

Computer-aided design, synthesis and evaluation of novel anti-chikungunya and anti-enterovirus compounds

A thesis submitted in accordance with the conditions governing
candidates for the degree of
Philosophiae Doctor in Cardiff University

by

Birgit Eva Christine Zonsics

June 2019

Cardiff School of Pharmacy and Pharmaceutical Science
Cardiff University

Statements and Declaration

STATEMENT 1

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

Signed (candidate) 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 it being submitted concurrently for any other degree or award (outside of any formal collaboration agreement between the University and a partner organisation)

Signed (candidate) Date

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available 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 (candidate) 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 are acknowledged by explicit references. The thesis has not been edited by a third party beyond what is permitted by Cardiff University's Use of Third Party Editors by Research Degree Students Procedure.

Signed (candidate) Date

WORD COUNT 59,874 words

(Excluding summary, acknowledgements, declarations, content pages, appendices, tables, diagrams and figures, references, bibliography, footnotes and endnotes)

Summary

RNA viruses present a large group of viruses that contains many important human pathogens. Chikungunya virus is an Alphavirus transmitted by tiger mosquitoes, causing a febrile disease that often leads to very disabling, sometimes chronic, joint and muscular pain that can last for several weeks up to months. The *Picornaviridae* family including enterovirus A71, coxsackievirus B3, poliovirus, enterovirus D68 and rhinoviruses cause various different clinical symptoms and diseases like hand-foot-and-mouth disease, poliomyelitis, or the common cold. For none of these viruses direct-acting antivirals are on the market yet, stressing the need to design novel compounds that could target these viruses and that may enter into (pre-)clinical development soon.

The replication cycle of RNA viruses requires specific viral proteins that replicate the viral genome and fulfil other crucial functions within the host-cell but are not packed into new viral particles. These non-structural proteins present excellent targets to inhibit the viral replication and were therefore investigated using computer-aided techniques in order to find novel antiviral compounds. Pharmacophore screening and docking were used to select molecules from large chemical libraries that were then tested in cell-based antiviral assays for their activities. Then the compounds were synthesised and improved using classic medicinal-chemistry modifications. For chikungunya several different compounds with low micromolar activity could be identified.

For the picornaviruses several inhibitors were reported, but the exact mode-of-action on their molecular target (2C protein) was unknown. Possible sites and interactions were explored using site identification tools, docking and molecular dynamics simulations. In collaboration with virologists and structural biologists this led to the clarification of the mode-of-action of fluoxetine, which exhibits a stereoselective activity on 2C. In addition, a series of novel inhibitors with broad-spectrum activity against the described picornaviruses was developed.

Acknowledgements

The three years of PhD were an interesting journey that has now come to an end. Here I am reflecting on that journey and find myself very grateful for the experiences I have made and the people that made this possible and shared this very intense and special time with me.

I would first like to thank my supervisor Prof Andrea Brancale, who gave me the possibility to join his group and assigned me to a great research project that occupied and challenged me for the last three and a half years. I am very grateful for his patience and his kind and calm attitude that gave me room to grow to become independent and confident on the way. I also want to thank him for his constant support when we started a new collaborative research project on top of the assigned one. It was a great and absolutely invaluable experience to initiate such a project from scratch, to collaborate with peers and experienced researchers equally and to get a glimpse of how academic research reality can work. I would also like to thank Andrea for allowing me to attend all the great conferences I attended, for involving me in the organisation of the Antivirals network meeting in Cardiff and for sending me to several outreach and public engagement events. His trust and belief in me encouraged me a lot.

I would like to thank the European Union, in particular the Marie Skłodowska Curie Actions under grant agreement No 642434, for funding my PhD within the Antivirals ETN. For the organisation of the network and the excellent trainings my thanks go particularly to Prof Frank van Kuppeveld and Dr Clazien Oomen. I feel furthermore blessed that Prof Bruno Coutard was assigned to me as a mentor during the time of the network. I would like to thank him for the possibility to work in his laboratory during a secondment and I am extremely grateful for the feedback, the long discussions and the collaboration with him. He always took his time to support me and to challenge my critical and independent thinking. Furthermore, I would like to thank Dr Marie-Louise Jung and the team at Prestwick Chemical for giving me the opportunity to pass a secondment in their company. The insights into a contract research organisation in the field of chemistry and drug discovery were very useful to broaden my horizon and I keep the time there in very good memory. My gratitude is also owed to Prof Thierry Langer from the University of Vienna, who made me aware of the Antivirals network, encouraged me to apply and kept in contact checking on me throughout the whole time, without him the last three years might have taken a very different turn.

I would like to thank all my collaborators during the different parts of the projects. In particular I want to thank the team of Bruno Coutard, especially my peer Ana-Sofia Ferreira Ramos, for the help during my secondment in Marseille and the collaboration on CHIKV macro domain. I thank Prof. Johan Neyts, Dr Leen Delang and Dr Rana Abdelnabi for testing my compounds at

KU Leuven and for the discussions and their excellent scientific advice. I would like to thank Prof Gilles Querat (AMU) and Sofie Jacobs (KU Leuven) for testing my compounds on a panel of different viruses.

I would like to thank all the participants of the 2C project for the exciting journey that we pursued and the excellent collaboration and exchange. It is incredible what is possible with a strong team! Thanks to Lisa Bauer, who presented the idea to work on the topic and Roberto Manganaro for working side-by-side with me on the computational and chemical part of the project. I want to thank Prof Frank van Kuppeveld, Prof Bruno Coutard and Prof Andrea Brancale once more for extensive exchange, incredible support and, above all, kindness and patience and to help us to grow into passionate scientists even more.

Moving to a different country for the duration of three years is a difficult decision and although the head says, “this is important and it’s good for your career”, the feelings can sometimes be overwhelming. I want to thank everybody who contributed to my Cardiff family. First of all, Martin who answered patiently all my questions before coming to Cardiff, who took it upon himself to show me around, and to help me with literally everything I encountered and didn’t have a solution myself. He was an endless source of information and care and I am very grateful to have him as a friend. Apart from Martin, I would like to thank all the members of Brancale’s lab, Marcella and Salvo who were always there for help and guidance and both great role models for me. My thanks also go to Cinzia, who still found time during the end of her own PhD project to help me with my questions and who was always there for me, in the lab and as a friend. Thanks to Roberto, my Antivirals sibling, with whom I went to many conferences and network meetings, with whom I could rant about the bureaucracy of the network and coordinate on all important deadlines. Thank you for challenging me but especially for becoming a really great collaborator and friend.

I would like to thank “the Italians” Gilda, Elisa, Gaia, Carmine, Giulio, Cecilia, both Alessandras, Silvia L, Massimo, Laura, Alberto, Antonio, Fabio, Vito and all the Erasmus students that were part of or associated to our big group during the three years for their persistence in being a critical mass, thus making me learn Italian in record time. I had great times with all of you and it made me feel like being in an Italian family.

Furthermore, would like to thank everybody in the School of Pharmacy and Pharmaceutical Sciences, Redwood building, both the scientific and administrative staff for the events they organised, the help they gave me, and the lovely working atmosphere. I will come back and visit in the future.

I am very happy to have met everybody in the Antivirals network, especially my 14 peers. Erion, Lisa, Kristina K, Natacha, Melissa, Ania, Ana Sofia, Nhung, Kristina L, Marion, Angelica, Hari, Clara and Roberto thank you for the great meetings, the endless chats, the

collaborations, the scientific exchange, the spontaneous unofficial network meetings, the conferences together, the assignments, the joy and the difficulties we mastered together. This network was amazing because of all of you! Big thanks also to the supervisors who took a lot of time out of their busy schedules to dedicate it to our development and training. Thank you all for sharing your experience, your career pathways, your scientific journeys, and for giving advice to the next generation of antiviral drug hunters.

I would also like to thank my friends, both from my home Vienna and those I found in Cardiff. I want to thank Silvia, Alessandra, Chiara, Elin and Hannah very deeply for becoming my friends and sharing great moments during the times in Wales. I want to thank the friends from Vienna who came to visit me in Cardiff for taking part in my PhD experience and those who never stopped calling me for their constant support and those who welcomed me back as if I had never been away, for being my safety net.

I would like to thank Riccardo for showing up on that Friday at the Pen and Wig, for the time that we spent together and forgot about both our work, the holidays and the hard times that we needed to spend apart. Thanks for believing in me, trusting me and supporting me in whatever I do.

Last but not least I would like to thank my family. I thank my parents, Margit and Erich, for encouraging me to start this journey and letting me go for such a long time, for coming and visiting me, sending me care packages helping me with many decisions even from afar. Thanks to my grandmother for getting a smartphone and learning how to use it, it meant a lot to me to stay in contact and to have you participate! I want to thank her also for supporting me financially in the difficult times of being on the move and writing up. Without all of them this PhD would not have been possible, their support kept me going and they cared for me especially during my writing period. I hope I can make them proud and thank them for their encouragement, their endless patience, their understanding and the curiosity they also passed on to me.

Table of Contents

DECLARATION AND STATEMENTS	III
SUMMARY	V
ACKNOWLEDGEMENTS	VII
TABLE OF CONTENTS	XI
LIST OF ABBREVIATIONS	XIX
AMINO ACID ABBREVIATIONS	XXI
LIST OF FIGURES	XXII
LIST OF TABLES	XXV
LIST OF EQUATIONS	XXVII
1 INTRODUCTION	1
1.1 ANTIVIRAL RESEARCH AND ANTIVIRAL DRUG DISCOVERY	3
1.1.1 HISTORY	3
1.1.2 VIRUS CLASSIFICATION	4
1.1.3 THE VIRAL REPLICATION CYCLE	5
1.1.4 ANTIVIRAL THERAPY	6
1.1.5 VACCINATION	8
1.2 DRUG DISCOVERY AND DESIGN - A COMPLEX PROCESS	9
1.3 MOLECULAR MODELLING	9
1.3.1 FUNDAMENTAL CONCEPTS IN MOLECULAR MODELLING	9
1.3.1.1 Molecular modelling beyond molecular representation	10
1.3.1.2 Quantum mechanics	11
1.3.1.3 Molecular mechanics and force field methods	12
1.3.1.4 Energy minimization	13
1.3.1.5 Conformation generation	14
1.3.2 APPLICATIONS OF MOLECULAR MODELLING	14
1.3.2.1 Homology modelling	15
	XI

1.3.2.2	Pharmacophore modelling	16
1.3.2.3	Docking and Scoring	17
1.3.2.4	Molecular dynamics	18
1.4	GENERAL AIMS	20
2	CHIKUNGUNYA VIRUS	21
2.1	INTRODUCTION	23
2.1.1	CLASSIFICATION AND TAXONOMY	23
2.1.2	EPIDEMIOLOGY	24
2.1.3	HOST	24
2.1.4	CHIKUNGUNYA FEVER	25
2.1.5	TREATMENT OPTIONS	25
2.1.6	DESCRIPTION OF THE VIRION	26
2.1.7	GENOME ORGANISATION	26
2.1.8	CHIKUNGUNYA REPLICATION CYCLE	28
2.1.9	VIRAL PROTEINS AS POTENTIAL TARGETS	29
2.1.9.1	Structural Proteins	29
2.1.9.1.1	Capsid	29
2.1.9.1.2	E1	30
2.1.9.1.3	E2	30
2.1.9.1.4	E3	31
2.1.9.1.5	6K	31
2.1.9.2	Non-structural Proteins	31
2.1.9.2.1	nsP1	32
2.1.9.2.2	nsP2	32
2.1.9.2.3	nsP3	33
2.1.9.2.4	nsP4	34
2.1.10	CURRENT RESEARCH ON SPECIFIC ANTIVIRALS AND VACCINES AGAINST CHIKV	35
2.1.10.1	Antivirals in development against CHIKV	35
2.1.10.1.1	Entry inhibitors	35
2.1.10.1.2	Replication inhibitors	37
2.1.10.1.3	Protein synthesis inhibitors and compounds directly targeting CHIKV proteins	39
2.1.10.1.4	Host targeting compounds against CHIKV	41
2.1.10.2	Vaccines	42
2.2	AIMS	45
2.3	TARGETING CHIKV NSP3 MACRO DOMAIN	47

