoshi, *Nature*, 2000, **407**, 599–605.

16 B. L. De Groot, A. Engel and H. Grubmüller, *FEBS Lett.*, 2001, **504**, 206–211.

17 D. Fu, A. Libson, L. J. W. Miercke, C. Weitzman, P. Nollert, J. Krucinski and R. M. Stroud, *Science (80-.).*, 2000, **290**, 481–486.

18 B. L. de Groot and H. Grubmuller, *Science (80-.).*, 2001, **294**, 2353–2357.

19 E. Tajkhorshid, P. Nollert, M. Jensen, L. J. W. Miercke, J. O'Connell, R. M. Stroud and K. Schulten, *Science (80-.).*, 2002, **296**, 525–530.

20 B. L. De Groot and H. Grubmüller, *Curr. Opin. Struct. Biol.*, 2005, **15**, 176–183.

21 Jin Sup Jung, G. M. Preston, B. L. Smith, W. B. Guggino and P. Agre, *J. Biol. Chem.*, 1994, **269**, 14648–14654.

22 D. Krenc, J. Song, A. Almasalmeh, B. Wu and E. Beitz, *Mol. Membr. Biol.*, 2014, **31**, 228–238.

23 N. Chakrabarti, E. Tajkhorshid, B. Roux and R. Pomès, *Structure*, 2004, **12**, 65–74.

24 R. Oliva, G. Calamita, J. M. Thornton and M. Pellegrini-Calace, *Proc. Natl. Acad. Sci.*, 2010, **107**, 4135–4140.

25 N. Chakrabarti, B. Roux and R. Pomès, *J. Mol. Biol.*, 2004, **343**, 493–510.

26 M. Kato, A. V. Pisliakov and A. Warshel, *Proteins Struct. Funct. Genet.*, 2006, **64**, 829–844.

27 H. Chen, B. Ilan, Y. Wu, F. Zhu, K. Schulten and G. A. Voth, *Biophys. J.*, 2007, **92**, 46–60.

28 G. Portella, P. Pohl and B. L. De Groot, *Biophys. J.*, 2007, **92**, 3930–3937.

29 L. Janosi and M. Ceccarelli, *PLoS One*, 2013, **8**.

30 A. Horner, F. Zocher, J. Preiner, N. Ollinger, C. Siligan, S. A. Akimov and P. Pohl, *Sci. Adv.*, 2015, **1**, 1–6.

31 J. S. Hub and B. L. de Groot, *Proc. Natl. Acad. Sci.*, 2008, **105**, 1198–1203.

32 R. H. Tunuguntla, Y. Zhang, R. Y. Henley, Y. C. Yao, T. A. Pham, M. Wanunu and A. Noy, *Science (80-.).*, 2018, **359**, 792–796.

33 A. Horner and P. Pohl, *Faraday Discuss.*, 2018, **209**, 9–33.

34 M. Jensen, E. Tajkhorshid and K. Schulten, *Structure*, 2001, **9**, 1083–1093.

35 M. O. Jensen, S. Park, E. Tajkhorshid and K. Schulten, *Proc. Natl. Acad. Sci.*, 2002, **99**, 6731–6736.

36 A. Spinello, A. De Almeida, A. Casini and G. Barone, *J. Inorg. Biochem.*, 2016,

160, 78–84.

37 A. De Almeida, A. P. Martins, A. F. Mósca, H. J. Wijma, C. Prista, G. Soveral and A. Casini, *Mol. Biosyst.*, 2016, **12**, 1564–1573.

38 A. De Almeida, A. F. Mósca, D. Wragg, M. Wenzel, P. Kavanagh, G. Barone, S. Leoni, G. Soveral and A. Casini, *Chem. Commun.*, 2017, **53**, 3830–3833.

39 O. S. Smart, J. G. Neduvelil, X. Wang, B. A. Wallace and M. S. P. Sansom, *J. Mol. Graph.*, 1996, **14**, 354–360.

40 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33–38.

41 G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, *Comput. Phys. Commun.*, 2014, **185**, 604–613.

42 F. Moraca, J. Amato, F. Ortuso, A. Artese, B. Pagano, E. Novellino, S. Alcaro, M. Parrinello and V. Limongelli, *Proc. Natl. Acad. Sci.*, 2017, **114**, E2136–E2145.

43 D. Wragg, A. de Almeida, R. Bonsignore, F. E. Kühn, S. Leoni and A. Casini, *Angew. Chemie - Int. Ed.*, 2018, **57**, 14524–14528.

44 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605–1612.

45 Dassault Systèmes BIOVIA, 2016.

46 Z. E. R. Newby, J. O'Connell, Y. Robles-colmenares, S. Khademi, L. J. Miercke, R. M. Stroud and J. O. C. Iii, *Nat. Struct. Mol. Biol.*, 2008, **15**, 619–625.

47 H. Sui, B. G. G. Han, J. K. J. K. K. Lee, P. Walian and B. K. B. K. K. Jap, *Nature*, 2001, **414**, 872–878.

48 C. C. Group, *Inc. Montr. Quebec, Canada.*, 2010.

49 D. Fu, A. Libson, L. J. W. Miercke, C. Weitzman, P. Nollert, J. Krucinski and R. M. Stroud, *Science (80-.).*, 2000, **290**, 481–486.

50 S. Jo, T. Kim, V. G. Iyer and W. Im, *J. Comput. Chem.*, 2008, 29, 1859–1865.

51 E. L. Wu, X. Cheng, S. Jo, H. Rui, K. C. Song, E. M. Dávila-Contreras, Y. Qi, J. Lee, V. Monje-Galvan, R. M. Venable, J. B. Klauda and W. Im, *J. Comput. Chem.*, 2014, **35**, 1997–2004.

52 A. K. Malde, L. Zuo, M. Breeze, M. Stroet, D. Poger, P. C. Nair, C. Oostenbrink and A. E. Mark, *J. Chem. Theory Comput.*, 2011, **7**, 4026–4037.

53 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindah, *SoftwareX*, 2015, **1**–**2**, 19–25.

54 S. Nosé, *Mol. Phys.*, 1984, **52**, 255–268.

55 M. Parrinello and A. Rahman, *Phys. Rev. Lett.*, 1980, **45**, 1196–1199.

Chapter 7 - G-quadruplex DNA and molecular dynamics

7.1 G-quadruplex structure

G-quadruplexes (G4s) are secondary DNA structures containing a high proportion of guanine bases. The high stability of G4s has been known since their discovery by Gellert, Lipsett and Davis in 1962[1], and has been attributed to the strong hydrogen bond network formed by the assembly of these guanine-rich sequences into planar layers consisting of four guanine units, known as G4-tetrads, bound *via* Hoogsteen hydrogen bonds. The G4s are further stabilized by π-π interactions between tetrads and by the presence of monovalent cations (e.g. K^+ and Na^+)[2] (Figure 7.1).

In fact, the cavity formed in the centre of each tetrad, first thought to be capable of accommodating water molecules[1], contains a metal cation, coordinated to the O6 of each of the four guanines, which further stabilises the structure. Repeating tetrad sequences are then stacked together to form the G4 unit[3]. The folding of G4s can lead to a multitude of different conformations which can be categorized using the orientation of the strand folding topologies as antiparallel, parallel and hybrid[4] (Figure 7.2), indicating the strand direction from the 5' to 3' end of the sequence. Folding conformations can be dependent on the sequence and environment, such as the cation type and concentration[5].

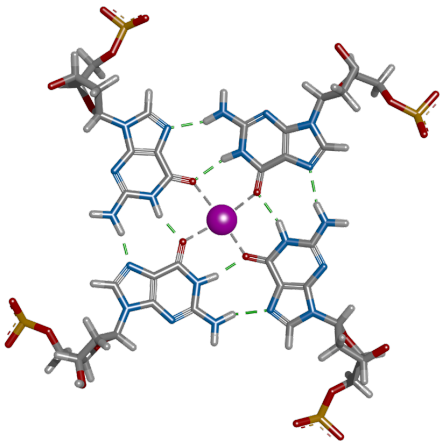


Figure 7.1 Single G-tetrad showing Hoogsteen-type hydrogen bonds and coordinated K^+ cation (shown in purple).

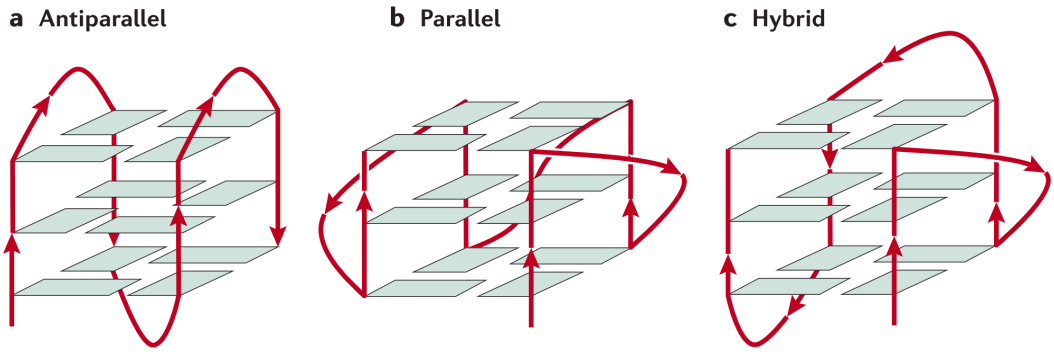


Figure 7.2 Representation of G4-DNA folding topologies from 5' to 3'. a, Antiparallel. b, Parallel. c, Hybrid (taken from Neidle 2017)[4].

7.2 G-quadruplex DNA and cancer

G-quadruplexes are present in both telomeres and promotor regions of oncogenes[6] and are known to be involved in a number of biological processes, such as telomere maintenance and replication as well as oncogene regulation[4,7] and have therefore, become an important target for cancer therapy[8].

A number of factors account for the variety in G-quadruplex structures observed. Base sequence order, solution conditions; such as intracellular concentration of K^+ and Na^+[9] or whether the G4 stack is formed from a single (intramolecular) or multiple (up to four) individual guanine rich strands[9] all have an effect on the overall topology. These factors in conjunction with parallel, antiparallel or hybrid topologies[10] (Figure 7.2) can lead to the exposure of multiple tetrad faces. In the case of Tel 23[11] (pdb ID 5CCW) both the upper and lower tetrads are exposed. Alternately, in the case of *h*Telo[12] (pdb ID 2HY9), and *C-KIT*1[13] (pdb ID 4WO2) (Figure 7.3), only the upper tetrad is exposed. Also, due to the extended loop observed on *h*Telo access to the upper tetrad is hindered.

Telomeric DNA is found at the 3' end of the chromosome with the final 100-200 bases forming a single strand[4]. With each replication cycle, telomere length is reduced until it reaches a finite limit, whereby, senescence and apoptosis are induced[14]. The main

focus of therapeutic telomer G4 stabilisation is indirect targeting of the reverse-transcriptase enzyme telomerase.

The promoter regions are found at the 5' end of the DNA sequences, and also within the sequence at sites of nuclease hypersensitivity between the gene coding sequences[15]. These areas above the transcription starting point of the gene sequence are able to dissociate from the double strand folding and act as binding sites for transcription factors such as polymerase, and also give them the flexibility to form G4 units. As a number of different gene sequences will be present on each chromosome, there will be a number of these promoter binding sites along the strand, providing many potential sites for G4 folding[16] and potentially gene-specific repression through the selective targeting of G4 units.[15]

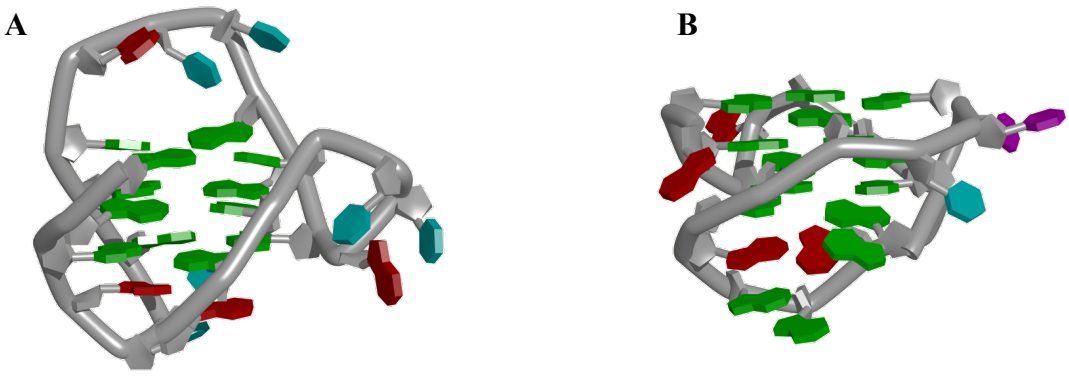


Figure 7.3 Crystal structures of two G-quadruplex models: **A**, *h*Telo (pdb ID 2HY9), a telomeric G4. **B**, *C-KIT*1(pdb ID 4WO2), a proto-oncogene promoter G4.

Most cancer types overexpress the reverse-transcriptase enzyme telomerase[4]. This enzyme is involved in the initiation of tumorigenesis and catalyses the addition to the 3' end of the chromosome TTAGGG repeats, thus, maintaining the telomer length and preventing apoptosis of the cancer cell. The addition of these guanine-rich repeats increases the likelihood of telomers folding into tandem G4 units[17] with the addition of just four TTAGGG repeats needed to form a stable G4 structure[18]. Telomeric G4s have been shown to inhibit telomerase's ability to interact with the telomer[18]

To allow replication and tumorigenesis to occur, the chromosomes need to be in the single strand state. By introducing small molecules with the high binding affinity and the ability to stabilise the folding of G4s structures, and stabilise these structures, it is

possible to indirectly inhibit these processes within cancer cells and therefore reduce both transcription and telomeric lengthening. Stabilisation is induced through π-π and π-alkyl stacking interactions between the aromatic compounds and the guanine residues forming the tetrad surface. Through stabilisation of the G4 structure, unfolding of the sequence is prevented, therefore replication of the chromosome cannot be achieved. The increased occurrence of G4 DNA structures has been shown to be a hallmark of cancer[8] and thus, provides a highly viable therapeutic target.