2.3.1	MACRO DOMAINS	47
2.3.2	PROPERTIES OF VIRAL MACRO DOMAINS	49
2.3.3	STRUCTURE OF CHIKV MACRO DOMAIN	51
2.3.4	BINDING AND ENZYMATIC FUNCTIONS OF CHIKV MACRO DOMAIN	53
2.3.5	AIMS	54
2.3.6	RESULTS AND DISCUSSION	54
2.3.6.1	Screening on CHIKV macro domain for new small molecule inhibitors	54
2.3.6.2	Building the Pharmacophore model for nsP3 macro domain	56
2.3.6.3	Running the Pharmacophore query	58
2.3.6.4	Docking	59
2.3.6.5	Consensus scoring	59
2.3.6.6	Visual inspection	60
2.3.6.7	Biological evaluation of the selected compounds	64
2.3.6.8	Synthesis and biological evaluation of 10 and its derivatives	65
2.3.6.8.1	First modifications	66
2.3.6.8.2	Second round of modifications	73
2.3.6.8.3	Third round of modifications	75
2.3.6.9	Compound 10 – confirmation of a promising hit compound	78
2.3.6.10	Determination of the EC ₅₀ of compound 10	78
2.3.6.10.1	CPE inhibition assay	78
2.3.6.10.2	Virus yield assay	79
2.3.6.11	Assays on the purified macro domain of CHIKV	80
2.3.6.11.1	Thermal Shift Assay	80
2.3.6.11.2	Co-crystallisation	83
2.3.6.12	Cell-based mode of action studies for 10	83
2.3.6.12.1	Time of addition assay	83
2.3.6.12.2	Chikungunya virus pseudo particle entry assay	84
2.3.6.13	Tests against other viruses	85
2.3.7	CONCLUSIONS	86
2.3.8	EXPERIMENTAL	89
2.3.8.1	Chemistry	89
2.3.8.1.1	Reagents	89
2.3.8.1.2	Thin layer chromatography - TLC	89
2.3.8.1.3	Column Chromatography	89
2.3.8.1.4	Automated flash column chromatography	89
2.3.8.1.5	UPLC-MS	89

2.3.8.1.6	NMR Spectra	90
2.3.8.1.7	Mass Spectra	90
2.3.8.1.8	General Procedures Synthesis	90
2.3.8.1.9	Purification by automated flash column chromatography	91
2.3.8.2	Synthesis of the compounds	92
2.3.8.3	Biology	136
2.3.8.3.1	Cells, viruses and compounds	136
2.3.8.3.2	Cytopathic effect (CPE) reduction assay	136
2.3.8.3.3	Virus yield assay	137
2.3.8.3.4	Quantitative reverse transcription PCR (qRT-PCR)	137
2.3.8.3.5	Determination of CCID ₅₀ per ml	137
2.3.8.3.6	Delay of treatment assay	138
2.3.8.3.7	Entry assay using CHIKV pseudoparticles (CHIKVpp)	138
2.3.8.3.8	Evaluation of 10 on different Alphaviruses and on West Nile Virus	138
2.4	TARGETING THE DISTAL RIBOSE POCKET OF CHIKV MACRO DOMAIN	141
2.5	AIMS	141
2.6	METHODS	141
2.6.1	HARDWARE AND SOFTWARE	141
2.6.2	PROTEIN PREPARATION	142
2.6.3	PHARMACOPHORE QUERIES AND SEARCHES	142
2.6.4	DOCKING AND SCORING	142
2.6.5	BIOLOGICAL EVALUATION	143
2.6.6	CROSS EVALUATION AGAINST HUMAN MACRO DOMAINS	143
2.7	RESULTS	143
2.7.1	LIBRARY PREPARATION	143
2.7.2	PHARMACOPHORE MODELLING	144
2.7.3	DOCKING	148
2.7.4	CONSENSUS RESCORING	149
2.7.5	VISUAL INSPECTION	149
2.7.6	BIOLOGICAL EVALUATION	153
2.7.7	COMPUTATIONAL EVALUATION AGAINST HUMAN MACRO DOMAINS	154
2.7.8	FUTURE PERSPECTIVES	156
2.8	MODELLING THE P23 PRECURSOR PROTEIN	158
2.9	AIMS	158
2.10	RESULTS AND DISCUSSION	159
2.10.1	ALPHAVIRUS UNIQUE DOMAIN (AUD)	159

2.10.1.1	Homology model of AUD	159
2.10.1.2	Evaluation of the homology model	162
2.10.2	EXPANDING NSP3	163
2.10.3	LINKING THE P23 PRECURSOR	164
2.10.4	FUTURE PERSPECTIVES	165
2.11	THE NSP2/NSP3 CLEAVAGE SITE POCKET AS TARGET	166
2.11.1	THE P23 CLEAVAGE SITE	166
2.12	AIMS	168
2.13	RESULTS AND DISCUSSION	168
2.13.1	SITE FINDER	168
2.13.2	VIRTUAL SCREENING WORKFLOW	169
2.13.3	SELECTED COMPOUNDS	171
2.13.4	BIOLOGICAL EVALUATION	174
2.14	DISCUSSION AND CONCLUSIONS	176
2.15	EXPERIMENTAL	177
2.15.1	HARDWARE	177
2.15.2	PROTEIN PREPARATION	177
2.15.3	SITE FINDER	177
2.15.4	SPECS LIBRARY	177
3	ENTEROVIRUSES	179
3.1	INTRODUCTION	181
3.1.1	CLASSIFICATION	181
3.1.2	EPIDEMIOLOGY AND PATHOGENESIS	182
3.1.3	TRANSMISSION	183
3.1.4	CURRENT TREATMENT	183
3.1.5	GENOME ORGANISATION AND REPLICATION CYCLE	184
3.1.6	REPLICATION CYCLE OF ENTEROVIRUSES	185
3.1.7	TARGETS AND INHIBITORS	186
3.1.7.1	Attachment and entry and inhibitors thereof	187
3.1.7.2	Protease inhibitors	188
3.1.7.3	3D ^{pol} - RNA-dependent RNA polymerase inhibitors	189
3.1.7.4	2B protein	190
3.1.7.5	2C protein and inhibitors	190
3.1.7.6	3A and 3AB proteins	190
3.1.7.7	3B/VPg protein	191

3.1.7.8	Assembly inhibitors	191
3.1.7.9	Host factors as antiviral targets	192
3.2	2C PROTEIN	194
3.2.1	INTRODUCTION TO THE NON-STRUCTURAL PROTEIN 2C	194
3.2.2	CHARACTERISATION OF THE 2C STRUCTURE	195
3.2.3	INHIBITORS OF 2C PROTEIN	197
3.2.3.1	Guanidine hydrochloride	198
3.2.3.2	2-(α -Hydroxybenzyl)-benzimidazole (HBB)	198
3.2.3.3	TBZE-029 and MRL-1237	199
3.2.3.4	Compounds from repurposing screenings	199
3.2.3.5	Novel 2C targeting compounds	200
3.3	AIMS	201
3.4	INVESTIGATIONS ON THE EV71 2C CRYSTAL STRUCTURE	203
3.5	POCKET FOR THE C-TERMINAL DOMAIN	204
3.5.1	HTVS	205
3.5.2	BIOLOGICAL EVALUATION	210
3.5.2.1	Protocol for the multicycle CPE-Reduction Assay	210
3.5.3	DISCUSSION AND CONCLUSIONS	211
3.6	INVESTIGATIONS ON 2C MODELS FOR DIFFERENT ENTEROVIRUSES	212
3.6.1	HOMOLOGY MODELS OF 2C MONOMERS	212
3.6.1.1	Procedure	212
3.6.1.2	Results	214
3.6.2	STRUCTURAL MODELS OF HIGHER ORDER	216
3.6.2.1	Dimer	216
3.6.2.1.1	Homology models	217
3.6.2.1.2	Molecular dynamics	218
3.6.2.1.3	Results	219
3.6.2.2	Hexamer	219
3.6.2.2.1	SymmDock	219
3.6.2.2.2	Molecular Dynamics	221
3.6.2.2.3	Results	222
3.6.2.2.4	Discussion	222
3.7	MODE OF ACTION OF KNOWN 2C INHIBITORS	223
3.7.1	SITE FINDER	223
3.7.2	DOCKING	225
3.7.3	MOLECULAR DYNAMICS OF (R)- AND (S)-FLUOXETINE	226

3.7.4	CONFIRMATION OF THE MODELLING RESULTS	228
3.7.4.1	Biological evaluation	228
3.7.4.2	Evaluation of six fluoxetine fragments for necessary chemical features	229
3.7.4.3	Biochemical evaluation	231
3.7.4.3.1	Thermal shift assay	231
3.7.4.3.2	Isothermal titration calorimetry	232
3.7.4.3.3	Evaluation of (S)-fluoxetine against a panel of enteroviruses	233
3.7.4.3.4	Mutations in the predicted pocket render CVB3 resistant to (S)-fluoxetine	233
3.7.5	PRELIMINARY CRYSTAL STRUCTURE	236
3.7.6	SUMMARY AND CONCLUSION	237
3.8	DESIGN OF NOVEL 2C INHIBITORS WITH IMPROVED AND BROAD-SPECTRUM ACTIVITY	238
3.8.1	FLUOXETINE ANALOGUES	238
3.8.2	GUANIDINE ANALOGUES OF FLUOXETINE	240
3.8.3	<i>N</i> -BENZYL- <i>N</i> -PHENYLFURAN-2-CARBOXAMIDE COMPOUNDS THEIR ANALOGUES	241
3.9	CONCLUSIONS	245
3.10	EXPERIMENTAL	246
3.10.1	COMPUTATIONAL STUDIES	246
3.10.2	VIROLOGICAL ASSAYS	247
3.10.2.1	Cell Culture	247
3.10.2.2	Viruses	247
3.10.2.3	Single-cycle virus Infection	247
3.10.2.4	Multicycle CPE reduction Assay	247
3.10.2.5	Calculations	248
4	GENERAL CONCLUSIONS	249
5	REFERENCES	253

List of Abbreviations

(-)ssRNA	negative sense single stranded RNA	DMF	Dimethylformamide
(+)ssRNA	positive sense single stranded RNA	DMSO	Dimethyl sulfoxide
(d)	density	DNA	Desoxyribonucleic acid
(H)CMV	Human Cytomegalovirus	Don	Hydrogen bond donor
¹³ C	Carbon 13	dsDNA	double stranded DNA
¹⁹ F	Fluor 19	dt	Doublet of triplets
¹ H	Proton 1	E	Energy
2D	two-dimensional	E-9	Echovirus-9
3D	tree-dimensional	E1	Envelope protein 1
Å	Angström	E2	Envelope protein 2
aa	amino acid	E3	Envelope protein 3
Acc	Hydrogen bond acceptor	EC ₅₀	Half maximum effective concentration
ADP	Adenosine diphosphate	ECSA	Eastern, South and Central Africa (strain)
ADPG	Adenosine diphosphate glucose	EHT	Extended Hückel theory
ADPR	Adenosine diphosphate ribose	Eq.	Equivalents
ALC1	Protein amplified in liver cancer 1	ER	Endoplasmic reticulum
AMP	Adenosine monophosphate	ESI	Electro spray ionisation
AMU	Aix-Marseille Université	ETN	European training network
Ani	Anionic	EV	Enterovirus
Aro	Aromatic	EV-A71	Enterovirus-A71
ART	Adenosine ribose transferase	F	Feature
ATP	Adenosine triphosphate	F	Fragment
AUD	Alphavirus unique domain	FBS	Fetal bovine serum
AUD	Alphavirus unique domain	FCS	Fetal calve serum
BGM cells	Buffalo green monkey cells	FDA	Food and Drug Administration
BHK cells	Baby hamster kidney cells	GB/VI	Generalised Born/Volume Integral
bs	broad singulet	GDP	Guanosine diphosphate
C	Capsid	GFP	Green fluorescent protein
CC	Cell control	GTP	Guanosine triphosphate
CC ₅₀	Half maximum cytotoxic concentration	H	Enthalpy
CDC	Center for disease control	HBV	Hepatitis B virus
cDNA	Copy DNA	HCV	Hepatitis C virus
CHIKV	Chikungunya virus	HEV	Hepatitis E virus
CHIKVpp	Chikungunya virus pseudo particles	HIV	human immunodeficiency virus
CNS	Central nervous system	HSP-90	Heat shock protein 90
CPE	Cytopathic effect	HTVS	High throughput virtual screening
CPU	Central processing unit	HVR	Hypervariable region
CTD	C-terminal domain	Hz	Hertz
d	doublet	IC ₅₀	Half maximum inhibitory concentration
DCE	Dichloro ethane	IC ₉₀	Inhibitory concentration of 90%
DCM	Dichloro methane	ICTV	International Committee of Taxonomy of Viruses
dd	doublet of doublets		

IFN	Interferon	P	Pressure
IL	Interleukin	P123	Polyprotein precursor nsP1-3
IMPDH	Inositol monophosphate dehydrogenase	P1234	Polyprotein precursor for nsP1-4
IND	Investigational new drug	P2/3	Cleavage site between nsP2 and nsP3
ITC	Isothermal titration calorimetry	P23	nsP2 and nsP3 precursor
<i>J</i>	Coupling constant	P62	=PE2
kb	kilobases	PAR	polyADPribose
<i>K_d</i>	Dissociation constant	PARG	Poly adenosine ribose glycohydrolase
KU	Katholieke Universiteit (Leuven)	PARP	Polyadenosineribosepolymerase
LC	Liquid chromatography	PBZ	PAR-binding zinc-finger
logP	Octanol/Water coefficient	PCL	Prestwick chemical library
m	multiplet	PDB	Protein databank
M	Molecule ion	PDB	Protein databank
MAR	Monoadenosine ribose	PE2	Precursor of E2 and E3
MD	Molecular dynamics	pH	Pondus hydrogeni
MEM	Minimal essential medium	Poa1p	ADP-ribose 1''-phosphate phosphatase
MERS-CoV	Middle east respiratory syndrome – Coronavirus	Poly I:C	Polyinosinic : polycytidylic acid
min	Minutes	PP	Pyrophosphate
MLV	Murine leukaemia virus	ppm	Parts per million
MMFF	Merck molecular force field	pro	Protease
MoA	Mode of action	PTM	Posttranslational modification
MOE	Molecular operating environment	q	quadruplet
MOI	Multiplicity of infection	qRT-PCR	real-time quantitative polymerase chain reaction
mRNA	Messenger RNA	RC	Replication complex
MT	Methyltransferase	RdRp	RNA-dependent RNA-polymerase
MT-like	Methyltransferase like	R _f	retarding-front
MTS/PMS	(3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium inner salt)/phenazine methosulfate	RLuc	<i>Renilla</i> luciferase
MV-CHIKV	Measels-vectored CHIKV vaccine	RMSD	Root mean square deviation
MW	Molecular weight	RNA	Ribonucleic acid
N	Number	RT	reverse transcriptase
NA	Not active	RT	Retention time
Na	Sodium	s	singulet
NAD ⁺	Nicotinamide adenine dinucleotide +	SAH	S-Adenosyl-L-homocysteine
NC	Nucleocapsid	SAR	Structure activity relationship
NMR	Nuclear magnetic resonance	SARS-CoV	Severe acute respiratory syndrome – Coronavirus
NNRTIs	non-nucleosidic reverse transcriptase inhibitors	sgRNA	Subgenomic RNA
NRTIs	nucleosidic reverse transcriptase inhibitors	shRNA	Small hairpin RNA
nsP	Non-structural protein	SI	Selectivity index
nt	nucleotide	SINV	Sindbis virus
ONNV	O'nyong-nyong virus	siRNA	Small interfering RNA
OPLS	Optimized Potential for Liquid Simulations	SlogP	Calculated octanol/water coefficient
ORF	Open reading frame	SM	Starting material
XX		SP	Standard precision
		ssDNA	single stranded DNA
		STAB	Sodium triacetoxo borohydride
		t	triplet

TARG1	Terminal adenosine ribose glycohydrolase 1	vdW	van der Waals
td	triplet of doublets	VEEV	Venezuelan equine encephalitis virus
TLC	Thin layer chromatography	VLPs	Virus-like particles
TMS	Trimethylsilate	WHO	World Health Organisation
Tof	Time of flight	WNV	West Nile virus
TPA	12- <i>O</i> -tetradecanoylphorbol-13- acetate	wt	Wild type
tRNA	Transfer RNA	WWE	Domain named after conserved aa WWE
TSA	Thermal shift assay	XP	Extra precision
UPLC-MS	Ultra performance liquid chromatography – mass spectrometer	YFV	Yellow fever virus
UTP	Uridine triphosphate	ZBD	Zinc binding domain
UTR	Untranslated region	ZIKV	Zika virus
UV	Ultraviolet	Zn ²⁺	Zinc ion
V	Volume	δ	Chemical shift
VC	Virus control	ΔT _m	Shift in melting temperature
		μ	Chemical potential
		μM	Micro molar

Amino acid abbreviations

Ala	A	Alanine	Leu	L	Leucine
Asn	N	Asparagine	Lys	K	Lysine
Asp	D	Aspartic acid	Met	M	Methionine
Arg	R	Arginine	Phe	F	Phenylalanine
Cys	C	Cysteine	Pro	P	Proline
Glu	E	Glutamic acid	Ser	S	Serine
Gln	Q	Glutamine	Thr	T	Threonine
Gly	G	Glycine	Trp	W	Tryptophan
His	H	Histidine	Tyr	Y	Tyrosine
Ile	I	Isoleucine	Val	V	Valine

List of Figures

Figure 1: Drug design workflow adapted from Klebe 2009 reprinted from Zonsics 2013	10
Figure 2: Energy landscape	13
Figure 3 Pharmacophore model (PDB: 3GPO, 3GPQ)	16
Figure 4 Molecular dynamics basic algorithm	19
Figure 5: Phylogenetic tree of the different CHIKV lineages described in Weaver and Forrester, 2015.	23
Figure 6: Transmission of CHIKV zoonotic and human cycle	25
Figure 7: Chikungunya particle A) envelope B) cross section C) core	26
Figure 8: Genome organisation of CHIKV and RNA replication and translation	27
Figure 9: Schematic description of CHIKV genome replication and life cycle.	28
Figure 10: Cycle of ADP-ribosylation	48
Figure 11: CHIKV macro domains 3GPO and 3GPQ superimposed	53
Figure 12: Screening workflow pursued within this project to select compounds for biological evaluation	55
Figure 13: Crystal structures 3GPO and 3GPQ with their ligands ADP-ribose (red) and a small fragment of RNA (blue)	56
Figure 14: Pharmacophore model for 3GPO and 3GPQ	57
Figure 15 The plots show the antiviral activity and the cytotoxicity in the cells of compounds 10 and 19	65
Figure 16 The plots show the antiviral activity and the cytotoxicity in the cells of compounds 22 and 26	65
Figure 17: Compound 10 and its building blocks for synthesis	66
Figure 18: General reaction scheme for the nucleophilic substitution	66
Figure 19: General scheme for the reductive amination	67
Figure 20: Synthesis of intermediate 27 and compound 10	67
Figure 21: Compound 28	68
Figure 22: Concentration dependent antiviral activity of 10 (AB-1244) as % inhibition of the virus induced cytopathic effects	79
Figure 23: Results of the virus yield assay	80
Figure 24: Time of addition of the antiviral compounds 10	83
Figure 25: Structure of a CHIKV pseudo particle	84
Figure 26: Dose dependent inhibition of the CHIKVpp entry into BGM cells for 10 (green) compared to chloroquine (red)	84
Figure 27: Cytotoxicity curves for 10 (red) and 48 (blue)	85
Figure 28: Virus yield assays for 10 and 48	86

Figure 29: Reduction of aldehyde in general procedure 2 when attempting a coupling with electron-deficient anilines and heterocyclic amines	91
Figure 30: Screening workflow describing the course of the compound selection	144
Figure 31: Query A on the distal ribose site	145
Figure 32: Example of query A with a compound (orange) from SPECS (left) and PCL (right)	146
Figure 33: Query B on the distal ribose site	146
Figure 34: Example of query B with a compound (orange) from SPECS (left) and PCL (right)	147
Figure 35: Query C on the distal ribose site	147
Figure 36: Example of query C with a compound (orange) from SPECS (left) and PCL (right)	148
Figure 37: Compound 81 docked into 3GPO	152
Figure 38: Assay 1	153
Figure 39: Assay 2	153
Figure 40: Superimposition of 4IQY (grey), 2X47 (cyan), 4J5R (pink), 4ABK (black) onto 3GPO (blue); Pocket region in yellow; ADP-ribose in red	155
Figure 41: Schematic description of P23, nsP2 and nsP3 with available structural information	159
Figure 42: nsP2 protease and methyl-transferase like domain (3TRK in red, 4GUA in green)	160
Figure 43: nsP3 macro domain (3GPO in blue, 4GUA in green)	160
Figure 44: nsP3 alphavirus unique domain (4GUA in green)	161
Figure 45: Alphavirus unique domain Homology model in blue and SINV template in red	162
Figure 46: nsP3 Macro domain in teal, AUD in blue, Connective loops in bright green	164
Figure 47: Assembly of the precursor P2 ^{pro} 3 ^{AUD} nsP2 is depicted in red, macro domain in teal and AUD in blue. Connective loops are shown in green	165
Figure 48: CHIKV P23 precursor (nsP2 part in red, nsP3 part in blue)	167
Figure 49: P23 cleavage site pocket	168
Figure 50: P23 cleavage site pocket with the central dummy atom in green	169
Figure 51: HTVS workflow to select compounds addressing the P23 cleavage site pocket	170
Figure 52: Enterovirus genome Capsid proteins in green, non-structural proteins in purple with permission of Denise Seitner	184
Figure 53: Replication cycle of enteroviruses and a selection of crucial host factors Image with permission of Denise Seitner	185
Figure 54: Semi-schematic depiction of 2C protein Walker A (green) and B (orange) motifs, motif C (turquoise) and the C-terminal helix (purple)	195
Figure 55: Dimer interface and close-up of the ATP binding pocket	196
Figure 56: 2C Dimer oligomerising with C-terminal helix (top); C-terminal helix (bottom-left) and the corresponding pocket (bottom-right)	205
Figure 57: High-throughput virtual screening workflow (Design by Roberto Manganaro)	206
Figure 58: Superimposition of the homology models coloured by conservation of residues	215
Figure 59: Identity (left) and similarity (right) between the homology models numeration according to Table 75	215
Figure 60: Superimposition of the homology models coloured by RMSD	216