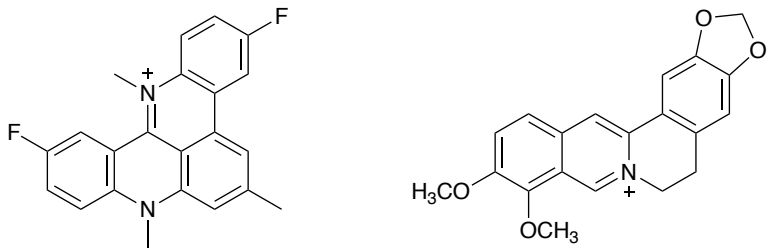
7.3 G-quadruplex stabilizers

A number of both organic and metal-based compounds have been developed with the ability to stabilise G-quadruplex DNA structures and show no affinity for duplex DNA. These molecules include key structural features that are essential to their binding selectivity: i) a heteroaromatic moiety capable of π-π stacking interactions with the planar G4 tetrad; ii) side-chains to increase binding affinity *via* van der Waals, electrostatic and H-bond interactions and cationic charge capable of delocalising the π electrons around the aromatic moiety, this positive charge also has an effect on increasing the affinity of the compounds for the negatively charged DNA backbone[19] iii) in the case of metal-based compounds, a central metal ion mimics the cation at the centre of each tetrad, helping facilitate interactions *via* a continuation of the cation chain[19].

The stabilising effect of new molecules can be quantified by different methods, including DNA melting-temperature assays[20]. This technique calculates the difference in melting temperature (ΔT_m) of the G4 structure with and without the ligands to estimate the extent of G4's stabilisation. A greater ΔT_m value obtained indicates that the selected compound has greater stabilising effect on the G4 structure. This experiment can also be run in the presence of excess duplex DNA to evaluate the selectivity of the G4 interactions.

Examples of organic G4 stabilisers include RHPS4[21], BRACO-19[22] and berberine[14] (Figure 7.4), all of which have shown anticancer activity. Whilst berberine, originally an antibiotic derived from Chinese herbal medicine[23], has undergone clinical trials as a glucose regulator[24], none have yet reached clinical trials as G4 stabilisers. Of note, BRACO-19 has been shown to induce DNA damage, as well as stabilising telomeric G4 structures, inducing uncapping of the telomer and breakdown of the telomeric T-loop strucure[25].

The ability of potential G4 stabilisers to differentiate between G4 and duplex DNA and interact exclusively with G4s is essential to avoid side-effects and toxicity. For example, the acridinium derivative RHPS4 (Figure 7.4) which, whilst showing promising *in vivo* anticancer activity in a number of animal models, also had off-target toxicity with β_2 adrenergic and hERG receptors[26].



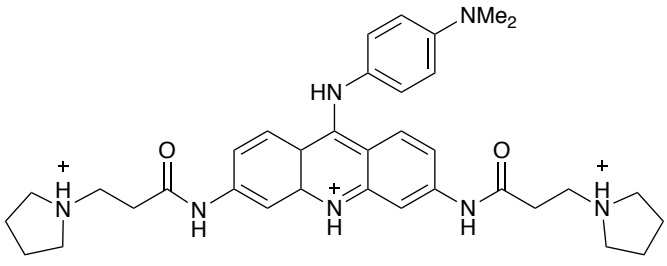
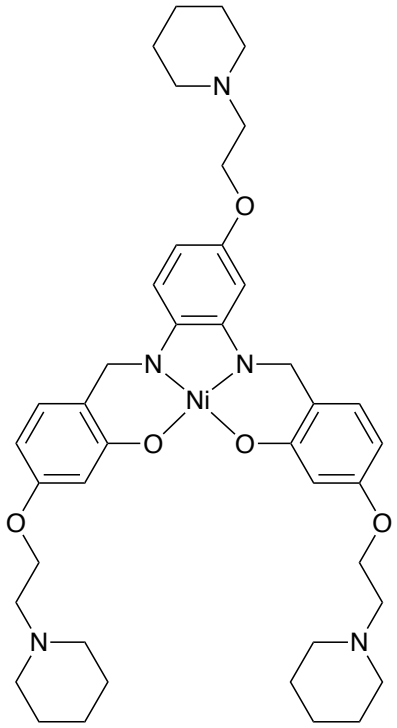


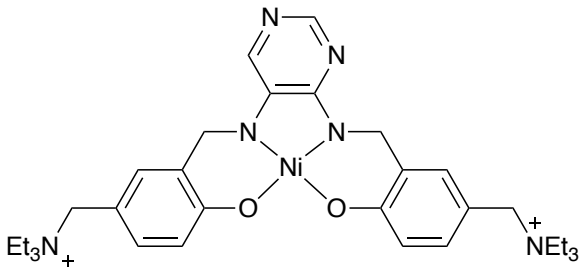
Figure 7.4 Examples of organic G4 stabilisers.

The majority of G4 inhibitors in development are organic ligands. However, with the known pharmacological activity of metal-based compounds such as cisplatin and its derivatives carboplatin[27] and oxaliplatin[28], the development of next generation anticancer drugs can take advantage of the chemically diverse scaffolds afforded by different families of organometallic and metal coordination complexes[29,30]. Thus, a

number of metal-based compounds have been identified as G4 stabilisers, including the Ni(II) tri-ethylpiperidine complex *N*,*N*0-bis[4-[[1-(2-ethyl)piperidine]oxy]salicyli-dene]-4-(2-(piperidin-1-yl)ethoxy)-1,2-phenylenediamine-Ni(II)[31] (Figure 7.5). Using the same ligand scaffold, copper, platinum and vanadium compounds were also synthesized to investigate the effects of the different metal ions on the G4 stabilization properties[31]. Whilst the nickel derivative proved to have the greater affinity for *h*Telo G4 with a ΔT_m of 26.8±1.5 °C[31], it was also shown to be less selective than the vanadium complex[31]in competition experiments with circulating tumour DNA (ct-DNA). Whilst the copper and platinum derivatives also showed excellent affinity with the *h*Telo G4, with ΔT_m values of of 22.9±1.5 °C and 25.4±1.5 °C respectively, the study did not perform the competition experiments to ascertain selectivity for these compounds[31].



Ni(II) tri-ethylpiperidine complex



Ni(II) Schiff base complex

Figure 7.5 Examples of Ni(II)-based G4 stabilisers.

The possibility to add functionalities to the ligand system make these coordination compounds ideal scaffolds for developing highly specific G4 stabilising ligands. Nickel(II) Schiff base complexes have also shown high biding affinity with telomeric

G4 structures over duplex DNA in competition experiments[30], further demonstrating the value of metal based complexes as selective inhibitors. In the case of the Salen-like derived Ni(II) compound 5-(triethylammoniummethyl)salicylaldehyde chloride (Figure 7.5), whilst it showed an selectivity for G-quadruplex DNA over duplex-DNA in competition experiments[30].

Among the organometallic complexes, featuring a direct metal-carbon bond, the gold(I) bis-carbene [Au(9-methylcaffein-8-ylidene)$_2$]$^+$ (AuTMX$_2$, Figure 7.6) has also been shown to bind preferentially to G4s *vs* duplex DNA, as well as to possess anticancer properties against human ovarian cancer cell lines[32].

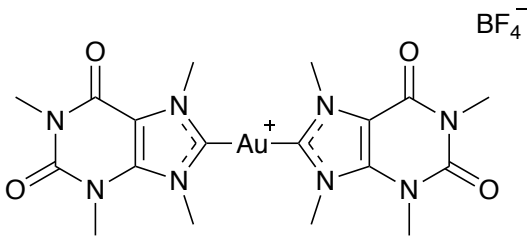


Figure 7.6 [Au(9-methylcaffein-8-ylidene)$_2$]$^+$ (AuTMX$_2$), an example of an organometallic G4 stabiliser.

In this case, the structure of the adduct of AuTMX$_2$ with the Tel 23 G4 has been solved by X-ray crystallography (Figure 7.7) [11]. The crystal structure shows that three individual AuTMX$_2$ are able to interact with the G4 at two distinct sites on both the top and bottom exposed tetrad surfaces. Moreover, this leads to the stacking of multiple G4 units with AuTMX$_2$ at the interface of both the upper and lower surfaces of adjacent G4s molecules.

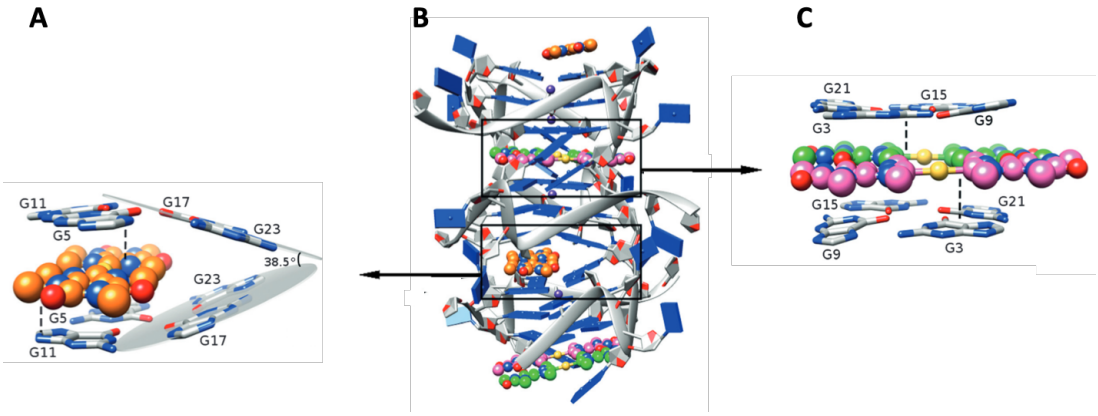


Figure 7.7 AuTMX$_2$ binding positions with Tel 23 (pdb ID 5CCW). **A**, 3'-3' end binding site showing the position of a single AuTMX$_2$ adduct (carbons shown in orange). **B**, Overall crystal structure of Tel 23 with six AuTMX$_2$ adducts. **C**, 5'-5' end binding site showing the positions of two AuTMX$_2$ adducts (carbons coloured green and pink). Dashed lines indicate interplanar distance of ca 3.4 A[11].

7.4 Molecular dynamics simulations of G-quadruplex DNA

G-quadruplex structures make up less than 2% of the DNA structures registered in the Protein Data Bank, with the majority of these being registered in the last 8 years (145/241 registered since 2010) showing G4s as an up and coming area of research. A number of molecular dynamics studies have been carried out with various G4 models to enhance our knowledge of their dynamic nature.

The majority of MD simulations concentrate on telomeric G4s[14,30,33] rather than promotor region G4s, providing and insight into the dynamics of G4-ligand integrations and lowest energy conformations through the use free energy calculations[33]. These findings indicate an optimal distance and orientation of the stabilising ligand with respect to the tetrad surface.

Studies on the affinity of Schiff base complexes with a telomeric G4 model shows the potential for metal-based complexes as therapeutics[30]. Unbiased simulations of Ni(II) Schiff base complex (Figure 7.5) show how the complex 'docks' with the telomeric G4 (*h*Telo (pdb ID 2HY9) models upper exposed tetrad due to the electrostatic attractions between the positively charged ligand and the negativity charged phosphate groups of the G4. Further π - π interactions were observed between the complex and guanines 6, 10, 14 and 22 of the upper tetrad as the complex positions itself with the nickel ion approximately in line with the central potassium ion (Figure 7.8).

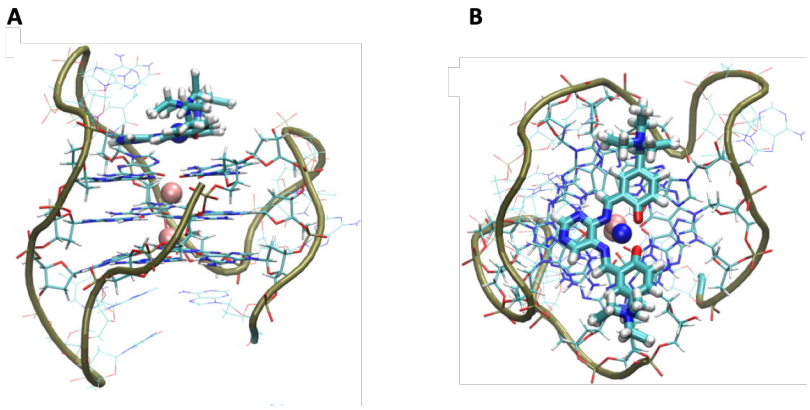


Figure 7.8 Ni(II) Schiff base 'docking' with *h*Telo (pdb ID 2HY9). **A**, Side view showing the complexes optimal distance above the upper tetrad. **B**, Top view showing alignment of the nickel ion centre with the central potassium ion of the tetrad[30].

Whilst these calculations allow the positioning of the ligand with respect to G-quadruplexes, they are unable to provide quantitative data on the binding affinities. The development and implementation of *in silico* methods suitable for calculating the binding affinities of small molecules with biological systems allows for greater comparison with experimental data.

One such method is metadynamics and its variants. The free energy of binding for organic ligands using funnel metadynamics[14] shows that state of the art *in silico* techniques can provide results matching those of experimental and provide a greater insight into the mechanistic details of ligand-G4 interactions. The binding mechanism of berberine with the human telomeric G4-DNA model (pdb ID 3R6R) was investigated with this method allowing direct comparison with experimental results[14].

The use of metadynamics allows the sampling and calculation of the free energy landscape of the system. The binding free energy of berberine with the telomeric G4 model was calculated using variant of metadynamics called funnel- metadynamics. In this case a funnel shaped potential is applied to the 3'-end of the G4-DNA complex to prevent the berberine ligand visiting areas of the system outside of the funnel (Figure 7.9). This increases the number of data points associated with the binding and unbinding mechanism without affecting the sampling data. The results show the positioning and orientation of the berberine with respect to the outer tetrad surface and the favourable π - π interactions between the complex and guanines 4 and 10 of the G4-DNA model as well as the stabilising salt bridge between the N^+ and the phosphate of the DNA backbone.