Figure 61: RMSD to the template (numeration according to Table 75)	216
Figure 62: ATP binding site between chain A and F of 5GRB In green the residues of the Walker A motif, Walker B in gold, C motif in turquoise on chain A in darker blue; R-finger(s) and residue T196 in yellow on chain F in lighter blue. ATPyS in orange surrounded by a light grey surface.	217
Figure 63: Dimer formed by chain A and F of EV-A71 and CVB3 superimposed EV-A71 5GRB (blue) CVB3 model (purple) ATPyS (orange); chain A dark coloured, chain F pastel	218
Figure 64: Screenshot of the SymmDock website	220
Figure 65: Hexamer output from SymmDock prepared for molecular dynamics simulations	221
Figure 66: Configuration of ATP between chain A and F after MD simulations	222
Figure 67: Site Finder results on the dimer of EV-A71 224-229 loop in yellow, sites as white and red spheres, site 4 and 11 are corresponding sites on different monomers	224
Figure 68: Site Finder results on dimer of CVB3 homology model ²²⁴ AGSINA ²²⁹ loop in yellow, site 2 and 7 are corresponding sites on the different monomers	225
Figure 69: Docking poses of (R)- and (S)-fluoxetine into CVB3 into site A and B	226
Figure 70: (S)-fluoxetine snapshot during molecular dynamics simulation in site A	228
Figure 71: CPE reduction assay of racemic, (S)- and (R)-fluoxetine against CVB3 Figure provided by Lisa Bauer	229
Figure 72: Single cycle assay of racemic, (S)- and (R)-fluoxetine against RLuc-CVB3 Figure provided by Lisa Bauer	229
Figure 73: Fluoxetine and its fragments (F)	230
Figure 74: CPE-reduction assay for fragment F1 (activity in black, cytotoxicity in grey)	230
Figure 75: TSA of the fluoxetine enantiomers and the racemic mixture	231
Figure 76: ITC measurements for (S)-fluoxetine and (R)-fluoxetine	232
Figure 77: Mutation studies for (S)-fluoxetine to validate the potential binding pocket Figure provided by Lisa Bauer	234
Figure 78: 2C protein with focus on site A and site B ²²⁴ AGSINA ²²⁹ loop blue, loop 175-183 pink, loop 158-163 turquoise	235
Figure 79: Close-up on site B with (S)-fluoxetine in MD1 frame 600 and MD3 frame 625 Surface colour lipophilicity (green lipophilic, pink hydrophilic), Ribbons coloured as in Figure 78	235
Figure 80: (A) Three potential binding sites in 2C (B) Superimposition of MDs on site A and site B with the preliminary crystal structure (C) Pocket view of the crystal structure with relevant residues	236
Figure 81: (S)-fluoxetine in frame 625 of MD2 interacting with D245	238
Figure 82: Compound 115 <i>N</i> -(4-Fluorobenzyl)- <i>N</i> -(4-methoxyphenyl)furan-2-carboxamide	241
Figure 83: 115 (turquoise) docked into site A after MD2 frame 625 (S)-fluoxetine as reference in purple	242

List of Tables

Table 1: Vaccine candidates in clinical trials phase II	44
Table 2: Macro domain mutations and their effects (adapted from Leung et al. 2018)	50
Table 3: Features of the pharmacophore model	58
Table 4: Final compound selection purchased from SPECS	61
Table 5: Modifications of ring A	68
Table 6: Synthesis of the intermediates with modified ring B	69
Table 7: Compounds with modified ring B	69
Table 8: Modifications of ring C	70
Table 9: Biological evaluation of synthesised compounds	71
Table 10: Series 1 of compound 10 analogues tested in shipment 2	73
Table 11: Nucleophilic substitution for intermediates 49-51 for compounds 52-56	73
Table 12 Reductive amination for compounds 52-56	74
Table 13: Biological activity of derivatives of series 2	74
Table 14: Synthesis of the intermediates 57-60 for compounds 61-64	75
Table 15: Conditions and yields to obtain compounds 61-64	76
Table 16: Compounds 65-67 with heterocycles in building block C	76
Table 17: Synthesis of 69 with heterocycle in building block A	77
Table 18: Biological evaluation of the compounds from the third round of modifications	77
Table 19: EC₅₀ summary of the assays performed for compound 10	80
Table 20: Compounds analysed in the thermal shift assay	82
Table 21: Chemicals for the preparation of 27	92
Table 22: Chemicals for the preparation of 10	93
Table 23: Second batch of compound 10	94
Table 24: Chemicals for the preparation of 28	95
Table 25: Chemicals for the preparation of compound 28	95
Table 26: Chemicals for the preparation of 29	96
Table 27: Chemicals for the preparation of 30	97
Table 28: Chemicals for the preparation of 31	98
Table 29: Chemicals for the preparation of 32	99
Table 30: Chemicals for the preparation of 33	100
Table 31: Chemicals for the preparation of 34	101
Table 32: Chemicals for the preparation of 35	102
Table 33: Chemicals for the preparation of 36	103
Table 34: Chemicals for the preparation of 37	104
Table 35: Chemicals for the preparation of 38	105

Table 36: Chemicals for the preparation of 39	105
Table 37: Chemicals for the preparation of 40	106
Table 38: Chemicals for the preparation of 41	108
Table 39: Chemicals for the preparation of 42	109
Table 40: Chemicals for the preparation of 43	110
Table 41: Chemicals for the preparation of 44	111
Table 42: Chemicals for the preparation of 45	112
Table 43: Chemicals for the preparation of 46	113
Table 44: Chemicals for the preparation of 47	114
Table 45: Chemicals for the preparation of 48	115
Table 46: Chemicals for the preparation of 49	116
Table 47: Chemicals for the preparation of 50	117
Table 48: Chemicals for the preparation of 51	118
Table 49: Chemicals for the preparation of 52	118
Table 50: Chemicals for the preparation of 53	119
Table 51: Chemicals for the preparation of 54	120
Table 52: Chemicals for the preparation of 55	122
Table 53: Chemicals for the preparation of 56	122
Table 54: Chemicals for the preparation of 57	123
Table 55: Chemicals for the reaction of 58	124
Table 56: Chemicals for the synthesis of 59	125
Table 57: Chemicals for the synthesis of 60	126
Table 58: Chemicals for the synthesis of 61	126
Table 59: Chemicals for the synthesis of 62	127
Table 60: Chemicals for the synthesis of 63	129
Table 61: Compounds for the synthesis of 64	130
Table 62: Chemicals for the synthesis of 65	131
Table 63: Chemicals for the synthesis of 66	132
Table 64: Chemicals for the synthesis of 67	133
Table 65: Chemicals for the synthesis of 68	134
Table 66: Chemicals for the synthesis of 69	135
Table 67: List of selected compounds for biological testing	150
Table 68: Identity and similarity of the human macrodomain containing proteins	155
Table 69 Compounds selected for the P23 binding site	171
Table 70: Activity and cytotoxicity of the tested compounds	174
Table 71: Enterovirus representatives	181
Table 72: Mutations of 2C inhibitors (Table adapted from supplementary material of Guan et al. 2018)	201
Table 73: Compound selection after HTVS	207
Table 74: Sequences of viruses retrieved for homology modelling	213

Table 75: Analysis of the monomer models on the SAVES server	214
Table 76: Binding energies of the protein ligand complexes during molecular dynamics simulations	227
Table 77: Evaluation of (R)-, (S)- and racemic fluoxetine against a panel of enteroviruses	233
Table 78: Fluoxetine analogues	239
Table 79: Guanidine analogues of fluoxetine	240
Table 80: First round of compound XX analogues synthesised and tested	243
Table 81: Evaluation of Zuo2 analogues on a panel of entero- and rhinoviruses	244

List of Equations

Equation 1: Summative energy calculation used in force fields	12
Equation 2: Schematic scoring function	17
Equation 3: Consensus scoring function sum of SIGN value for each scoring function (A, B or C) calculated as the quartile minus the score of the pose X of each scoring function.	60

1 Introduction

1.1 Antiviral research and antiviral drug discovery

1.1.1 History

In 1892, when bacteria were already discovered as microorganisms causing disease to plants and humans, and technology was evolved enough to remove all so far known bacteria by filtration, Dimitrii Ivanovsky made the observation that the pathogen causing tobacco mosaic disease could not be filtered out of the solution and the filtrate could infect healthy tobacco leaves dependent on the concentration of the filtrate solution (Lechevalier 1972). Similar results were independently obtained by Martinus Beijerinck, who reasoned that the cause must have been a distinct agent too small to be retained in the filters. In 1898 Friedrich Loeffler and Paul Frosch observed the same with the agent causing foot-and-mouth disease. Investigating the matter further they observed that the concentration of the infectious agent did not increase in a medium that supported the growth of bacteria or cells, therefore they hypothesized the parasitic nature of the pathogen and its host dependency. The discovery of viruses is often attributed to Ivanovsky, but only Beijerinck, Loeffler and Frosch suspected distinct agents of parasitic nature, smaller than bacteria, to be the pathogens that we now know as viruses (Bos 1999).

Historical evidence of several viruses that we know to date can already be found in ancient documents from Egyptian hieroglyphs and drawings to legal documents, reports and of course in legends and stories. It is notable that highly virulent viruses are only likely to have occurred after the settlement of humans in bigger agglomerations, where there is a sufficient number of naïve and susceptible population to sustain virus spread before leading to immunization of the subjects (Baron et al. 1996). Ancient reports of diseases together with genetic analysis of finds from excavations all around the world are parts of a big puzzle of information to decipher the history of human, animal and plant pathogenic viruses (Thèves et al. 2016).

Curiously, antiviral therapy and vaccination began before the discovery of viruses as the disease-causing agents. Treatment of various diseases with natural remedies existed ever since. The concept of immunity was observed for diseases that only affected people that were naïve to the disease, but once they survived they were protected for their lifetime. The best example for controlled countermeasures against viral diseases would be smallpox. In a procedure called variolation, material scratched out of pustules of infected patients was applied into a little scratch on the arm of healthy individuals. Many of them contracted a milder form of the disease and the risk of death was drastically reduced compared to naturally infected patients (NIH US National Library of Medicine & Public Health Service Historian 2002).

In the 1790s, in the UK Edward Jenner recognised that milkmaids that were exposed to cowpox were widely protected against smallpox. He conducted experiments of variolation with cowpox in humans to test the protective qualities of this much milder disease against the severe smallpox. The term vaccination comes from this practice as “vacca” in latin means cow (Smith 2011). Vaccinia virus that was widely used to immunize against smallpox in the past, is a derivative of the original cowpox vaccine that might have transformed within the human host (Riedel 2005). Louis Pasteur in 1885 discovered a way to attenuate the virulence of viruses by passaging them in animals. He infected rabbits with material of cows with rabies disease. Extracting the spinal cord of the infected rabbits and infecting others with the obtained extract finally reduced the symptoms of the disease to a much milder form but still produced immunity against the disease (Smith 2012). In the 1930s attenuation of viral vaccines was possible in a much easier way and on a larger scale, leading to the vaccines against yellow fever and influenza. These advances were due to the identification of the viruses as intracellular parasitic and pathogenic entities (Plotkin 2014).

With the advances of physics, especially microscopy, and with the invention of the electron microscope it was finally possible to visualise entities smaller than bacteria and cells. In this time many viruses were morphologically characterised and that opened the door to the first rational classification of viruses (Haguenau et al. 2003). The discovery of bacteriophages, viruses that infect bacteria, led to quick advances in molecular biology and in the research of viruses. This was crucial for the studies on the viral replication cycle (Keen 2015).

With a better understanding of the biology and replication of viruses the possibility to interfere with these processes and to find agents that will stop viruses from replicating and spreading became achievable. The field of virology was born (Flint et al. 2009).

1.1.2 Virus classification

Since viruses were discovered and especially since they could be characterised, there was a need for a classification system. Viruses are not living organisms but they do have a genome and other characteristics that allow them to be classified in a taxonomic system. The institution that is responsible for the classification of viruses is the International Committee on Taxonomy of Viruses (ICTV). Its function is to classify viruses into orders, groups, genera etc. and to provide a unified nomenclature for all known and new viruses. The ICTV takes into account newly generated data and new findings, thus the classification continues to vary. As research advances fast in the field and new techniques to analyse the relations between different viruses are being developed and used, the need for the committee is now higher than ever. This also explains the constant changes in the reports of the committee. Suggestions for

names of new viruses are reviewed as well as advances in genome analysis or bioinformatics that can help characterisation. All changes and efforts since the 1970s can be found on the ICTV website <https://talk.ictvonline.org> (Lefkowitz et al. 2018).

A simplified classification was used previously and can still be used today: the Baltimore classification. It divides the viruses into seven different groups according to the type of their genomes and their mode of replication.

Baltimore classification:

Group I: dsDNA viruses

Group II: ssDNA viruses (+ strand) DNA

Group III: dsRNA viruses

Group IV: (+)ssRNA viruses (+ strand) RNA (e.g. Picornaviruses, Togaviruses)

Group V: (-)ssRNA viruses (- strand) RNA

Group VI: ssRNA-RT viruses (+ strand) RNA with DNA intermediate in replication cycle

Group VII: dsDNA-RT viruses DNA with RNA intermediate in replication cycle

Originally the classification contained only six groups but was adapted when dsDNA-RT viruses were discovered. The classification still applies and its benefit is that knowing the type of nucleic acid helps the virologist to immediately connect the virus to a replicative cycle, because of the necessary steps during transcription (Baltimore 1971).

The viruses discussed within this work are part of group IV ((+)ssRNA viruses) and their replication cycles share at least some similarities. Several Picornaviruses and the Togavirus chikungunya virus will be the main focus of this work.

1.1.3 The viral replication cycle

In general, the viral replication cycle is characterized by several crucial steps that are common for all viruses (Goulding 2016). It starts with the virus entry. The interaction between the host and the virus can be mediated by specific receptors and co-receptors that serve as gateways for the virus to break into the host cell. Another way for a virus to enter its host cell is for example endocytosis, where the viral particle is ingested by the cell. Thereafter, the outer shell of the virus must break down to release the genetic information into the cytoplasm. This can happen via membrane fusion processes and virus capsid disassembly. When the virus managed to successfully enter the cytoplasm, it usually hijacks ribosomes to translate the protein-coding parts of the genome into viral proteins. Once produced, the viral non-structural proteins deal with most of the necessary processes to make new viruses. Non-structural proteins function as replication machinery and have auxiliary functions, whereas structural

proteins will assemble the new virions. But not only proteins need to be produced, also the genome needs to be copied. This is a task usually carried out by non-structural proteins. Then the genome is incorporated into the nascent virions. Some viruses require the maturation of certain proteins, either still within the host cell or after the budding off the host cell. For some viruses cell-lysis is observed instead of budding.

Additional steps in the viral replication cycle depend largely on the viruses and their genomes and if they replicate their genomes in the cytoplasm or if import and export to and from the nucleus is required. The viruses discussed in this work are (+)ssRNA viruses and their replication cycle will be discussed separately in the corresponding sections. For the general discovery of antiviral agents and the history of antiviral drug design many specific viral proteins and processes are important. Some of them are described in the following section (Flint et al. 2009).

1.1.4 Antiviral therapy

The previous section exemplified the complexity of viral replication. With this understanding and better insights into virus biology, many crucial steps in the replication cycle of viruses are suitable targets for antiviral drug design. But the task remains challenging. Viruses are intracellular parasites and therefore rely on the host cells. Many processes that could stop viral replication are also harmful to the host cell. Given that viral infections very often are self-limiting and non-chronic, it is of utmost importance that antiviral drugs are safe and non-toxic. The development of new antivirals and especially the phases of clinical testing are further hampered by the fact that newly discovered viruses must quickly be characterised. If there was a drug ready to be tested, clinical trials needed to be granted permission quickly. In the cases of Ebola or Zika for example, there were urgent needs for medication when the outbreaks occurred, but logistics and bureaucracy proved difficult. Only after the peak of the epidemic trials could be initiated. Some promising compounds and a vaccine against Zika are now awaiting a future outbreak in order to complete their test phase and can eventually be approved. Both WHO and FDA or other medical agencies have to address the ethical concerns behind “emergency clinical trials”, as well, and data from experimental treatment has to be evaluated with caution (Whitehead et al. 2016).

In 1963 the first antiviral drug Idoxuridine was approved against herpes simplex virus type 1 (Prusoff 1959)(Kaufman 1962). It marked the start of a list of over 90 compounds of marketed antiviral drugs we have to our disposition to treat viral infections. But this list is not equally distributed over the classes of known viruses. Only 9 viral diseases can be treated with them. Due to similarities in replication cycle, genome or some target proteins, off-label use and trials

to expand the spectrum of the known and approved compounds are one big hope for patients that are infected with new viruses, for which a specific antiviral has not yet been found. Nevertheless, the applicability and the success are limited.

In their review, De Clercq and Li give an extensive overview over the developments in the antiviral drug discovery field of the past 50 years. The achievements are great but challenges remain. The nine treatable viral diseases are human immunodeficiency virus (HIV), hepatitis B virus (HBV), hepatitis C virus (HCV), herpesvirus infections, influenza virus infections, human cytomegalovirus infections (HCMV), varicella-zoster virus infections, respiratory syncytial virus infections and external anogenital warts caused by human papillomavirus infections (De Clercq & Li 2016). The classes of virus therapeutics are referred to the viral life cycle of the corresponding viruses. Thirteen functional mechanisms are exploited so far to target the different viruses. (I) 5-substituted 2'-desoxyuridines target viral DNA synthesis in the nucleus, (II) nucleoside analogues target viral DNA synthesis in the nucleus and reverse transcription, (III) pyrophosphate analogues are interfering with the viral DNA synthesis as well. (IV) Nucleoside reverse transcriptase inhibitors (NRTIs) and (V) non-nucleoside reverse transcriptase inhibitors (NNRTIs) target the reverse transcriptase by two different mechanisms, (VI) protease inhibitors target viral proteases, (VII) integrase inhibitors inhibit the integration of viral DNA into the host genome, (VIII) entry inhibitors block the virus entry into the host cell, (IX) acyclic guanosine analogues target the viral DNA polymerase, as do (X) acyclic nucleoside phosphonate analogues. (XI) Last, HCV NS5A/NS5B polymerase inhibitors and (XII) influenza virus inhibitors build the two distinct classes of approved drugs (De Clercq & Li 2016).

The most recent FDA approvals between 2016 and November 2018 are the following four different combinations for HCV infections: Zepatier (elbasvir + grazoprevir), Epclusa (sofosbuvir + velpatasvir), Vosevi (sofosbuvir + velpatasvir + voxilaprevir), Mavyret (glecaprevir + pibrentasvir). In 2017 Prevymin was approved containing Letemovir against HCMV infections in transplant patients. Early 2018 Biktarvy, a new triple therapy for HIV and Trogarzo, a non-immunosuppressive monoclonal antibody binding CD4, as entry inhibitor against HIV, were approved. TPOXX (tecovirimat), an envelope inhibitor against smallpox, was approved in July 2018 (a measure to protect against potential bioweapon threat). Pifeltro with its compound doravirine presents a new NNRTI against HIV-1, and finally Xofluza (baloxavir morboxil), the first cap-dependent endonuclease inhibitor for influenza virus, was approved end October 2018 (FDA U.S. Food & Drug Administration 2016; FDA U.S. Food & Drug Administration 2017; FDA U.S. Food & Drug Administration 2018).

There is a growing number of new compounds in various different stages of clinical or pre-clinical evaluation. One of the challenges of developing an antiviral drug is the size of the

patient population and the duration of treatment. For the patient's benefit, an antiviral cure should be effective in a couple of days or weeks and then eliminate the virus from the system and restore the patient's health. The treatment should be affordable and the drug production cheap and easy. For the treatment of tropical diseases in remote areas it would be highly beneficiary to assure the stability of the compound at elevated temperature or humidity, especially without the necessity of a cold chain.

Unfortunately the revenue is therefore clearly not as long lasting, and as constant, as for a drug against diabetes or hypertension, for example. Therefore, from a commercial point of view, bringing an antiviral drug to the market involves high development costs but also a high risk in terms of predictable revenue.

1.1.5 Vaccination

For several viruses vaccination proved a suitable strategy to protect the population from devastating diseases. Several seriously life-threatening viruses will hopefully be eradicated by vaccination in the future. The only worldwide-eradicated virus due to extensive vaccination efforts and the nature of the viral disease is smallpox. In 1980 the WHO declared smallpox eradicated (WHO 2018b). But even with safe and efficient vaccines antiviral drugs are needed to control the virus within the people that are already infected and carry the virus. Vaccines can prove disadvantageous or problematical for viruses with many different closely related serotypes as in the case of dengue virus, where vaccination against one serotype could aggravate the outcome when infected with a different serotype (WHO 2018a). Efficacy of vaccines against quickly evolving viruses might be compromised. In addition, geographical and distribution challenges, and the need for a cold chain for many vaccine preparations, pose another problem when considering outbreaks in remote areas, which is often the case for tropical diseases. Predicting, producing and stockpiling an adequate amount of doses has proven difficult in the past as illustrated by the Yellow Fever outbreak in 2016. The WHO temporarily issued a statement that doses could be fractioned to a fifth and still prove effective. This was an emergency measure that should not become regularity but gives hope that global efforts in distribution can help in cases of severe outbreaks and at least prevent further spread (WHO 2016).

1.2 Drug discovery and design - a complex process

The workflow in Figure 1 shows the various factors contributing to the discovery and design of new drugs nowadays. It does not include the details of the phases of clinical evaluation, which are still crucial steps for a new drug to succeed or fail and take the biggest chunk of time and money within the whole process. The figure highlights complexity and the multidisciplinary character of preclinical development. In most of the steps, computers are used to support the researcher in dealing with large amounts of experimental data, but also by the availability of specific software applications to replace some of the chemical or biological experiments, which are usually cost and time consuming and require a lot of resources. Especially in medicinal chemistry and drug design, computers have become an essential tool to support the chemist and to enable a range of new rational drug design techniques that fall under the terms molecular modelling or computer-aided drug design (Klebe 2009).

1.3 Molecular modelling

1.3.1 Fundamental concepts in molecular modelling

Molecular modelling roots in the description and visualisation of chemical and biological molecules. Two-dimensional drawing is most commonly used and a very practical tool for the representation of small molecules in chemistry. It is especially useful when handled by a chemist, because humans can intellectually interpret its meaning and degree of abstraction. The first ones to draw molecular representations in 2D were Loschmidt and Kekulé who can be seen as fathers of chemical representations as we know it now (Loschmidt 1861). But molecules are not two-dimensional and, especially for the interactions with a biological target, the occupied three-dimensional space and the arrangement of the chemical features within this space are of great importance. The first three-dimensional models were made of balls with holes and sticks that represent the bonds between the atoms or of spheres that were overlapping to build the model of a linear molecule that is space-filling. Such models were crucial for the discovery of the alpha-helix as protein fold by Pauling and colleagues (Pauling et al. 1951) and the structural elucidation of the DNA double helix by Watson and Crick that yielded them the Nobel price in 1962 (Meinel 1992).