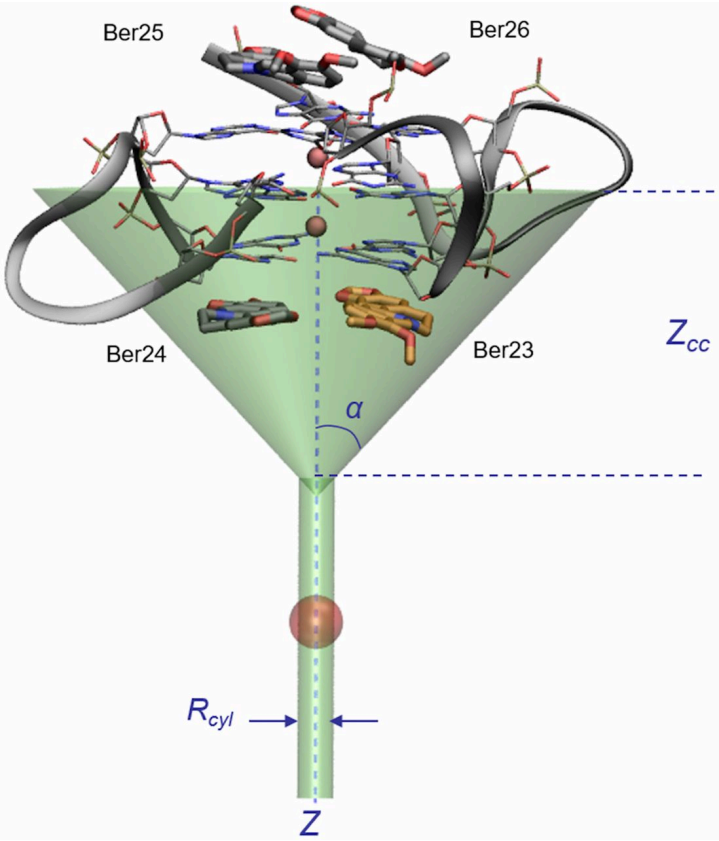


Figure 7.9 The funnel shaped potential restraint used my Moraca *et al*[14]. The angle α is 0.75 rad, distance z_{cc} is 16Å and R_{cyl} is 1Å.

These interactions can be seen as energy basins in the calculated free energy surface based on the collective variables (CV), in this case the distance between berberine and the tetrad and the angle formed between the two were chosen as the CVs. The data calculated for the binding affinity and position of berberine with the G4-DNA model match those found experimentally from X-ray crystallography and steady state fluorescence titration experiments whilst revealing the mechanism at an atomistic level and providing validation of the computational techniques used.

Overall, the above-mentioned examples provide trajectories of a relevant timescale for the interactions under investigation. Whilst both rely on the accuracy of the force field parameters to provide relevant data, by working in parallel with experimental techniques this accuracy can be constantly assessed to ensure such accuracy.

7.4 References

1 M. Gellert, M. N. Lipsett and D. R. Davis, *Proc. Natl. Acad. Sci. U. S. A.*, 1962, **48**, 2013–2018.

2 D. Bhattacharyya, G. Mirihana Arachchilage and S. Basu, *Front. Chem.*, 2016, **4**, 1–14.

3 S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd and S. Neidle, *Nucleic Acids Res.*, 2006, **34**, 5402–5415.

4 S. Neidle, *Nat. Rev. Chem.*, 2017, **1**, 1–10.

5 S. L. Noer, S. Preus, D. Gudnason, M. Aznauryan, J. L. Mergny and V. Birkedal, *Nucleic Acids Res.*, 2016, **44**, 464–471.

6 V. S. Chambers, G. Marsico, J. M. Boutell, M. Di Antonio, G. P. Smith and S. Balasubramanian, *Nat. Biotechnol.*, 2015, **33**, 877–881.

7 D. Rhodes and H. J. Lipps, *Nucleic Acids Res.*, 2015, **43**, 8627–8637.

8 R. Hänsel-Hertsch, D. Beraldi, S. V Lensing, G. Marsico, K. Zyner, A. Parry, M. Di Antonio, J. Pike, H. Kimura, M. Narita, D. Tannahill and S. Balasubramanian, *Nat. Genet.*, 2016, **48**, 1267.

9 G. Arrabito and L. Wang, Eds., *DNA Nanotechnology for Bioanalysis: From Hybrid Dna Nanostructures To Functional Devices*, World Scientific Europe Ltd, 2017.

10 H.-Q. Yu, D. Miyoshi and N. Sugimoto, *J. Am. Chem. Soc.*, 2006, **128**, 15461–15468.

11 C. Bazzicalupi, M. Ferraroni, F. Papi, L. Massai, B. Bertrand, L. Messori, P. Gratteri and A. Casini, *Angew. Chemie - Int. Ed.*, 2016, **55**, 4256–4259.

12 J. Dai, C. Punchihewa, A. Ambrus, D. Chen, R. A. Jones and D. Yang, *Nucleic Acids Res.*, 2007, **35**, 2440–2450.

13 D. Wei, J. Husby and S. Neidle, *Nucleic Acids Res.*, 2015, **43**, 629–644.

14 F. Moraca, J. Amato, F. Ortuso, A. Artese, B. Pagano, E. Novellino, S. Alcaro, M. Parrinello and V. Limongelli, *Proc. Natl. Acad. Sci.*, 2017, **114**, E2136–E2145.

15 A. Siddiqui-jain, C. L. Grand, D. J. Bearss and L. H. Hurley, 2002, **99**, 11593–11598.

16 I. Minn, M. E. Menezes, S. Sarkar, K. Yarlagadda, S. K. Das, L. Emdad, D. Sarkar, P. B. Fisher and M. G. Pomper, in *Emerging Applications of Molecular Imaging to Oncology*, eds. M. G. Pomper and P. B. B. T.-A. in C. R. Fisher, Academic Press, 2014, vol. 124, pp. 131–169.

17 A. Bugaut and P. Alberti, *Biochimie*, 2015, **113**, 125–133.

18 T. M. Fletcher, D. Sun, M. Salazar and L. H. Hurley, *Biochemistry*, 1998, **37**, 5536–5541.

19 R. Vilar, C. L. Ruehl, A. H. M. Lim, T. Kench and D. J. Mann, *Chem. – A Eur. J.*, 2019, 1–11.

20 R. Rodriguez, S. Müller, J. A. Yeoman, C. Trentesaux, J.-F. Riou and S. Balasubramanian, *J. Am. Chem. Soc.*, 2008, **130**, 15758–15759.

21 C. Leonetti, M. Scarsella, G. Riggio, A. Rizzo, E. Salvati, M. D'Incalci, L. Staszewsky, R. Frapolli, M. F. Stevens, A. Stoppacciaro, M. Mottolese, B. Antoniani, E. Gilson, G. Zupi and A. Biroccio, *Clin. Cancer Res.*, 2008, **14**, 7284 LP – 7291.

22 N. H. Campbell, G. N. Parkinson, A. P. Reszka and S. Neidle, 2008, 6722–6724.

23 M. Franceschin, L. Rossetti, A. D'Ambrosio, S. Schirripa, A. Bianco, G. Ortaggi, M. Savino, C. Schultes and S. Neidle, *Bioorganic Med. Chem. Lett.*, 2006, **16**, 1707–1711.

24 J. et al Yin, *Metabolism*, 2008, **57**, 712–717.

25 G. Zhou, X. Liu, Y. Li, S. Xu, C. Ma, X. Wu, Y. Cheng, Z. Yu, G. Zhao and Y. Chen, *Oncotarget*, 2016, **7**, 14925–14939.

26 S. Iachettini, M. F. G. Stevens, M. Frigerio, M. G. Hummersone, I. Hutchinson, T. P. Garner, M. S. Searle, D. W. Wilson, M. Munde, R. Nanjunda, C. D'Angelo, P. Zizza, A. Rizzo, C. Cingolani, F. De Cicco, M. Porru, M. D'Incalci, C. Leonetti, A. Biroccio and E. Salvati, *J. Exp. Clin. Cancer Res.*, 2013, **32**, 68.

27 R. Canetta, M. Rozencweig and S. K. Carter, *Cancer Treat. Rev.*, 1985, **12**, 125–136.

28 D. Townsend, *xPharm Compr. Pharmacol. Ref.*, 2011, **3**, 1–4.

29 Ö. Karaca, S. M. Meier-Menches, A. Casini and F. E. Kühn, *Chem. Commun.*, 2017, **53**, 8249–8260.

30 R. Bonsignore, A. Terenzi, A. Spinello, A. Martorana, A. Lauria, A. M.

Almerico, B. K. Keppler and G. Barone, *J. Inorg. Biochem.*, 2016, **161**, 115–121.

31 N. H. Abd Karim, O. Mendoza, A. Shivalingam, A. J. Thompson, S. Ghosh, M. K. Kuimova and R. Vilar, *RSC Adv.*, 2014, **4**, 3355–3363.

32 B. Bertrand, L. Stefan, M. Pirrotta, D. Monchaud, E. Bodio, P. Richard, P. Le Gendre, E. Warmerdam, M. H. De Jager, G. M. M. Groothuis, M. Picquet and A. Casini, *Inorg. Chem.*, 2014, **53**, 2296–2303.

33 E. Fadrná, N. Špačková, R. Štefl, J. Koča, T. E. Cheatham and J. Šponer, *Biophys. J.*, 2004, **87**, 227–242.

Chapter 8 - On the mechanism of gold NHC compounds binding to DNA G- quadruplexes elucidated by combined metadynamics and biophysical methods

Based on the paper:

Darren Wragg, Andreia de Almeida, Riccardo Bonsignore, Fritz E. Kühn, Stefano Leoni and Angela Casini. Angewandte Chemie - Int. Ed., 2018, 57, 14524–14528.

DOI: 10.1002/anie.201805727

FRET melting assays performed by Dr Riccardo Bonsignore.

Abstract: G-quadruplexes (G4s) are four-stranded DNA structures that are over-represented in gene promoter regions and are viewed as emerging therapeutic targets in oncology. The binding modes and free-energy landscape of two Au(I) N-heterocyclic carbene complexes interacting with G4s, namely a human telomeric (hTelo) and a promoter (*C-KIT1*) sequence, are studied here for the first time using metadynamics. The theoretical results are validated by FRET DNA melting assays and provide an accurate estimate of the absolute gold complex/DNA binding free energy. This advanced *in silico* approach is valuable to achieve rational drug design of selective G4s binders.

8.1 Introduction

DNA is able to adopt different structures other than the canonical right-handed double helix (B-DNA), and numerous structural studies have shown that guanine-rich DNA sequences can form secondary structures termed G-quadruplexes (G4s)[1]. To form G4s, four guanine bases assemble into a pseudoplanar tetrad (G-quartet) which is held together by one or more nucleotide strands and stabilized by metal ions. Recent bioinformatics studies have shown that there are ca. 716,000 DNA sequences in the human genome with the potential to form G4 structures[2]. These sequences are present in telomeres and promoter regions of oncogenes, such as *C-MYC* and *C-KIT1*. These non-canonical DNA structures have been the subject of intense study over the past 10 years due to their association with a number of biological processes such as telomere maintenance, gene regulation, and replication[3,4].

It has been suggested that formation of the quadruplex structure in promoter regions can control transcription and, as a consequence, the expression of the corresponding oncogenes.[5] Furthermore, stabilizing G4s in telomeres results in indirectly inhibiting telomerase activity, affecting cancer mortality.[4]

In this context, G4s emerge as promising targets for anticancer drug development, while their roles in cancer biology have yet to be completely elucidated. A number of

studies report on the stabilization of G4s by small molecules with associated anticancer effects[6]. One example is the tri-substituted acridine derivative, BRACO-19, a telomeric G4 stabilizer, which has shown *in vitro* anticancer activity in prostate cancer.[7] Of note are two quinolone molecules, CX-3543 and CX-5461, selectively stabilize G4s structures, and have now entered clinical trials[8].

In addition to organic molecules, the development of metal-based compounds have also shown to be promising experimental G4 stabilizers, including several Schiff-base metal complexes (mainly Ni^{2+}[9–14], Cu^{2+}[11–13], Zn^{2+}[9,11,15], Pt^{2+}[16], and Pt^{4+}[16], as well as some metallo-supramolecular DNA-binders[17–19]. Despite the great advances in the development of G4 stabilizers, important challenges still remain. For example, highly selective binding of small molecules to specific quadruplex structures over duplex DNA and other G4s, is a major issue which needs to be addressed if this approach is to make it into the next generation of anticancer drugs.

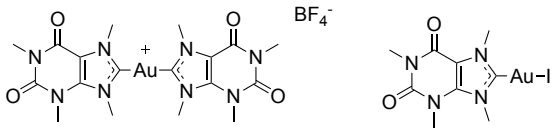
8.2 Results and discussion

Our pioneering work into G4 stabilizers identified small-molecule organometallic Au(I) compounds, featuring N-heterocyclic carbenes (NHCs) ligands, as potent, selective stabilizers of telomeric G4s.[20–22] Thus, we recently reported on a bis-NHC gold(I) complex - $[Au(9\text{-methylcaffein-8-ylidene})_2]^+$ (**$AuTMX_2$**, Figure 8.1) - containing the natural product caffeine as NHC ligand[20]. The compound was shown to have selective antiproliferative effects on human ovarian cancer cells over non-tumorigenic ones. X-ray diffraction analysis of the adduct formed by **$AuTMX_2$** and Tel23, a 23-nucleotide telomere repeat sequence, indicated that the compound binds non-covalently between neighbouring G4 units[21].

Based on this promising work, our research involves developing new organometallic Au(I) NHC complexes capable of targeting specific G4 structures for possible in therapeutic and/or imaging applications. In order to rationally achieve the required selectivity, *in silico* methods, including molecular dynamics (MD) approaches, are highly valuable as they can elucidate both the structural and energetics underlying the

ligand/target recognition process. In fact, a number of classical MD studies on the adducts of G4s structures with a number of different stabilizers have been performed providing atomistic support for the interpretation of the binding mechanism to G4-DNA[9,11,15,23,24].

Funnel-metadynamics has recently been shown to be successful method for calculating the free energy surface of organic ligands and their interactions with G4s[25]. Thus, we applied metadynamics to evaluate the binding of **AuTMX$_2$** to two different G-quadruplex structures, the human telomeric sequence hTelo (pdb 2HY9[26]) and the *C-KIT1* oncogene promoter sequence (pdb 4WO2[27]). The results have been compared with those calculated for the neutral mono-carbene complex **AuTMX-I** (Figure 8.1). To validate the accuracy of our calculations we also performed gold complexes/G4 binding assays using FRET (fluorescence resonance energy transfer) DNA melting.



AuTMX$_2$ **AuTMX-I**

Figure 8.1 Chemical structures of the two Au(I) NHCs investigated in this study.