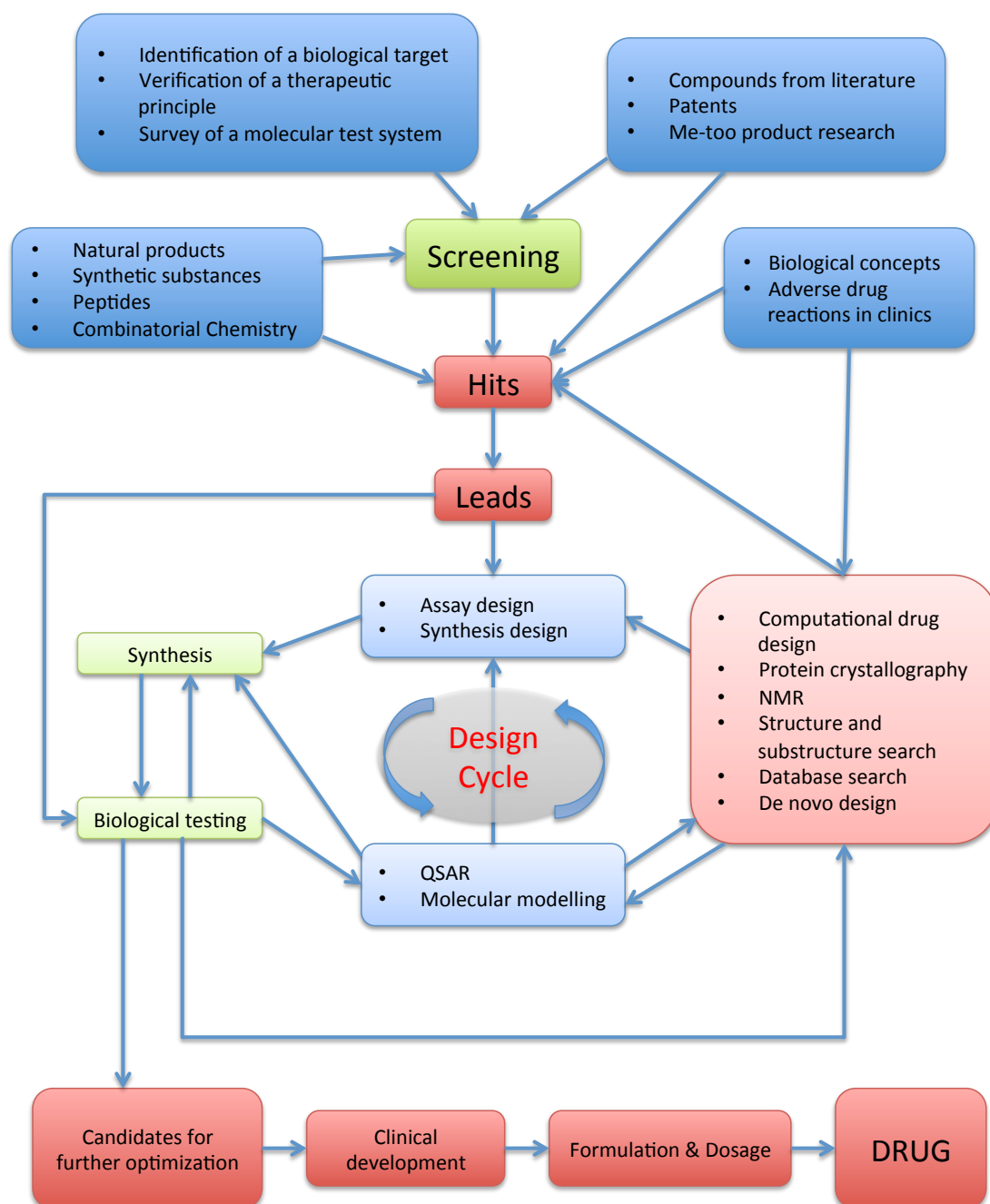


Figure 1: Drug design workflow adapted from Klebe 2009 reprinted from Zonsics 2013

1.3.1.1 Molecular modelling beyond molecular representation

The terms molecular modelling and computational chemistry are somehow used synonymously, nowadays. The computational calculating capacity and advances in graphical representation have clearly facilitated the visualisation of molecular structures and enhanced the understanding of molecular interactions in the three-dimensional space. With greater computational power it became possible to carry out complex physicochemical calculations on bigger molecular systems that can be applied to problems in drug design. This saves time and

money of wet lab experiments and requires only access to a powerful computer and the adequate software.

The most common tasks in molecular modelling comprise the representation of 3D molecules, the modelling of small molecules with respect to their 3D structure generation, conformation analysis, calculation of physicochemical parameters, the comparison of small molecules and their alignment. For the modelling on proteins, sequence comparison, homology modelling and protein-folding simulations are among the most frequently used queries. With the advance of graphic cards and parallel computing even molecular dynamics simulations of protein complexes became feasible in a reasonable time frame. The calculation or prediction of binding constants and energies between proteins and ligands and the docking and scoring of small molecules are among the most important tasks for drug discovery. With the handling of big compound databases and experimentally determined data from biological assays, high-throughput virtual screening has now a fixed place in the discovery of new potential hits in the drug discovery process (Klebe 2009).

The techniques used in computational chemistry are fundamentally based on the following two principles: 1) Quantum mechanics calculations that use the basic laws of quantum physics to accurately calculate molecular properties from scratch (*ab initio*) 2) empirical methods, which rely on the vast amount of experimental data collected in specific repositories and create models to extrapolate therefrom. A combination of both are semi-empirical methods, which are simplifying the quantum chemical calculations by using empirical data in the approximations of otherwise computationally expensive terms in the equation (Leach 2001).

1.3.1.2 Quantum mechanics

Quantum mechanics describes the physics of energy and matter of particles on a very small scale. This is exactly the case when looking at atoms and bonds in chemical molecules. The physicochemical properties of a molecule are dependent on the states of protons and electrons. Thus, quantum mechanics is the most precise description of the molecular energy and can be used in systems where electronic contributions are relevant. There are several theories about how to treat a molecular system and the most widely used is the molecular orbit theory (Coulson 1956). Other theories like the extended Hückel theory (EHT) (Hoffmann 1963) are implemented within software packages used in the present work, but in depth discussion of the different theories is beyond the scope of this thesis. The Schrödinger equation is the fundamental equation in which all discussions about quantum mechanics are rooted. It is a partial differential equation that can be analytically solved only for very simple

cases like the hydrogen atom. The key lies in the mathematical determination of the wave function for a given molecule. Once determined, the energy of a molecular system can be calculated based on this wave function. For systems with more atoms and electrons several approximations have to be made, some of which based on empirical observations, resulting in semi-empirical approaches (Leach 2001).

1.3.1.3 Molecular mechanics and force field methods

Molecular mechanics falling under the empirical methods and using equations derived from classical physics and mechanics, usually describing phenomena on the macro scale, and applying them to molecular systems. Also referred to as force field methods, these techniques do not take into account the contributions of the electrons as specific entities within the molecular system. The concept simplifies the molecular system to atom nuclei (often represented as balls) and bonds (represented as sticks that connect the atoms to each other). Thus, equations can be solved in a fraction of the time compared to those of quantum mechanics, and they are often still sufficiently accurate. To calculate the energy of a molecular system with molecular mechanics, a so-called force field has to be parameterised. It stores the atom and bond types in a table of standard parameters, which were observed in experiments. All atoms in the given system need to be assigned to one of the predefined categories for the force field, to be applicable on the given problem (Klebe 2006).

In order to calculate the energy of the system any difference between the standard parameter and the actually found solution results in a penalty for the energy value. A classical force field calculates the energy in a summative equation comparable to the following:

$$E = E^{\text{bondlength}} + E^{\text{angle}} + E^{\text{torsion}} + E^{\text{improper}} + E^{\text{ele}} + E^{\text{vdW}}$$

Equation 1: Summative energy calculation used in force fields

Each of these contributions to the total energy has its own sub-equation and, depending on the parameters, additional considerations might be accounted for in the equation. Bondlength, angle, torsion and improper (for improper dihedral angle bending) are intramolecular properties that apply to atoms that are covalently bound to each other. Electrostatic and van der Waals terms are taking into account interactions between different molecules (Riniker 2018).

Several different force fields exist and are applicable to different molecular systems. Force fields like the MMFF are parameterised for small molecules. AMBER force fields are used for proteins and nucleic acids. And more modern force fields take into account both components

at the same time. The force fields mainly used within this work are AMBER12:EHT (EHT standing for extended Hückel Theory) within the software package MOE (Chemical Computing Group Inc. 2016) and OPLS-2005 or OPLS3 within the Schrödinger suite (Harder et al. 2016; Jorgensen et al. 1996; Schrödinger LLC 2018).

1.3.1.4 Energy minimization

Energy minimization is crucial when drawing or generating new small molecules, but also in protein preparation energy minimisation is used to find the best conformation for example of protein sidechains. Energy minimisation is a classical application of molecular mechanics. In the process, slight alterations of bondlength or angle or the rotation around single bonds are performed and the energy of the resulting molecule is assessed and compared to the original molecule. If the energy is more favourable (negative), the conformation is kept for the next round of alterations. This process is repeated until changes do not result in smaller energy values anymore and this point is considered as an energy minimum. There are two types of minima: local minima and the global minimum (Figure 2). Energy minimisation with classical methods finds the closest minimum to a given input conformation. To reach the overall global minimum on the energy landscape, “hills” have to be overcome. In a classical minimisation algorithm, only the closest minimum is reached because the program never goes “up” the energy hills but only down (Leach 2001; Schneider & Baringhaus 2008).

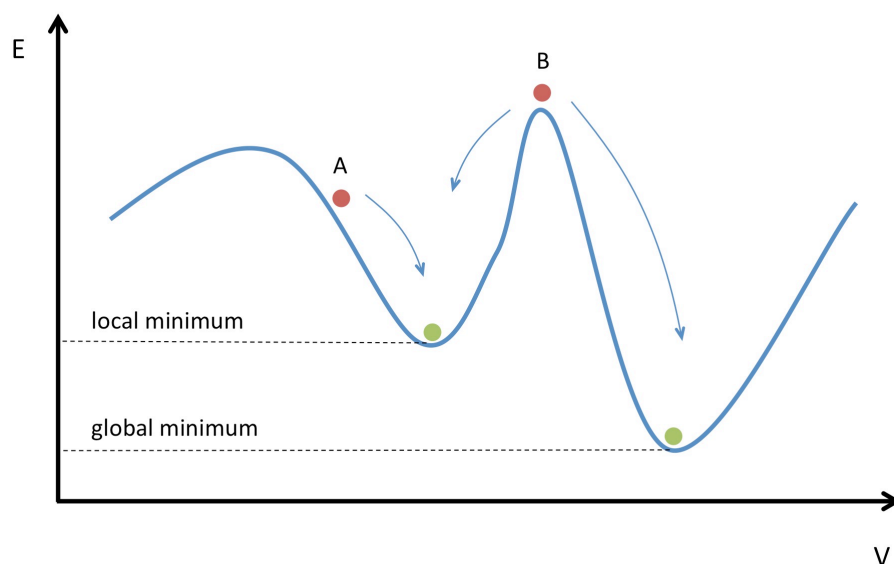


Figure 2: Energy landscape

Starting conformations in red, minimum conformations in green; E = energy, V = variation

1.3.1.5 Conformation generation

An ancient and somehow simplified concept of small molecules interacting with their target is the principle of lock and key postulated by Emil Fischer in 1894 (Klebe 2009). The molecule is the key and it needs to fit into the binding pocket in order to exert its function. Like a key a molecule has a three-dimensional structure and the chemical groups or features need to be in a certain arrangement in order to be complementary to the receptor. Molecules are not rigid and can therefore adopt different so-called conformations, one of which might be the active one. In order to find the active conformation of a molecule on the computer, many different conformations have to be generated. Molecules usually favour a conformation that resides within an energy minimum. To reach each possible minimum and therefore all stable conformations of a small molecule, it is necessary to start from a diverse set of initial conformations and then minimize each of them. These conformations can then be used in a docking study or a pharmacophore search, depending on the task. In Figure 2 the red dots can be seen as two different conformations of one molecule generated by the program. Conformation B could reach the global minimum, whereas conformation A will not overcome the steep energy barrier and only reach the local minimum. The active conformation does not necessarily have to be the global minimum, but at least the conformation must be stable (Leach 2001).

1.3.2 Applications of molecular modelling

Generally speaking, applications in molecular modelling distinguish between structure- and ligand-based approaches. For the former, the protein of interest is known and its structure is resolved either by X-ray crystallography, NMR or cryo-electron microscopy or other techniques that allow visualisation of the protein structure sometimes up to the atomic level. If the structure is not resolved but a structure of a homologous protein is, a homology model can be created and treated similarly to a crystal structure in the drug discovery or design process (Schmidt et al. 2014).