The X-ray structure of Tel23, the telomeric-G4 adduct, with **AuTMX$_2$** (pdb 5CCW[21]) was used as reference to run a first set of metadynamics simulations. These initial simulations allowed us to calculate the compound's most thermodynamically stable positions and provided a starting point for the free energy calculations with hTelo and *C-KIT1* (see methodology section for details). This allowed the validation and initial positioning of the interaction of **AuTMX$_2$** with the selected G4 models to be ascertained. The X-ray structure shows seven molecules of **AuTMX$_2$**, some of which are intercalated between the different G4 stacks (Figure 8.2). In our study, the Gibbs-free energy (at 300 K, ΔG_{MD}) was determined for all the seven poses (Table 8.1) and showed each one has a different binding energy. The starting position for hTelo and *C-KIT1* corresponds to the **AuTMX$_2$** interaction with the topmost tetrad, where **AuTMX$_2$** interacts exclusively with one G4 stack (pose 1, Figure 8.2), This choice is more representative of the situation occurring in solution, where model G4s monomers are present.

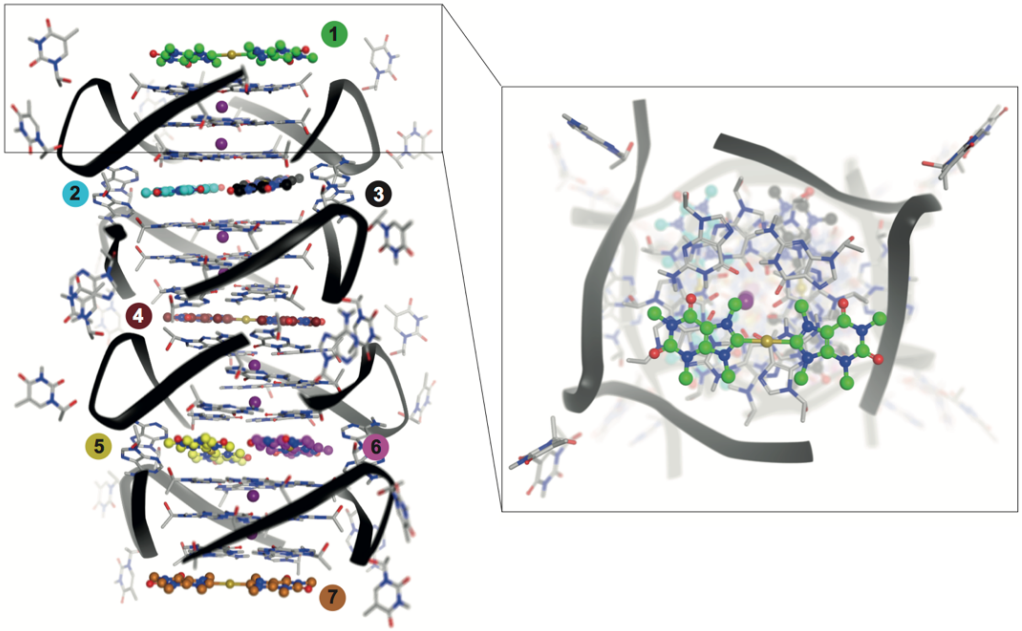


Figure 8.2 X-Ray crystallographic structure of **AuTMX$_2$** interaction with Tel23 (pdb 5CCW), showing the different poses of the seven complexes with the G4 stack. **AuTMX$_2$** positions are numbered from **1-7** and complexes are colored accordingly. K^+ ions are shown as purple spheres. On the right panel a top view of the G4 stack, showing pose 1 (colored in green) is shown. This pose chosen as a starting position for the further metadynamics simulations.

Table 8.1 Gibbs-free energy (ΔG_{300}) values (kJ/mol) for **AuTMX$_2$** interaction with Tel23 (pdb 5CCW). For poses, refer to Figure 8.2.

Pose	ΔG (kJ/mol)
1	-33 ± 9
2	-53 ± 4
3	-51 ± 1
4	-58 ± 9
5	-51 ± 3
6	-47 ± 8
7	-32 ± 13

Moraca et al. used funnel metadynamics to constrain the ligand within a specific binding area determined to be the tetrad surface, allowing multiple iterations of binding/unbinding within a single large calculation[25]. However, in our study, the gold

complexes were not constrained to one area and were enabled to find the most energetically favourable interactions with the entire G4 models, including loops and top and bottom tetrads, allowing for possible further interactions to be identified. Thus, five 50 ns trajectories were calculated for each combination of compound and G4 model (providing a total of 4 experimental conditions and a total of 20 simulations). The first set of calculations were performed using a simple distance collective variable (CV) between the Au^+ centre of the complex and the K^+ at the centre of the uppermost tetrad (see methodology section for details), resulting in a free-energy (ΔG_{MD}, at 300 K) profile output, based on the Au^+-K^+ distance (Figure 8.3, Table 8.2). To closely investigate the molecular mechanism of interaction of the gold complexes, further multi-CV calculations were run on the same systems. This involved adding a second CV for the torsion angle between the complexes and the uppermost tetrads (Figure 4).

Table 8.2 Gibbs-free energy values, experimental (ΔGexp) and calculated by metadynamics (ΔG_{MD}), for **AuTMX$_2$** and **AuTMX-I** interactions with hTelo and *C-KIT1*. ΔG values are expressed in kJ/mol and obtained considering T = 300K. [a] Data are obtained from simulations using a simple distance collective variable.

	G4 model			
	hTelo		***C-KIT1***	
	ΔG_{MD}[a]	ΔG_{exp}	ΔG_{MD}[a]	ΔG_{exp}
AuTMX$_2$ (state I)	-37 ± 7	-39 ± 2	-45 ± 3	-37.6 ± 0.4
AuTMX$_2$ (state II)	-14 ± 3	-12.1 ± 0.4	-12 ± 3	-15 ± 1
AuTMX-I	-28 ± 3	-23.2 ± 0.4	-30 ± 5	-29 ± 3

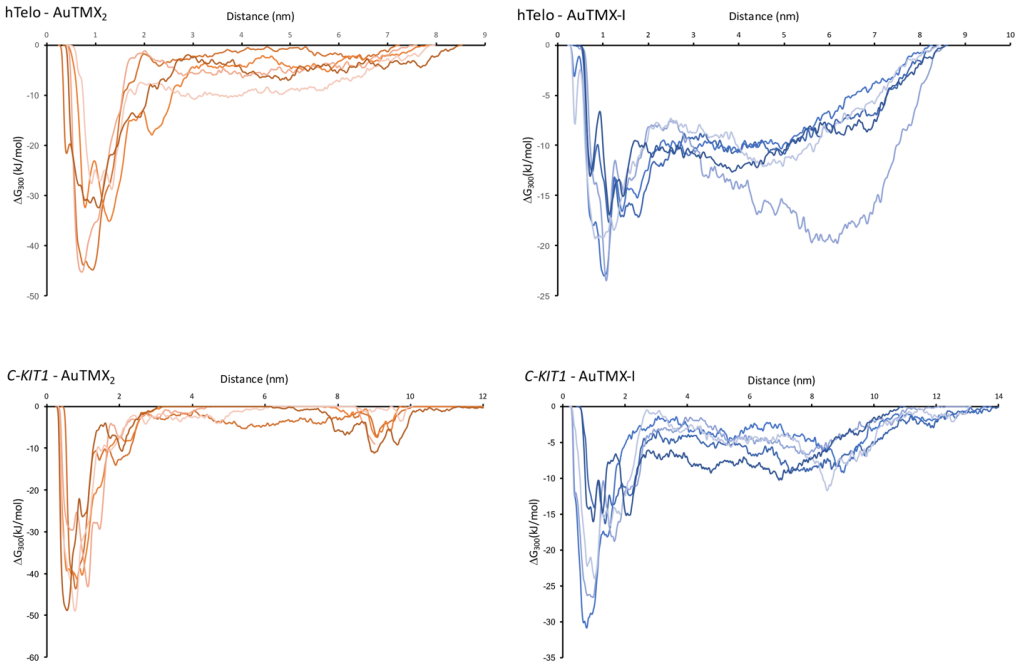


Figure 8.3 Graphical overlay of the free energy surface (FES) single collective variable (CV) plots for hTelo and *C-KIT1* with **AuTMX₂** (orange) and **AuTMX-I** (blue). Each overlay consists of the five individual FES for each system.

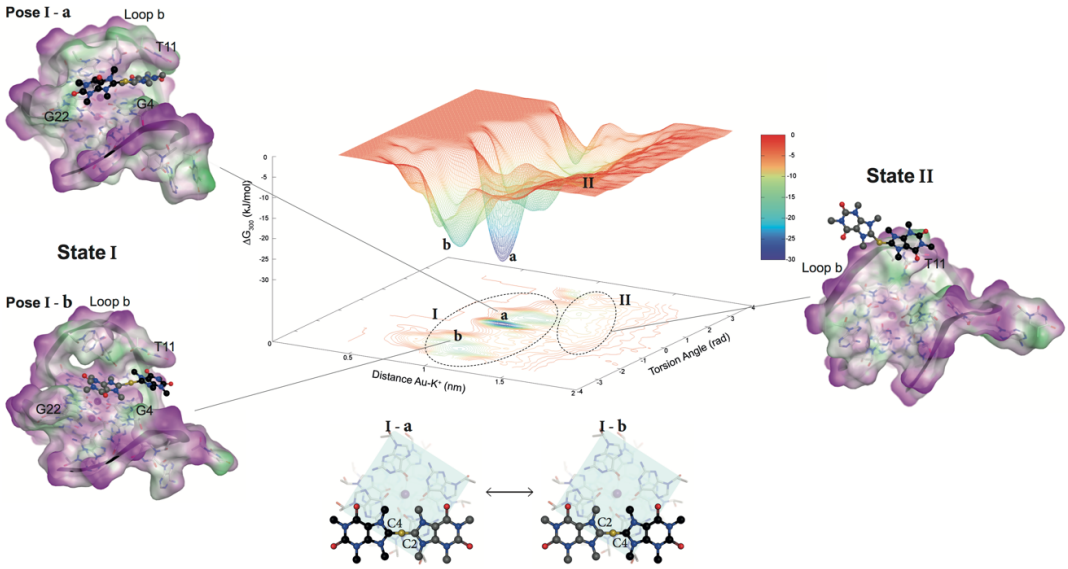


Figure 8.4 Multiple collective variable (CV) plot of free energy surface of **AuTMX₂** interactions with hTelo (centre). CVs correspond to distance (nm) between Au^{+} in **AuTMX₂** and K^{+} in upper tetrad and torsion angle (rad). Two states are highlighted (**I** and **II**) and two poses for state **I** are shown as **a** and **b**. States **I-a, b** and **II** are shown in translucent molecular surface, coloured according to lipophilicity (green: lipophilic, pink: hydrophilic). G4 structure is shown as sticks and ribbon, with hidden backbone for clarity. **AuTMX₂** is shown in ball and stick, with each caffeine ligand coloured differently (black and grey). C2 and C4 highlight the carbon atom positions in **AuTMX₂** in each of the related poses **I-a** or **I-b**.

As metadynamics allows for the exploration of the whole free energy surface of an interaction, rather than just one minimum, it is possible for further meta-stable positions to also be observed. In fact, this was the case for hTelo's trajectories with **AuTMX$_2$** which shows two possible binding sites (state I and II) with the first one (state I) having the lowest energy (ca. -37 KJ/mol, Table 8.2). Figure 8.4 shows the multiple collective variable (CV) plot of the free energy surface for **AuTMX$_2$** interactions with hTelo. Interestingly, state I shows two minima (**a** and **b**), corresponding to the same Au^+-K^+ distance (ca. 0.8 nm) but with two different torsion angles. The latter are related to **AuTMX$_2$** being in virtually the same position above the tetrad but with the gold complex rotating around its centre, thus, resulting in the same pose with two different torsion angles (see position of the caffeine ligands in Figure 8.4).

In state I, **AuTMX$_2$** is interacting with both an adenine (A13) in the loop b region, and two of the guanine bases in the upper tetrad (G4 and G22), with strong π-stacking interactions between the NHCs of the gold complex and the aromatic rings of G22 (Figure 8.4). Instead, the higher energy state II (ca. -14 KJ/mol, Table 8.2) corresponds to a position whereby the gold complex does not interact with any of the guanine bases in the upper tetrad, but exclusively with the loop thymine (T11) (Figure 8.4). In this second state, the loop covers the top of the G4-tetrad, hindering any possible interactions between the gold complex and the G-tetrad.

Interestingly, a similar behaviour was observed for **AuTMX$_2$** binding to *C-KIT1*, with two states I and II (Figure 8.5, Table 8.2). *C-KIT1* structure and surface are very different to that of hTelo: while hTelo has a prominent flanking loop capable of covering the top of the tetrad, *C-KIT1* top surface is virtually flat. This structural feature leads the gold complex to interact with the top of the tetrad and the rings of flanking bases (A1) (Figure 8.5) with no obstruction. Thus, state I corresponds to **AuTMX$_2$** stabilized by π-stacking with the guanine rings (lowest energy), while in state II, it interacts with both A1 and G6 *via* π-stacking (Figure 8.5).

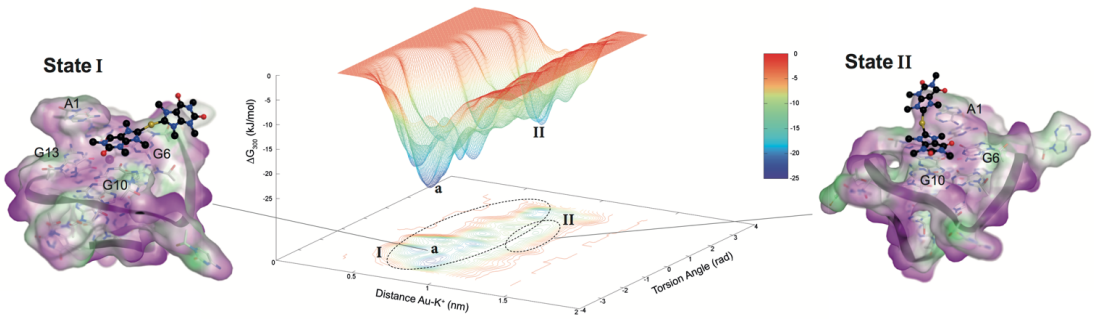


Figure 8.5 Multiple collective variable (CV) plot of free energy surface of **AuTMX₂** interactions with *C-KIT1* (center). CVs correspond to distance (nm) between Au in **AuTMX₂** and K^+ in upper tetrad and torsion angle (rad). Two states are highlighted (**I** and **II**), which correspond to **AuTMX₂** interaction with the top of the tetrad (**I**) or with both the top of the tetrad and flanking base A1 (**II**). In state **I**, one representative pose (**a**) is highlighted, where the complex is interacting with the top of the tetrad. States **I** and **II** are shown in translucent molecular surface, colored according to lipophilicity (green: lipophilic, pink: hydrophilic). **AuTMX₂** is shown in ball and stick.

Multi-CV calculations were also run for the neutral mono-NHC complex **AuTMX-I** with both hTelo and *CKIT1*. The compound was shown to interact *via* π-π and π-alkyl interactions with the guanine tetrad (G22) and the loop (A13) in hTelo (Figure 8.6), as observed for **AuTMX₂** (state I). However, as expected, the calculated ΔG_{MD} was lower with respect to that of AuTMX₂, due to the lack of the second caffeine ligand (Table 8.2).

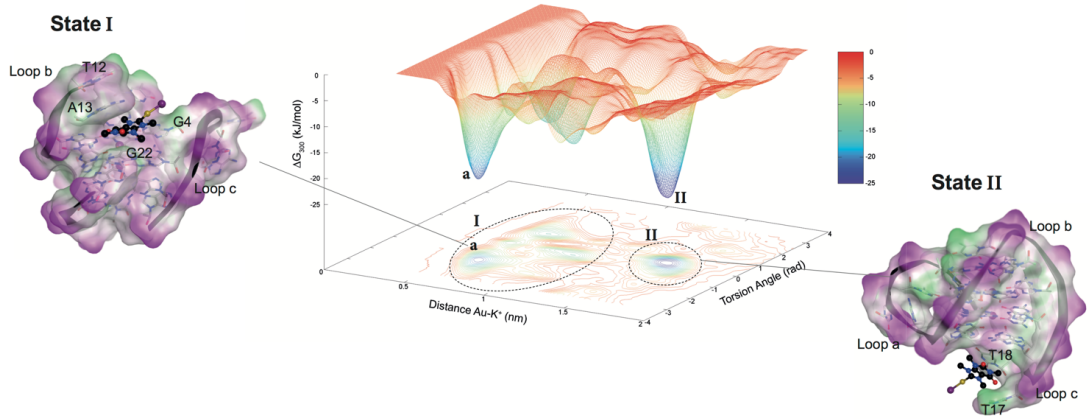


Figure 8.6 Multiple collective variable (CV) plot of free energy surface of **AuTMX-I** interactions with hTelo (center). CVs correspond to distance (nm) between Au in **AuTMX-I** and K^+ in upper tetrad and torsion angle (rad). Two states are highlighted (**I** and **II**), which correspond to **AuTMX-I** interaction with the top of the tetrad (**I**) or the groove in loop c (**II**). In state **I**, one representative pose (**a**) is highlighted, where the complex is interacting the top of the tetrad and A13 in loop b. States **I** and **II** are shown in translucent molecular surface, colored according to lipophilicity (green: lipophilic, pink: hydrophilic). **AuTMX-I** is shown in ball and stick.

The enhanced efficiency provided by multiple collective variable simulations proved extremely valuable in identifying a second, unexpected binding mode of **AuTMX-I** (state II), with a similar energy to state I. In this second case the complex interacts within a groove in loop c (Figure 8.6) by π-π stacking the caffeine moiety with T18. This interaction was only observed when using the multiple collective variable for the calculations.

In the case pof *C-KIT1*, **AuTMX-I** shows a single binding mode with the compound π-stacking with G6 of the uppermost tetrad (Figure 8.7). Moreover, the iodido ligand tends to position itself outside the G4 structure, in both hTelo and *C-KIT1* adducts.

The interesting observations of multiple binding modes of **AuTMX$_2$** led us to further investigate the stabilization properties of the two gold-based complexes. This was achieved through the determination of the Gibbs-free energy (ΔG_{exp}) from the DNA FRET melting profiles. Thus, **AuTMX$_2$** and **AuTMX-I** were synthesized by adapting published protocols[20,28]. starting from their methylated precursors and fully characterized. The difference in DNA melting temperature (ΔT_m, in °C) of hTelo and *C-KIT1* induced by the binding of the two Au(I) NHC complexes was readily monitored through the modification of the FRET phenomenon and enabled an easy quantification of the compounds' stabilization properties of G4-DNA.

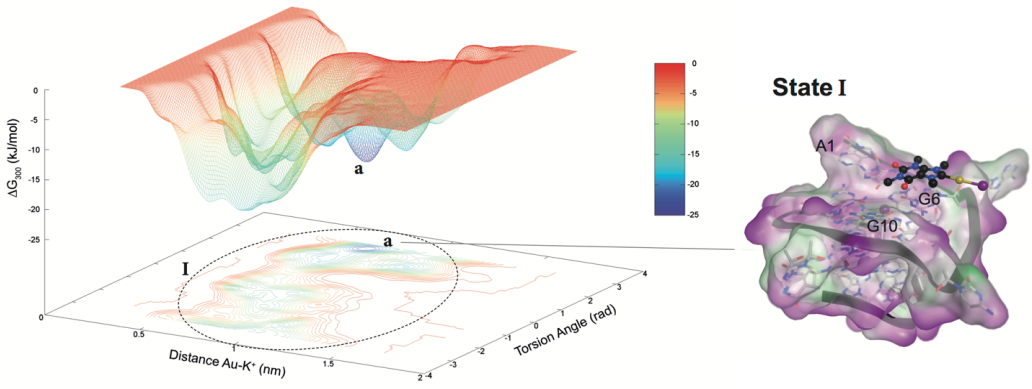


Figure 8.7 Multiple collective variable (CV) plot of free energy surface of **AuTMX-I** interactions with *C-KIT1* (center). CVs correspond to distance (nm) between Au in **AuTMX-I** and K^+ in upper tetrad and torsion angle (rad). Only one state is observed (**I**), which corresponds to **AuTMX-I** interaction with the top of the tetrad. One representative pose (***a***) is highlighted, where the complex is interacting with G6, on the top of the tetrad. Pose **a** is shown in translucent molecular surface, colored according to lipophilicity (green: lipophilic, pink: hydrophilic). **AuTMX-I** is shown in ball and stick.

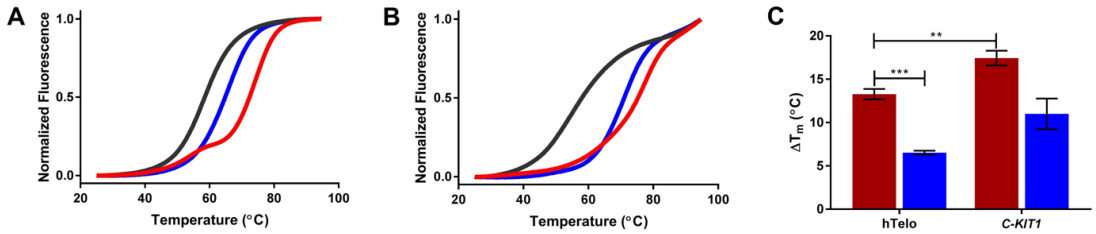


Figure 8.8 Representative FRET melting profiles of a 0.2 μM hTelo (A) and *C-KIT1* (B) G-quadruplex DNA solutions in 60 mM potassium cacodylate buffer in absence (black solid lines), and in presence of 1 μM of **AuTMX$_2$** (red lines) and **AuTMX-I** (blue lines). C) ΔT_m of hTelo and *C-KIT1* G4s in the presence of **AuTMX$_2$** (red) and **AuTMX-I** (blue). Data is shown as mean ± SEM of three independent experiments; ** $p<0.005$, *** $p<0.001$. Results for **AuTMX-I** hTelo vs *C-KIT1* are significantly different (significance not shown in the plot), * $p<0.05$.

The compounds were then incubated with fixed amounts of each G4 for 10 min and the DNA melting profile recorded. As shown in Figure 8.8, both NHC complexes stabilize the hTelo and *C-KIT1* structures, with the strongest effects being observed for **AuTMX$_2$**. As previously reported[22], **AuTMX$_2$** shows a characteristic melting profile for hTelo, featuring a two-step melting pattern, where a small increase in fluorescence is initially observed before the steep increase after ca. 65 ºC (Figure 8.8A, red trace). Alternatively, **AuTMX$_2$** with *C-KIT1*, investigated here for the first time, shows a gradually incrementing curve, rather than an initial steep ramp or two-step curve. In the case of the mono-caffeine derivative, **AuTMX-I**, the ΔT_m of both G4s is approximately 2-fold lower than the one found for **AuTMX$_2$** (Figure 8.8, blue trace). (6.5 ± 0.2 ºC for hTelo and 11.0 ± 1.8 ºC for *C-KIT1*, respectively). Interestingly, AuTMX-I does not have an effect on the shape of the melting profiles of either DNA sequence, which is similar to the control DNA (Figure 8.8A-B, black and blue traces).

In order to calculate the energy of binding (ΔG_{exp}) of the gold compounds to each G4, the experimental data were normalized to folded fraction (θ) of G4-DNA and fitted according to Eq. 1 and 2[29], where the enthalpy (ΔH) for the process was derived from the resulting fit[30]. In order to fit a two-step melting profile of **AuTMX$_2$** with hTelo, a second equation (Eq. 4) was used, allowing us to take into account the upper and lower limits of the sigmoid curve used for fitting. This allowed us to treat the data fits as two independent melting curves. Thus, ΔG_{exp} was calculated for both compounds *vs* each G4 structure, and also for hTelo's two-step melting curve (Table 8.2). The resulting

fitting curves are shown in Figure 8.9. From the reported results, as expected, a trend could be identified in which the greater the stabilization of the G4 structure by the compound and lower is the energy of binding (when considering the most stable mode, state I for hTelo).

$$\textbf{Eq. 1: } \theta(T) = \varphi(x) = \frac{e^x}{1+e^x}, \text{ where}$$

$$\textbf{Eq. 2: } x = \left(\frac{\Delta H}{R}\right)\left(\frac{1}{T} - \frac{1}{T_m}\right)$$

$$\textbf{Eq. 3: } \Delta G_{300} = \Delta H.\left(\frac{T_{MD}}{T_m} - 1\right)$$

$$\textbf{Eq. 4: } \varphi(x) = \frac{e^x}{1+e^{x-Lim_{upper}}} + Lim_{lower}$$

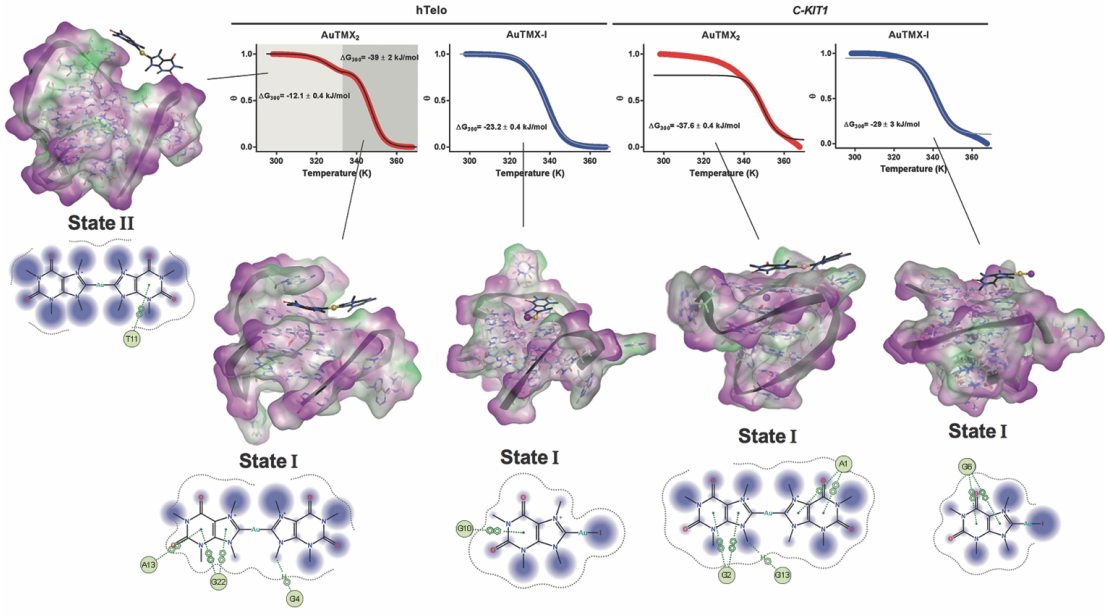


Figure 8.9 Folded fraction (θ) of hTelo and *C-KIT1* with either **AuTMX$_2$** (red traces) or **AuTMX-I** (blue traces). Corresponding fits are shown as black or grey traces, and calculated Gibbs-free energy (ΔG_{300}) is shown in each plot. The most representative poses for each combination of G4-gold complex are shown as hydrophobic (green)/ hydrophilic (magenta) surfaces, with the gold complex and the DNA bases shown in stick representation. Gold atoms, iodido ligand and K^+ are shown in ball-and-stick representation, with phosphate backbone in black ribbon. Bellow each pose the interactions observed between the complexes and the G4s, corresponding to arene-arene or arene-H interactions are shown, where A – adenine, T – thymine, G – guanine.

Most importantly, the ΔG_{exp} are in perfect accordance with the ΔG_{MD} values obtained from the metadynamics calculations (Table 8.2). Analysis of the melting curves and ΔG_{exp} values of **AuTMX$_2$** with the two G4 models clearly suggest the existence of two distinct binding modes, in line with the computational results. Thus, considering hTelo, the first binding mode corresponds to the lower energy state I (Figures 8.4 and

8.9), exclusively featuring compound's interaction with the guanines in the tetrad. The second binding mode corresponds to the higher energy (state II) involving loop/flanking base interactions and/or interactions with part of the tetrad (Figures 8.4 and 8.9). As previously observed, the flatness of the top tetrad of *C-KIT1* allows the compound to probe the entire top-surface of the G4 without hindrance. Since the stacking of the complex with the guanines is more favourable (state I), it would be expected that this is the interaction most likely to occur *in vitro*.

Notably, our metadynamics results also point towards the existence of a second binding site for **AuTMX-I** with hTelo, involving the loop C. This interaction may be exploited in the future for further optimization of selective hTelo stabilizers.