Ligand-based design, on the contrary, lacks the information of the specific receptor and only molecules exerting a certain function are known and form the basis of the drug design process. In the ligand-based design process, compound and conformation libraries are of outstanding importance and the calculation or prediction of physicochemical parameters of new molecules is a crucial step in generating new active compounds.

All molecular modelling techniques are predictions to a certain degree and can only be an aid for the selection of molecules to be tested in *in vitro* assays. In ligand-based design the chemical synthesizability might be predicted and this needs to be tested by a chemist that actually makes the compound.

1.3.2.1 Homology modelling

Homology modelling is a method to build the 3D structure of a protein with a known amino acid sequence using one or more similar (homologous) proteins with existing crystal structures as template(s). The resulting 3D structure of the query protein is called homology model and can be used in the same manner as a crystal structure, i.e. to pursue a structure-based drug design approach (Cavasotto & Phatak 2009).

The generation of a new homology model requires four distinct steps. First the crystal parameters of conserved regions between the template and the query are copied as they are, while only the structure of the backbone is kept for all other regions. Then the program takes into account all missing or additional regions for which the template and the query sequence cannot be matched. For additions that have to be made – usually loops – the homology model tool has a specific feature that can also be used on its own: the loop modeller. By comparing the connective residues or the short sequence that has to be bridged, for which no structural information is provided in the template, the software package Molecular Operating Environment (MOE) (Chemical Computing Group Inc. 2016), which was predominantly used in this project, has implemented a database search referring to the protein data bank (PDB) (Bernstein et al. 1977), in order to identify the most similar and most plausible connective loop out of high resolution structures, taking into account the root mean square deviation (RMSD) distance between the model and the template loop. In the third step the loops are evaluated and then the geometry of the sidechains is optimised. Before the final refinement of the model, hydrogens are added to the systems, valences are calculated and serious steric problems are highlighted for further evaluation. The last step is the refinement of the best intermediate model. The user can choose between several refinement techniques and the selection usually depends on the input parameters and the available information about the structure.

For the evaluation of the homology models, a tool within the software package MOE can be used as well as an online platform for structural evaluation of proteins called SAVES (Molecular Biology Institute at the University of California 2016), compiling different tools to evaluate crystal structures and homology models. After the generation of a homology model, it should always be evaluated to guarantee the plausibility of the new structural information. The quality of the input crystal structure and the structure alignment are of utmost importance when generating a homology model. Bad input results in a bad model (Venclovas & Margelevičius 2005).

1.3.2.2 Pharmacophore modelling

Our actual understanding of a pharmacophore was first created by Paul Ehrlich in 1909 describing “a molecular framework that carries (phoros) the essential features responsible for a drug’s (pharmacon) biological activity (Ehrlich 1909). A pharmacophore model is therefore a 3D arrangement of electronic features describing the crucial interaction points of a bioactive molecule with its target (Wermuth et al. 1998). The classical pharmacophore features are hydrogen bond donors, hydrogen bond acceptors, hydrophobic, and positively and negatively ionized areas.

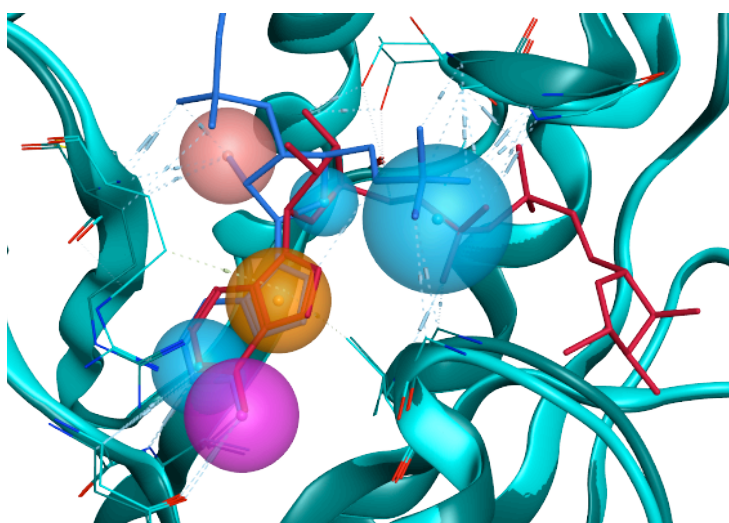


Figure 3: Pharmacophore model (PDB: 3GPO, 3GPQ)

purple: donor, blue: acceptor, orange: aromatic, pink: donor or acceptor, dotted lines: hydrogen bonds

A pharmacophore model containing a certain number of pharmacophore features to represent potential new molecules can be used to screen a database of compounds or conformations in order to find new molecules matching the pharmacophore features of the query whilst exhibiting a new scaffold or different functional groups.

The classical pharmacophore approach is used to describe a set of compounds, which are known to be active (and/or inactive) against a specific target, in order to find common and essential features to address the structurally unknown target. But as many computational techniques are applicable in a wider context than their original purpose, pharmacophore modelling and pharmacophore search can also be used to describe the features of co-crystallised ligands within a protein-ligand complex, in order to search for new molecules that might fit the template query. This approach was used in the current report using the pharmacophore query within the software package MOE.

1.3.2.3 Docking and Scoring

In order to see how a set of conformations fit into a designated binding pocket it is necessary to place them into the receptor pocket in a meaningful way and subsequently assess the position. These two problems are referred to as a) the docking problem and b) the scoring problem (Schneider & Baringhaus 2008).

Many automated docking programs have been developed to address these problems. In the current project Glide and PLANTS were used as docking programs (Friesner et al. 2004; Korb et al. 2009). The previously generated conformations are placed within the receptor in different docking poses and each pose needs to be evaluated. The evaluation is carried out by a scoring function. Both the ligand and the receptor each possess six degrees of translational and rotational freedom (Cui et al. 2011). Typically, and especially when docking a big library of compounds, only the conformational space of the ligand is explored, and therefore it is considered flexible in the docking, while the receptor is kept rigid.

Usually the docking program generates a vast number of poses for each ligand. Most of them can be sorted out immediately because of energetic clashes with the receptor, but the remaining poses have to be assessed and ranked for their quality compared to all other poses of all other possible ligand molecules within the query database.

The most common approach to rank ligands is to take into account the calculated free binding energy considering the various contributions to the energy function of a receptor-ligand complex: solvent effects, conformational changes in protein and ligand, specific protein-ligand interactions, rotational freedom, translational and rotational movements of the ligand, vibration energy. Such an equation is symbolically depicted below:

$$\Delta G_{\text{bind}} = \Delta G_{\text{solvent}} + \Delta G_{\text{conf}} + \Delta G_{\text{int}} + \Delta G_{\text{rot}} + \Delta G_{\text{t/r}} + \Delta G_{\text{vib}}$$

Equation 2: Schematic scoring function

Possible contributions to the binding free energy (ΔG_{bind}): $\Delta G_{\text{solvent}}$ contribution of solvation energy, ΔG_{conf} contribution of conformational changes, ΔG_{int} contribution of protein-protein interactions, ΔG_{rot} contribution due to changes in rotational freedom, $\Delta G_{\text{t/r}}$ contribution due to changes in translational and rotational movements, ΔG_{vib} contributions of vibration energy

The various scoring functions take into account different contributions to the binding free energy in a different way. Therefore, values for each pose scored with a different scoring function vary. In order to get an unbiased result, the docking poses generated with Glide were first evaluated by the standard precision scoring function GlideSP, and then further cross-evaluated and rescored with three other scoring functions: GlideXP, FlexX and PLANTS (Friesner et al. 2004; Rarey et al. 1996; Korb et al. 2009).

To facilitate the decision as to which compounds to take forward in the selection process, a consensus score including all three scoring results of the aforementioned scoring functions was calculated.

1.3.2.4 Molecular dynamics

Molecular dynamics (MD) is a simulation that investigates the movement of atoms in a molecular system over time. Examples of processes of interest in molecular modelling which can be investigated by molecular dynamics simulations, are protein folding, conformational changes, molecular interactions between proteins and their ligands, interactions of proteins with membranes, ion transport by ion channels, and others (Klebe 2009). An MD simulation will result in a trajectory, similar to a movie, of the modelled system over the given simulation time. Processes that occur in a fraction of the simulation time can be followed and analysed.

In order to run an MD simulation a start geometry needs to be obtained. This can be a PDB structure, a homology model or a topology obtained from *ab initio* calculations. The structure must be prepared because errors in the input can lead to artefacts and drastic mistakes in the output. Most physiological processes occur in an aqueous environment and at room temperature or body temperature, both of which have to be modelled appropriately. Therefore, in the setup of the system the molecule is usually submerged in a box of water. Using explicit waters requires more computational power and time, but it is more accurate and yields better results. In the second step the forces acting on each atom in the simulation box have to be determined. The potential energy can be deduced from the structure, usually employing a force field. The force field needs to be accurately parameterised but then allows the calculation to be quite quick. The third step, then, numerically solves Newton's law of motion by first randomly assigning an initial velocity to each atom. This will set the system in motion and determine the movement of the particles. In very small time steps (for the relevant cases in drug discovery usually femtoseconds) the system is analysed. The new positions of the atoms are calculated as well as the forces and the acceleration. Each time step can be called a snapshot or frame. The time step has to be smaller than the time employed by the movements sought to be investigated. Furthermore, if the time steps chosen are too far apart from each other, particles in the simulation might occupy the same space and the simulation "explodes". The time step is also the speed-limiting factor, because the numerical integration of the equations has to be carried out for each time step separately, for the entire duration of the simulation (Leach 2001; Patrick 2009; Hospital et al. 2015).

Workflow for a MD simulation:

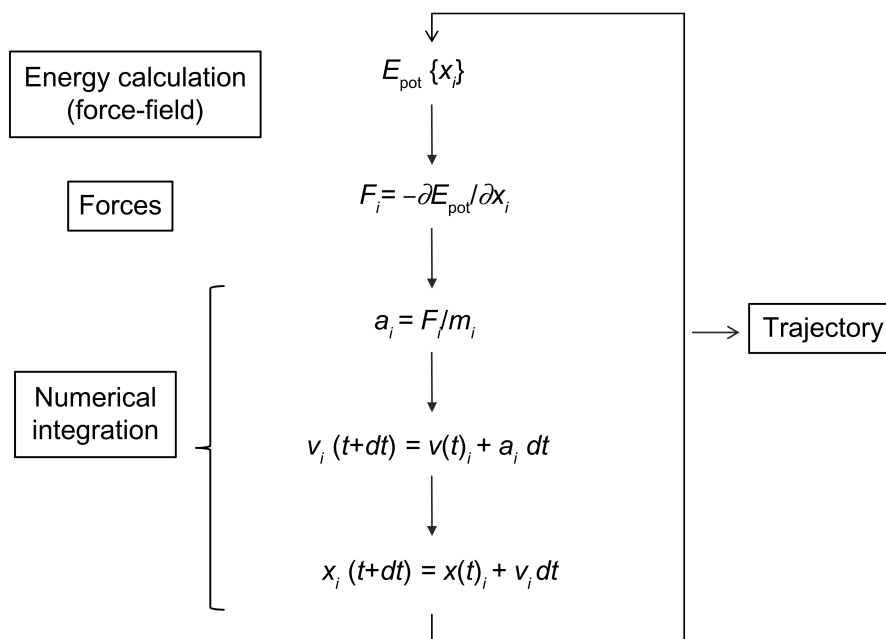


Figure 4: Molecular dynamics basic algorithm

Notes: The simulation output, the trajectory, is an ordered list of $3N$ atom coordinates for each simulation time (or snapshot). **Abbreviations:** E_{pot} , potential energy; t , simulation time; dt , iteration time; For each spatial coordinate of the N simulated atoms (i): x , atom coordinate; F , forces component; a , acceleration; m , atom mass; v , velocity. (Diagram and Legend reprinted according to the CC-BY license from (Hospital et al. 2015))

MD simulations can be performed under different physical conditions called ensembles, a concept from thermodynamics. Five ensembles are commonly used to ensure constant conditions throughout the simulation. Different values can be calculated from the different conditions. Gibbs defined three different ensembles: the microcanonical (NVE), the canonical (NVT) and the grand canonical ensemble (μ VT). Other two ensembles are of use in MD simulations: the isothermal-isobaric ensemble (NPT) and the isoenthalpic-isobaric (NPH; $H=E+PV$). N stands for particle number, V for volume, E for energy, T for temperature, μ for the chemical potential, P for pressure and H for enthalpy. With the different ensembles different structural, energetic and dynamic properties can be calculated from the fluctuations in the system (Hinchliffe 2003).