8.3 Conclusions

Overall, the obtained computational results fully support the experimental data revealing two ligand binding modes of **AuTMX$_2$** on the two G4s structures. Moreover, they are providing further structural and energetics information on the ligand binding mechanism, including a free-energy landscape quantification. The validation of the binding energy of gold(I) complexes-G4s adducts calculated by FRET DNA melting assays was also achieved by metadynamics methods. This investigational approach can be extended to other types of molecules as G4 stabilizers, including organic compounds, highlighting selectivity features essential to orient the drug design. Considering that putative G4-forming sequences have been found in other mammalian genomes, yeasts, protozoa[31], bacteria and viruses, the present integrated approach is certainly of value in other research disease areas.

8.4 Methodology Section

To validate the method, free energy calculations were first carried out using the crystal structure for Tel23 (pdb ID 5CCW) containing four identical G4 units stacked with seven **AuTMX$_2$** (Figure 8.2). A single entire stack was taken from the crystal structure and used for the MD simulation.

Further free energy calculations were carried out on the adducts of [Au(9-methylcaffein-8-ylidene)2]$^+$ (**AuTMX$_2$**) and Au(9-methylcaffein-8-ylidene) Iodide (**AuTMX-I**) with G-quadruplex models hTelo (pdb ID 2HY9) and *C-KIT1* (pdb ID 4WO2) were run with both [Au(9-methylcaffein-8-ylidene)2]$^+$ (**AuTMX$_2$**). The parameters for both the complexes were defined using Gaussian 9[32] using PBE0 hybrid functional and the LanL2DZ basis set. The information taken from the DFT calculations were used to define the bonded parameters for each complex, this included all the bond lengths, angles, dihedrals and charges.

Each system was set-up in a 10x10x10 nm box with periodic boundary conditions (PBC). The systems were solvated with approximately 32000 TIP3P water molecules dependent on G4 model used. Ions were added to each system to neutralize the charge and add a salt concertation of 0.15M using KCl to mimic physiological conditions.
All simulations were run using the GROMACS 2016.3 software[33] and the Amber99sb-ILDN-slipid forcefield. Particle-mesh Ewald method was used for calculating electrostatic interactions. The Verlet cut-off scheme with a cut-off distance of 0.84 nm was used for short range repulsive and attractive interactions and Lincs was used to constrain all bond lengths. Nose-Hoover[34] temperature coupling was used to maintain the temperature of the system (τ = 0.5 ps) at 300 K. The Parrinello-Rahman[35] algorithm was used to maintain the pressure of the system at 1 bar with a coupling constant of τ = 1.0 ps. Simulations were equilibrated for 100 ps before production. Total energy of the system was monitored using gmx_energy program within Gromacs. In the equilibrated system fluctuations remained within less than 0.5 % of the average value as shown in Figure 8.10.

Well-tempered metadynamics simulations were run for 50000000 steps with a 1fs time step (50ns) using the Plumed plugin[36] for Gromacs. The distance collective variable (CV) was used between the Au of the TMX complex and the K of the top tetrad of each G4 model. In the case of the Tel23 stack, each of the seven **AuTMX$_2$** adducts were given a separate plumed control within the same data file, allowing us to calculate the free energy for each within the same simulation. The computational parameters were set with a Gaussian height of 0.05 kJ/mol and a Gaussian width of 0.025 Å. Gaussians were added every 2000 steps (2 ps) giving a deposition rate of 0.025 kJ/mol.ps. The bias factor was set to 12 thus, the ΔT was 3600K. Eight further well-tempered metadynamics simulations were run for 150000000 steps with a 1 fs time step (150 ns) using the Plumed plugin[36] for Gromacs with and additional CV for torsion angle on each of the four systems(two for each system).The torsion angle was defined by either the two carbons either side of the gold centre for AuTMX2 (C2 and C4) or the carbon and iodine either side of the gold centre for AuTMX-I (C2 and I) and the N7 of G10 and N1 of G4 of hTelo and N1 of G2 and N1 of G10 for C-KIT1. The computational parameters were set with a Gaussian height of 0.05 kJ/mol, a Gaussian width of 0.025 Å and a torsion angle of 0.08 rad (π). Gaussians were added every 2000 steps (2 ps) giving a deposition rate of 0.025 kJ/mol.ps. The bias factor was set to 12 thus the ΔT was 3600K. The evolution of the free energy as a function of time for a representative simulation is shown in Figures 8.11 and 8.12.

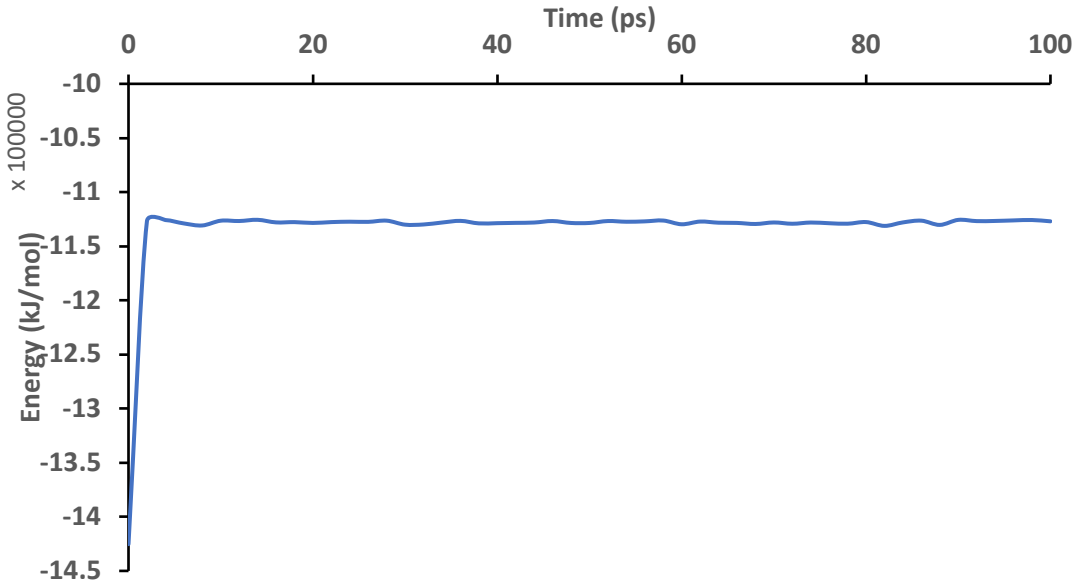


Figure 8.10: Total energy of system during 100 ps equilibration step for hTelo with **AuTMX$_2$**.

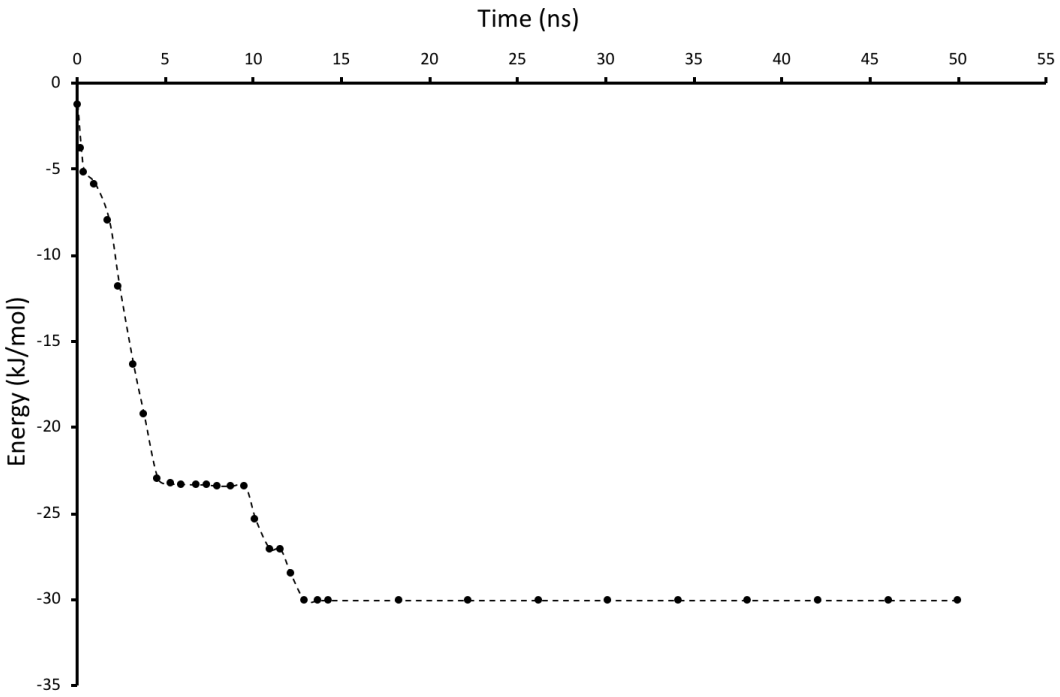


Figure 8.11: Evolution of the energy minimum of the first peak (see Figure 8.3, *C-KIT1*-**AuTMX$_2$**) as function of time. Its value does not change anymore after 15 ns.

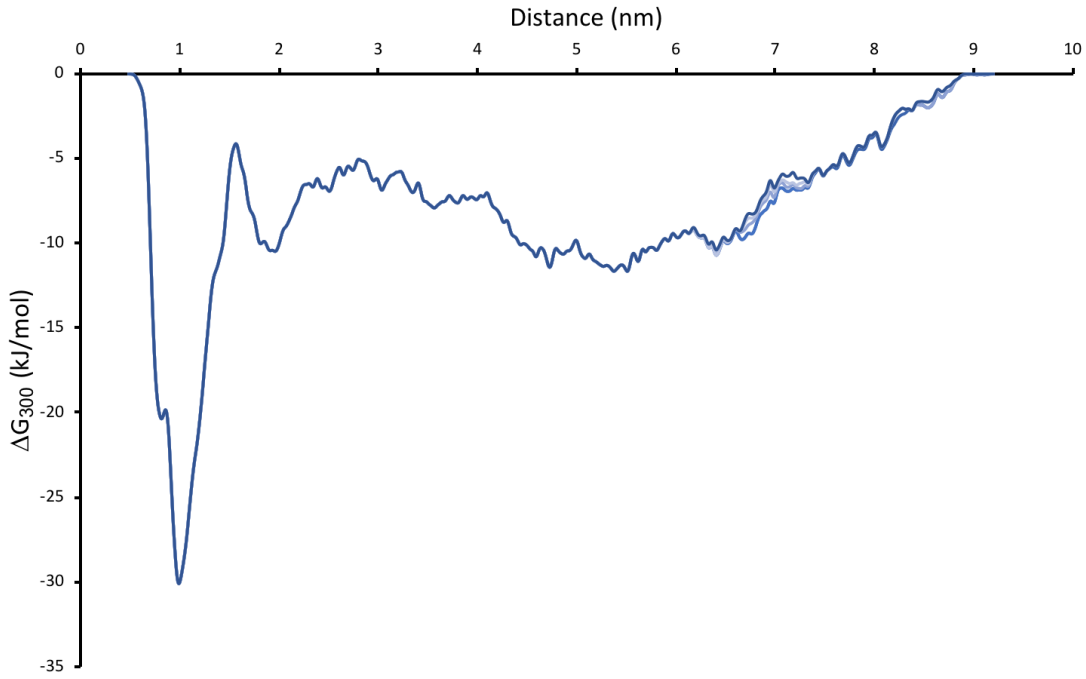


Figure 8.12: Overlays of the last 5 ns of the simulations. The 5 free energy curves are calculated with a 1 ns delay. No changes are observed in the distance range $0 < d < 3.0$ nm.

8.5 References

1 D. Monchaud and M. P. Teulade-Fichou, *Org. Biomol. Chem.*, 2008, **6**, 627–636.

2 V. S. Chambers, G. Marsico, J. M. Boutell, M. Di Antonio, G. P. Smith and S. Balasubramanian, *Nat. Biotechnol.*, 2015, **33**, 877–881.

3 D. Rhodes and H. J. Lipps, *Nucleic Acids Res.*, 2015, **43**, 8627–8637.

4 A. Bernal and L. Tusell, *Int. J. Mol. Sci.*, 2018, **19**, 1–21.

5 S. Neidle, *Therapeutic Applications of Quadruplex Nucleic Acids,* Academic Press, Boston, 2012.

6 S. Balasubramanian, L. H. Hurley and S. Neidle, *Nat. Rev. Drug Discov.*, 2011, **10**, 261–275.

7 C. Incles, C. Schultes, H. Kempski, H. Koehler, L. R Kelland and S. Neidle, *Mol. Cancer Ther.*, 2004, **3**, 1201–1206.

8 H. Xu, M. Di Antonio, S. McKinney, V. Mathew, B. Ho, N. J. O'Neil, N. Dos Santos, J. Silvester, V. Wei, J. Garcia, F. Kabeer, D. Lai, P. Soriano, J. Banáth, D. S. Chiu, D. Yap, D. D. Le, F. B. Ye, A. Zhang, K. Thu, J. Soong, S. Lin, A. H. C. Tsai, T. Osako, T. Algara, D. N. Saunders, J. Wong, J. Xian, M. B. Bally, J. D. Brenton, G. W. Brown, S. P. Shah, D. Cescon, T. W. Mak, C. Caldas, P. C. Stirling, P. Hieter, S. Balasubramanian and S. Aparicio, *Nat. Commun.*, 2017, **8**, 14432.

9 R. Bonsignore, F. Russo, A. Terenzi, A. Spinello, A. Lauria, G. Gennaro, A. M. Almerico, B. K. Keppler and G. Barone, *J. Inorg. Biochem.*, 2018, **178**, 106–114.

10 R. Bonsignore, A. Terenzi, A. Spinello, A. Martorana, A. Lauria, A. M. Almerico, B. K. Keppler and G. Barone, *J. Inorg. Biochem.*, 2016, **161**, 115–121.

11 A. Terenzi, R. Bonsignore, A. Spinello, C. Gentile, A. Martorana, C. Ducani, B. Högberg, A. M. Almerico, A. Lauria and G. Barone, *RSC Adv.*, 2014, **4**, 33245–33256.

12 A. Terenzi, D. Lötsch, S. Van Schoonhoven, A. Roller, C. R. Kowol, W. Berger, B. K. Keppler and G. Barone, *Dalt. Trans.*, 2016, **45**, 7758–7767.

13 N. H. Campbell, N. H. A. Karim, G. N. Parkinson, M. Gunaratnam, V. Petrucci, A. K. Todd, R. Vilar and S. Neidle, *J. Med. Chem.*, 2012, **55**, 209–222.

14 C. Zhou, T. Liao, Z. Li, J. Gonzalez-Garcia, M. Reynolds, M. Zou and R. Vilar, *Chemistry*, 2017, **23**, 4713–4722.

15 R. Bonsignore, A. Terenzi, A. Spinello, A. Martorana, A. Lauria, A. M. Almerico, B. K. Keppler and G. Barone, *J. Inorg. Biochem.*, 2016, **161**, 115–121.

16 S. Bandeira, J. Gonzalez-Garcia, E. Pensa, T. Albrecht and R. Vilar, *Angew. Chemie - Int. Ed.*, 2018, **57**, 310–313.

17 P. Wu, D. L. Ma, C. H. Leung, S. C. Yan, N. Zhu, R. Abagyan and C. M. Che, *Chem. - A Eur. J.*, 2009, **15**, 13008–13021.

18 S. Ghosh, O. Mendoza, L. Cubo, F. Rosu, V. Gabelica, A. J. P. White and R. Vilar, *Chem. - A Eur. J.*, 2014, **20**, 4772–4779.

19 X. H. Zheng, H. Y. Chen, M. L. Tong, L. N. Ji and Z. W. Mao, *Chem. Commun.*, 2012, **48**, 7607–7609.

20 B. Bertrand, L. Stefan, M. Pirrotta, D. Monchaud, E. Bodio, P. Richard, P. Le Gendre, E. Warmerdam, M. H. De Jager, G. M. M. Groothuis, M. Picquet and A. Casini, *Inorg. Chem.*, 2014, **53**, 2296–2303.

21 C. Bazzicalupi, M. Ferraroni, F. Papi, L. Massai, B. Bertrand, L. Messori, P. Gratteri and A. Casini, *Angew. Chemie - Int. Ed.*, 2016, **55**, 4256–4259.

22 L. Stefan, B. Bertrand, P. Richard, P. Le Gendre, F. Denat, M. Picquet and D. Monchaud, *ChemBioChem*, 2012, **13**, 1905–1912.

23 D. M. Engelhard, L. M. Stratmann and G. H. Clever, *Chem. - A Eur. J.*, 2018, **24**, 2117–2125.

24 M. Kanti Si, A. Sen and B. Ganguly, *Phys. Chem. Chem. Phys.*, 2017, **19**, 11474–11484.

25 F. Moraca, J. Amato, F. Ortuso, A. Artese, B. Pagano, E. Novellino, S. Alcaro, M. Parrinello and V. Limongelli, *Proc. Natl. Acad. Sci.*, 2017, **114**, E2136–E2145.

26 J. Dai, C. Punchihewa, A. Ambrus, D. Chen, R. A. Jones and D. Yang, *Nucleic Acids Res.*, 2007, **35**, 2440–2450.

27 D. Wei, J. Husby and S. Neidle, *Nucleic Acids Res.*, 2015, **43**, 629–644.

28 A. Kascatan-Nebioglu, A. Melaiye, K. Hindi, S. Durmus, M. J. Panzner, L. A. Hogue, R. J. Mallett, C. E. Hovis, M. Coughenour, S. D. Crosby, A. Milsted, D. L. Ely, C. A. Tessier, C. L. Cannon and W. J. Youngs, *J. Med. Chem.*, 2006, **49**, 6811–6818.

29 A. Böttcher, D. Kowerko and R. K. O. Sigel, *Biophys. Chem.*, 2015, **202**, 32–39.

30 A. De Almeida, A. F. Mó, D. Wragg, M. Wenzel, P. Kavanagh, G. Barone, S. Leoni, G. Soveral and A. Casini, *Chem. Commun.*, 2017, 3830–3833.

31 N. Smargiasso, V. Gabelica, C. Damblon, F. Rosu, E. De Pauw, M.-P. Teulade-Fichou, J. A. Rowe and A. Claessens, *BMC Genomics*, 2009, **10**, 362.

32 and D. J. F. G. W. T. M. J. Frisch, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A, 2013.

33 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindah, *SoftwareX*, 2015, **1**–**2**, 19–25.

34 S. Nosé, *Mol. Phys.*, 1984, **52**, 255–268.

35 M. Parrinello and A. Rahman, *Phys. Rev. Lett.*, 1980, **45**, 1196–1199.

36 G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, *Comput. Phys. Commun.*, 2014, **185**, 604–613.

Chapter 9 - Conclusions and future work

9.1 Conclusions and Perspectives

The growth of molecular modelling and free-energy calculations for drug discovery and development have provided a platform for the work undertaken here. The known affinity of certain metals for binding biological targets, such as the propensity of gold ions to coordinate to sulfur donors in amino acid residues, has guided our research towards the characterisation of the interactions between metal-based compounds and proteins or non-canonical DNA structures. Such interactions result in either protein inhibition or DNA stabilization, respectively, which are essential to achieve the desired metallodrugs' pharmacological effects. Thus, the ability of MD and metadynamics approaches to calculate microsecond timeframes of our biological targets has allowed for greater understanding of these mechanisms at a molecular level.

This thesis work is organized into two main sections. In the first one, we have studied the mechanism of permeation and inhibition of human aquaglyceroporins (AQPs) *via* a number of molecular dynamic techniques, such as umbrella sampling and metadynamics. AQPs are essential membrane proteins involved in cell homeostasis as well as in different pathologies, including cancer. Thus, the effects of one of the most potent AQP3 inhibitors, an Au(III) compound, on the protein conformation upon metal binding to Cys40 were investigated using steered molecular dynamics. Furthermore, with the aim of identifying key amino acids involved in the physiological transport and regulation mechanisms of aquaglyceroporins, two additional computational studies have been conducted: i) on the pH induced gating mechanism of human AQP7 and its effect on both water and glycerol permeation; ii) the mechanisms of glycerol and water permeation *via* human AQP3.

In the second part of the project, we have focused our research on the understanding of the mechanisms of interactions between organometallic Au(I) complexes and non-canonical G4 DNA structures with pharmacological relevance. Thus, the stabilising

effect of Au(I) N-heterocyclic carbenes (NHCs) towards different G4s were evaluated using metadynamics and compared to FRET DNA melting measurements to determine the free-energy of binding and the mechanism of interaction.

Based on this overall structure, the thesis is organized into different chapters, including an introductory one (chapter 1) and one including the main aims of the work (chapter 2).

While chapter 3 provides a specific introduction on the topic of AQPs and their relevance to physiology and disease mechanisms, as well as information on the state-of-the-art in terms of AQPs inhibition by gold complexes and molecular dynamics studies, in chapter 4 we have evaluated the effect of the Au(III) complex Au(PbImMe)Cl_2]PF_6 on the permeation of glycerol *via* AQP3.
A centre of mass (COM) pull code was used to encourage four glycerol molecules to pass through the AQP3 tetramer pores. Two models were used, and the results compared. The first model, constituted by the 'native' human AQP3 tetramer, was used as the control and the parameters set to consistently calculate the passage of all four glycerol molecules through the tetramer (one glycerol per pore). The second model contained the [Au(PbImMe)Cl_2]PF_6 directly bound, after detachment of a chloride ligand, to the sulfur of the Cys40 residue of one monomer of AQP3. The same pull parameters were used in both cases and the effects of the compound binding on the protein structure and permeation were observed. The calculation showed interesting distortions of the protein conformation, most notably Arg218 was prevented from H-bonding with the protein backbone and pushed outward into the channel, narrowing the pore and preventing both water and glycerol conductance. Further analysis showed changes to the extracellular surface of protein channel upon binding of the Au(III) compound, increasing the exposure of hydrophobic side-chains which are likely to affect how glycerol approached the pore entrance. Overall the MD calculations within this study showed that the binding of [Au(PbImMe)Cl_2]PF_6 did not cause direct steric hindrance of the pore, preventing glycerol and water conductance, but actually induced protein conformational changes which indirectly affected the substrate's approach and passage through AQP3.

In chapter 5 we investigated the effect of pH change on human AQP7. Experimental results showed AQP7 ‘gating’ from fully permeable at physiological pH or higher to virtually closed at acidic pH, suggesting that changes in the extracellular environment can have a dramatic change on solute permeability. The calculations were performed on two AQP7 model systems. The first system contained an AQP7 tetramer protonated at pH 7, the second system contained an AQP7 tetramer protonated at pH 5. Analysis of the models showed a difference of just four residues protonated at pH 5 which are not at pH 7 (residues Glu40, His92, His140 and Glu202). As before, a centre of mass (COM) pull code was used to encourage four glycerol molecules to permeate through the aquaporin tetramer of both the pH 5 and pH7 models. A number of longer simulations were also performed to determine changes in water conductance between the two systems. The result showed a significant reduction of both glycerol and water permeation between the pH 7 and pH 5 model, in agreement with the experimental results.

Further analysis of the models showed no significant reduction in average pore size between the two models, ruling out any conformational changes between the proteins in the channel region. The main differences were highlighted in the electrostatic surface of the tetramer, with an overall increase in positive charge intensity of the extracellular surface and a reversal of charge, from negative to positive, on the intracellular surface at pH 5. As the surface interactions are an important factor in the conductance rate of solutes through the pore, the observed pH changes may affect the hydrogen bond network between glycerol and the protein, causing the reduction of glycerol permeation at pH 5, rather than an open/closed gating mechanism.

In chapter 6, we focused on the investigation of the mechanism of AQP3 permeation by water and glycerol using both an established potential of mean force technique in umbrella sampling and a state-of-the-art free energy method, metadynamics. The main focus of the chapter is the use of metadynamics to calculate the free energy for the conductance of water and glycerol over a physiologically relevant timeframe in which the solute molecules can explore the entirety of the model. Results show the interactions between the protein residues lining the pore surface, water and glycerol during both uptake and efflux. Hydrogen bond residence time was used to ascertain these interactions, showing the pathway taken by glycerol and water as they pass through the protein pore and highlighted the importance of the ar/R and NPA

selectivity filters and the role they play. Free energy surface results show the NPA electrostatic barrier to be the more energetically costly rather than the steric ar/R selectivity filter, as previously discussed. It should be noted that the NPA motif is responsible for the established flipping mechanism observed within the water chain. Therefore, it is not surprising that this region is the one with the highest energy barrier. A third area of interest was observed within the pore around His81, located just below the NPA motif. Both the free energy and hydrogen bond data show both solutes to have a strong interaction with His81 and the surrounding residues. The simulations also show the bi-directional behaviour of water through the pore at equilibrium conditions and the importance of water in glycerol transportation.

Ongoing work

Further ongoing work on AQPs concerns elucidating the transport mechanism of H_2O_2 *via* human AQP3. In this case, a combination of both biased and unbiased *in silico* techniques was used to study the permeation of H_2O_2 through the pore. H_2O_2, a reactive oxygen species (ROS) known for causing cellular damage when overproduced, is thought to induce a signalling cascade leading to cell migration and metastasis in certain cancer cell lines, with AQP3 knockout studies in breast cancer cell lines showing reduced H_2O_2 uptake[1]. By knowing the permeation pathway of H_2O_2 along with the energies involved in this substrate's conductance, we hope to be able to identify key residues for this substrate transport which may support the design of highly selective inhibitors.

Initial studies involved using the established umbrella sampling technique to investigate H_2O_2 permeation *via* an homology model of AQP3, based on the structure of GlpF (pdb 1FX8). The potential of mean force (PMF) results is shown in figure 9.1 for both uptake and efflux pathways. An average ΔG value of 7.5 kJ mol can be calculated for a permeation event.

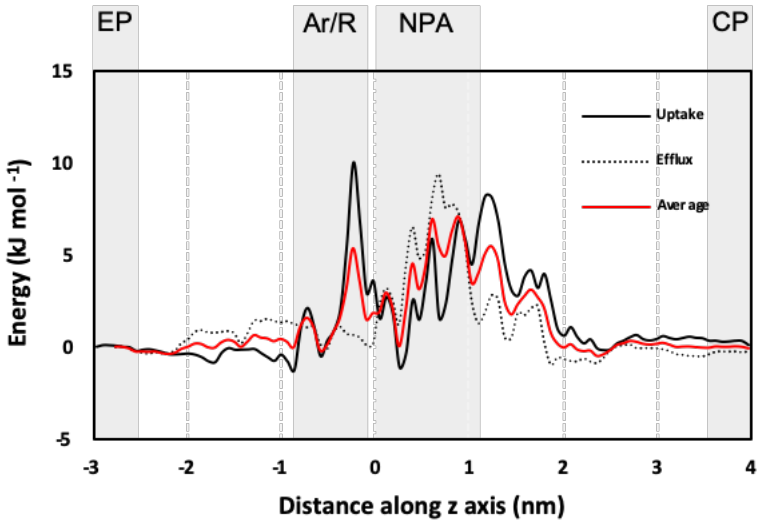


Figure 9.1. Free-energy of H_2O_2 uptake (solid line) and efflux (dashed line), calculated using umbrella sampling. Averaged curve shown in red.

Whilst umbrella sampling provides an insight into the energies involved, as previously discussed, this method does not relevant mechanistic data. Therefore, we have applied two further approaches to explore the system at a more physiological timeframe. The methods chosen were transition path sampling (TPS), to provide unbiased trajectory information, and metadynamics to provide both free energy and mechanistic trajectories of conductance events.

TPS calculations are currently ongoing with the model built using a single AQP3 monomer and one H_2O_2 molecule in a solvated system. As we are focusing the calculations on the pathway, having a smaller system allows for greater efficiency of the calculation.

The metadynamics calculations consist of a full AQP3 tetramer embedded in a POPC lipid bilayer, as was also used for the umbrella sampling. Multiple 200 ns simulations have been run to capture as many permeation events as possible. In all 15 uptake and 14 efflux events were successfully observed from a total of 1.2 μs of combined simulation time. The averaged free-energy results for both uptake and efflux can be seen in figure 9.2. Whilst the free-energy curve calculated by umbrella sampling and metadynamics do not match energetically, they do contain similar motifs in the calculated averages with the NPA region shown to be the highest energy barrier during H_2O_2 conductance. The calculated absolute ΔG values for H_2O_2 match the trend seen

with glycerol and water permeation in chapter 6, with H_2O_2 values between those of glycerol and water (Table 9.1).

Table 9.1. Free energies ΔG (kJ/mol^{-1}) for water, glycerol and H_2O_2 permeation through the NPA filter, for both uptake and efflux, calculated by metadynamics. Data shown as mean ± SEM. *n* = number of simulations. Data are calculated from the absolute ΔG of each successful permeation event and averaged.

	ΔG (kJ/mol^{-1})		
	Water	Glycerol	H_2O_2
Uptake	26 ± 5 (*n* = 7)	40 ± 4 (*n* = 8)	29±10 (*n*=15)
Efflux	21 ± 5 (*n* = 14)	35 ± 10 (*n* = 10)	37±9 (*n*=12)

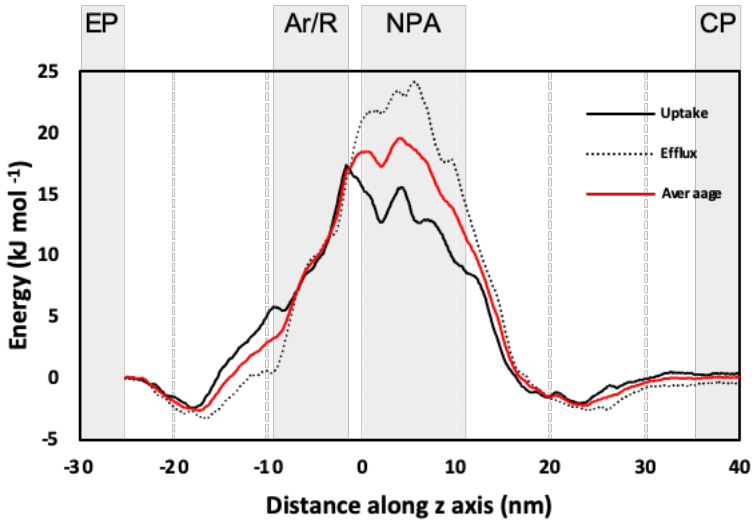


Figure 9.2. Average free-energy surface of H_2O_2 uptake (solid line) and efflux (dashed line), calculated using metadynamics sampling. Averaged curve shown in red.

Analysis of the H-bond network for H_2O_2 permeation shows similar interactions within the ar/R and NPA regions as seen with both water and glycerol. Again, the previously seen long residence times (ca. 1.1 ns for water and ca. 2.1 ns for glycerol) with the highly conserved His81 of ca. 2.0 ns, located below the NPA motif, are also observed. The main differences are evidenced in the extracellular pocket, where less defined interactions are observed for hydrogen peroxide in comparison to glycerol, and a different approach pathway is also noticed. This could be due to the similarity between H_2O_2 and water, meaning hydrogen peroxide is able to insert into the water

chain more easily than glycerol and have less of a detrimental effect on the bidirectional water flux.

Preliminary results of this study provide an insight into H_2O_2 conductance and are consistent with our previous on glycerol and water. The TPS analysis will, we hope, increase our understanding of the conductance pathway.

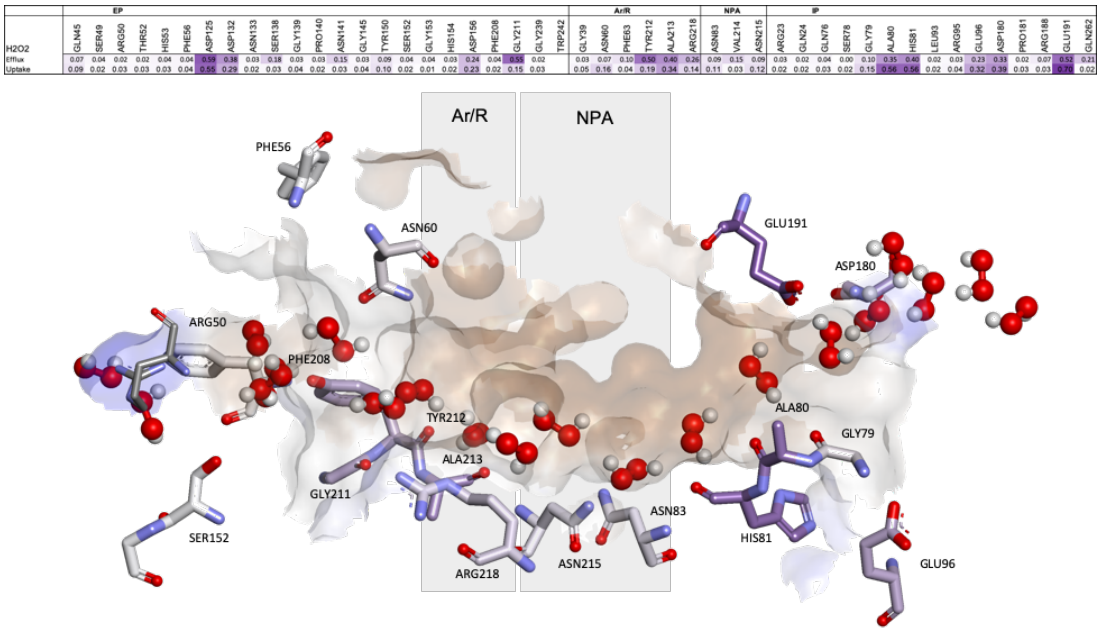


Figure 9.3. H_2O_2 permeation routes and H-bond patterns and RT (%) from metadynamics calculations. Average H-bond RT (%) of glycerol during uptake and efflux (on a scale of 0-0.8%) shown in a gradient purple colour, with the strongest colour indicating the highest RT. Multiple H_2O_2 molecules snapshots, taken from one representative simulation, are overlaid in one structure to create one single path. Amino acids that form crucial H-bonds are explicitly shown. Carbons are colour-mapped (white to purple) according to their corresponding RT (short to long). Pore colour representation based on hydrophobicity of the pore surface, blue = hydrophilic, brown = hydrophobic.

Concerning the second part of the thesis, chapter 7 provides a concise introduction on the structure and function of G4 DNA structures, as well as on the most representative families of G4s stabilizers, including metal coordination complexes. In chapter 8, we used metadynamics to calculate the binding affinities of two Au(I) N-heterocyclic carbene (NHC) complexes with two G-quadruplex model sequences, namely hTelo, a human telomeric sequence, and *C-KIT1*, a promoter sequence.

By calculating the binding affinities of the Au(I) NHC complexes with the G4 models we could ascertain which compound had the greater effect on G4 stabilisation by combining the *in silico* work with experimental FRET DNA melting assays. Through

mathematical modelling we were able to calculate the experimental ΔG value from the FRET data and have a direct comparison to the metadynamics calculations. Remarkably, the calculated binding free-energy values perfectly matched the experimental ones. As metadynamics allows the calculation of further metastable states, and not only the main free energy minima, we observed how the binding mode of the bis-carbene Au(I) complex with *h*Telo is a two stage process, as also hypothesized based on the FRET data. In addition, we were able to associate different binding modes of the two compounds with the G4s structures in correspondence to the different energy minima. Further studies will require the validation of these binding modes via different experimental techniques, such as NMR and CD spectroscopies.

Overall, throughout this work we have shown how the use of *in silico* methods can enhance the knowledge and understanding of biological systems, helping validate and explain experimental results, and vice versa. We have used state of the art techniques, either rarely or not yet used for such complex systems, furthering the knowledge of our group, and through published work, the wider scientific community.

9.2 Future work

The work carried out here has provided a proof of concept for a number of biological systems and computational methods. In particular the use of metadynamics with large, complex transmembrane proteins and the elucidation of their mechanism of permeation. To further validate our approach, this work should be expanded to cover other aquaporin isoforms such as AQP8 and AQP10, to ascertain if the permeation mechanism we have observed in AQP3 is true for these isoforms.

Previous work by Wambo et al (2017)[1] have shown the osmotic driving force of aquaporin solute conductance. In this work a large system consisting of four GlpF tetramers embedded into a POPE lipid bilayer was constructed. A hybrid periodic boundry conditions (PBC) was used whereby a ridged boundary was implemented parallel to the membrane to prevent solvent mixture along the z-axis unless transported *via* one of the 16 pores. A different salinity was modelled in each of the solvated sites

by adding an additional salt concentration to the extracellular side, thus creating an osmotic gradient. Previous studies involved simulating an osmotic gradient by applying hydraulic pressure to the system[2]. The method employed by Wambo et al (2017) is a more elegant solution as it does not bias the system and therefore allows for more reliable free energy results. Taking this work as inspiration, we proposed an alternative model to study the effect of osmotic gradients created by solutes such as glycerol and H_2O_2 concentration on the extracellular and intracellular sides of an aquaporin. In this case a large stacked double membrane/AQP system providing distinct solvated regions capable of mixing only *via* pore permeation could be built whilst still maintaining PBC. This would allow us to simulate the intra and extracellular regions, and the osmotic effects, more accurately based on biological conditions used experimentally. Depending on where the solute molecules are placed in the system it would be possible to model osmotic shocks on aquaporins and help to better understand their bidirectional nature. This system would allow us to calculate competition experiments between permeable solutes such as glycerol, H_2O_2 and urea. Whilst still being a simplified, and therefore, a manageable system, it would allow for a greater insight into the role played by aquaporins in the redistribution of solutes due to osmotic pressures.

The further parametrisation of metals for use with force fields and the refining of Au(I) and Au(III) would allow the evolution of the growing catalogue of organometallic and coordination complexes to be evaluated *in silico*. Taking the work with these compounds to a higher theory level, such as density function tight binding DFTB[3]. Whilst being more established for materials chemistry calculations, this method has proved successful at calculating the conformational properties of a number of biomolecules such as peptides and DNA bases[4]. Work undertaken on this method within our group has led us to believe DFTB could allow us to further incorporate and understand the complex electronic interactions between organometallic compounds and their biological targets and thus, enhance our calculations and knowledgebase.

9.3 References

1 T. O. Wambo, R. A. Rodriguez and L. Y. Chen, *Biochim. Biophys. Acta - Biomembr.*, 2017, **1859**, 1310–1316.

2 F. Zhu, E. Tajkhorshid and K. Schulten, *Biophys. J.*, 2002, **83**, 154–160.

3 C. M. Goringe, D. R. Bowler and E. Hernández, *Reports Prog. Phys.*, 1997, **60**, 1447–1512.

4 M. Gaus, Q. Cui and M. Elstner, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2014, **4**, 49–61.

9.4 Publication List

1. De Almeida, A. F. Mósca, **D. Wragg**, M. Wenzel, P. Kavanagh, G. Barone, S. Leoni, G. Soveral and A. Casini, The mechanism of aquaporin inhibition by gold compounds elucidated by biophysical and computational methods, Chem. Commun., 2017, 53, 3830–3833. DOI: 10.1039/C7CC00318H (In this thesis)

2. **D. Wragg**, A. de Almeida, R. Bonsignore, F. E. Kühn, S. Leoni and A. Casini, On the mechanism of gold NHC compounds binding to DNA G- quadruplexes elucidated by combined metadynamics and biophysical methods, Angew. Chemie - Int. Ed., 2018, 57, 14524–14528. DOI: 10.1002/anie.201805727 (In this thesis)

3. L. Rocard, **D. Wragg**, S. A. Jobbins, L. Luciani, J. Wouters, S. Leoni and D. Bonifazi, Templated chromophore assembly on peptide scaffold: structural evolution of the system, Chem. - A Eur. J., 2018, 24, 16136–16148. DOI: 10.1002/chem.201803205

4. Mósca, A. de Almeida, **D. Wragg**, A. Martins, F. Sabir, S. Leoni, T. Moura, C. Prista, A. Casini and G. Soveral, The pH regulation of human aquaporin-7 investigated by biophysical and molecular dynamics approaches, Cells, 2018, 7, 207. DOI: 10.3390/cells7110207 (In this thesis)

5. **D. Wragg**, A. de Almeida, A. Casini and S. Leoni, Unveiling the mechanisms of aquaglyceroporin-3 glycerol permeation by metadynamics, Chem. – A Eur. J., 2019, 25, 8713–8718. DOI: 10.1002/chem.201902121 (In this thesis)

In preparation

D. Wragg, J. Rundle, E. McShane, Angela Casini and Stefano Leoni, Understanding the mechanism of H_2O_2 conductance *via* aquaglyceroporin-3 by advanced molecular dynamic techniques.

9.5 Conference List

1. Chemistry: the science around us Symposium. Sheffield University. Sheffield, UK

Poster contribution: MD studies of H_2O_2 permeation *via* Aquaporin-3. (09/09/16)

2. Symposium on interdisciplinary collaborations. Life Sciences Hub Wales.

Oral contribution: MD studies of Hydrogen peroxide and glycerol permeation of AQP3. (24/11/16)

3. Conference on Atomistic Simulations of Biomolecules: towards a Quantitative Understanding of Life Machinery. the Abdus Salam International Centre for Theoretical Physics (ICTP) Trieste, Italy.

Oral and poster contributions: Molecular dynamic studies of H_20_2 permeation *via* Aquaporin-3. (06/03/17 – 10/03/17)

4. 16^{th} Annual Cardiff Chemistry Conference. Cardiff University. Cardiff, UK.

Poster Contribution: The Mechanism of Aquaporin Inhibition by Gold Compounds Elucidated by MD Simulations. (15/05/17)

5. Southwest Structural Biology Meeting, Cardiff University.

Poster Contribution: The Mechanism of Aquaporin Inhibition by Gold Compounds Elucidated by MD Simulations. (14/09/17)

6. TUM-IAS workshop “Inorganic Chemistry meets Medicine. Die Technische Universität München (TUM), Munich, Germany.

Oral contribution: Advanced molecular dynamics approaches for drug discovery. (24/10/17 – 26/10/17)

7. Structure Activity relationships. UK QSAR/MGMS meeting, Cardiff University, Cardiff, UK.

Poster contribution: On the mechanisms of DNA G-quadruplexes targeting by gold(I) N-heterocyclic compounds: new insights by combined meta-dynamics and biophysical methods. (11/04/18 – 12/04/18)

8. 17th Annual Cardiff Chemistry Conference. Cardiff University. Cardiff, UK.

Oral contribution: On the use of advanced molecular dynamics methods to design drugs targeted to non-canonical DNA structures. (15/05/18)

9. EuroBIC 14, University of Birmingham, Birmingham, UK.

Poster contribution: On the mechanisms of DNA G-quadruplexes targeting by gold(I) N-heterocyclic compounds: new insights by combined meta-dynamics and biophysical methods. (26/08/18 – 30/08/18)

10. Milestones in Molecular Simulations, Imperial College London, London, UK

Poster contribution: On the mechanisms of DNA G-quadruplexes targeting by gold(I) N-heterocyclic compounds: new insights by combined meta-dynamics and biophysical methods. (21/06/19)