1.4 General aims

Many viral diseases were discovered in the past and for some of them countermeasures were found and applied successfully as described before. But there are still many viruses that pose a threat to human health that are not treatable or preventable at present. New viruses are being discovered and their level of danger to the human population in the future is predictable yet. Therefore, antiviral research and the discovery of new chemical entities against viral targets are important areas of research.

This work aims at targeting non-structural proteins of chikungunya virus, an Alphavirus that is transmitted by tiger mosquitoes, to design and synthesise new antiviral molecules that would be tested in cell-based antiviral assays and further characterised for their mode-of-action. The two non-structural proteins targeted in this thesis are nsP2 and nsP3. Several pharmacophore screenings and docking studies were used to select computational hit compounds that would then be purchased and tested for their antiviral activity. Then the most promising hit compounds and analogues thereof would be synthesised and tested again. To evaluate the activity against the target, assays on the purified protein would be performed.

The second target of this work is the 2C protein of selected picornaviruses. Human enteroviruses, coxsackieviruses, poliovirus and rhinoviruses are all part of the *Picornaviridae* family, and many members are human pathogens with a high socioeconomic impact on societies worldwide. A crystal structure of the 2C protein was resolved in 2017 for the first time, making it an interesting new target for structure-based drug discovery. Furthermore, 2C protein is one of the most conserved proteins within the *Picornaviridae* family. Therefore, it might be a suitable target to develop inhibitors against more than one virus simultaneously. Homology models would be created based on the enterovirus A71 crystal structure of 2C and known inhibitors targeting this protein would be investigated for potential binding modes. New 2C inhibitors would be screened with computational methods and evaluated in collaboration with virologists and structural biologists in cell-based antiviral assays and binding assays on the purified protein.

2 Chikungunya Virus

2.1 Introduction

2.1.1 Classification and Taxonomy

Chikungunya virus (CHIKV) is an arthropod-borne (Arbovirus) enveloped (+)ssRNA virus that belongs to the genus *Alphavirus* within the family of *Togaviridae*. The genus *Alphavirus* can be subdivided into three groups: the Sindbis-, the Ross River- and the Semliki Forest group with CHIKV falling into the latter one (Schlesinger et al. 2011).

Chikungunya virus has evolved several different lineages initially corresponding to the region of isolation, but in the newer outbreaks these lineages have separately spread out to new territories via international travellers. The four main lineages are the East/Central/South Africa (ECSA) lineage, the Indian Ocean lineage, the Asian Lineage, and the West African lineage (Weaver & Forrester 2015).

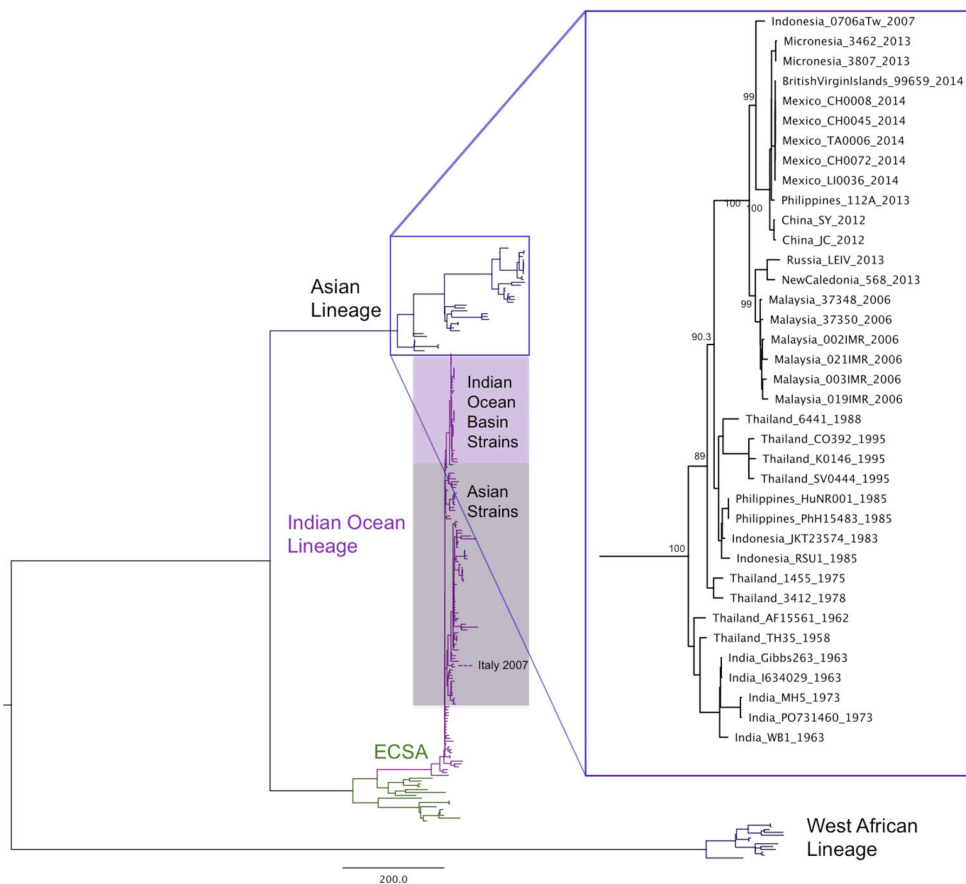


Figure 5: Phylogenetic tree of the different CHIKV lineages described in Weaver and Forrester, 2015.

2.1.2 Epidemiology

Chikungunya virus is the causing agent of Chikungunya fever, which was first described in 1952 in Tanzania during the course of an outbreak (Lumsden 1955). The first CHIKV strand was isolated in 1953 from a febrile patient in that region (Ross 1956). The name “Chikungunya” refers to a term in Makonde language, meaning “the one which bends up” describing the posture which a person adopts due to severe joint pain associated with the acute and chronic CHIKV infection (Kondekar & Gogtay 2006).

Since the first characterisation in 1953, CHIKV has repeatedly caused minor outbreaks in Africa whilst major epidemics occurred in India and Southeast Asia in the 1960s and 1970s. For the following 30 years only a few cases of CHIKV were reported until a large outbreak in 2004 in Kenya (Pulmanusahakul et al. 2011). From there the virus then spread towards wide regions in the Indian Ocean area, India and Southeast Asia (Staples et al. 2009).

In 2005/6 a severe outbreak hit the island La Réunion in the Indian Ocean and spread out towards Madagascar, India, and other islands in the area (Schuffenecker et al. 2006). In 2007 an outbreak in the south of Italy marked the first outbreak in Europe (Rezza et al. 2007) and from 2013 on the virus spread towards the Americas, finding perfect replication conditions, climate and vectors, also in the New World (Weaver 2014). In roughly one year from 2013-2014 CHIKV spread over 43 countries and infected around 1.1 million people. This brings the count to 3.4 million cases worldwide with many more people at risk. Areas inhabited by the mosquito vector *Aedes albopictus* are growing, expanding the area where the disease could be introduced and spread (Powers 2018).

2.1.3 Host

Chikungunya is an arthropod-borne virus with different transmitting mosquitoes. It is transmitted through a sylvatic cycle where monkeys build the reservoir for the virus, which is taken up by mosquitoes during their blood meal and subsequently transmitted to monkeys or humans. The other transmission route is predominantly found in urban and population dense areas where humans build the reservoir and mosquitoes pass on the virus within humans (Her et al. 2009).

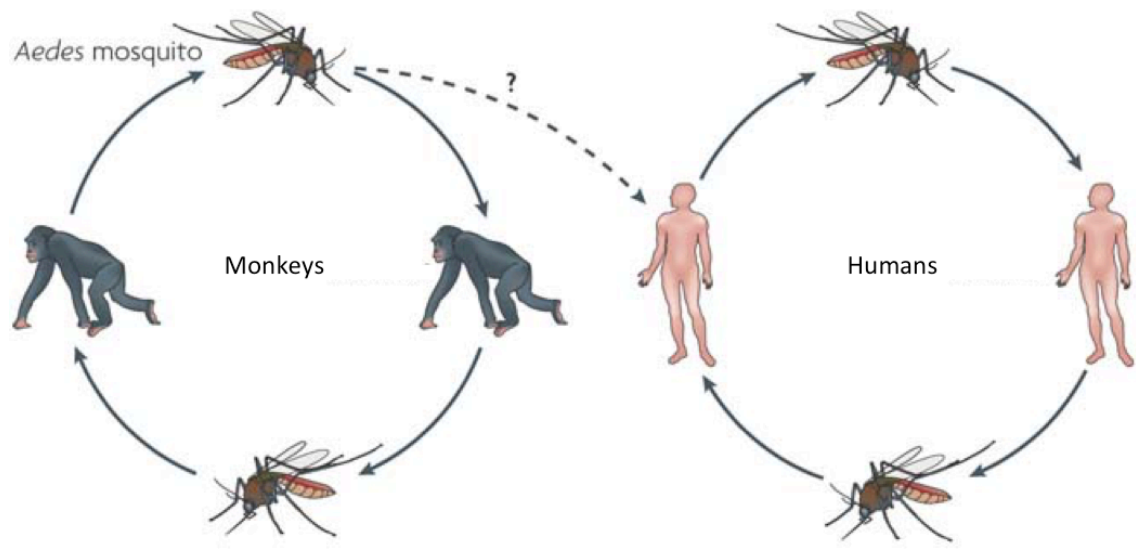


Figure 6: Transmission of CHIKV zoonotic and human cycle (Weaver & Barrett 2004).

The original vector was the *Aedes fuscifer* mosquito predominant in the African forests, which transferred the virus to primates and also caused small outbreaks in the local human population (Diallo et al. 1999). In Asia the main vector is *Aedes aegypti* (Weaver 2014). Mutations in the envelope glycoproteins E1 and E2 during the outbreak in 2005/6 in La Réunion lead to an adaptation to *Aedes albopictus* rendering the virus more efficient for a wider spread in the areas inhabited by that mosquito: Madagascar, Indian Ocean, Africa, Southern Europe and the Americas (Tsetsarkin et al. 2009) (Singh & Unni 2011). Considering the climate change the areas where *Aedes albopictus* is endemic might grow and, as such, also the potential areas for CHIKV and similar arthropod-borne infections (Gould & Higgs 2009).

2.1.4 Chikungunya fever

Typically 3-7 days after the bite by an infected mosquito a quick onset of the characteristic symptoms of chikungunya fever with high fever, painful polyarthralgia, asthenia, headache, vomiting, skin rash and myalgia occurs (CDC 2016a; CDC 2016b). These symptoms are usually persisting during the acute phase of 1-10 days. The chronic phase chikungunya fever can last up to months presenting with predominantly joint pain and myalgia (Schilte et al. 2013).

2.1.5 Treatment options

A lot of research was conducted especially since the outbreak in the Indian Ocean region in 2005/6. At the moment the only possibility is to treat CHIKV is symptomatic with anti-inflammatory and antipyretic drugs especially paracetamol. It is noteworthy that CHIKV co-locates often with dengue virus especially in the Americas and distinguishing between the two

viruses is still a challenge for local healthcare practitioners. Therefore classical NSAIDs are avoided in order not to enhance haemorrhagic complications especially with dengue virus infected patients (Bettadapura et al. 2013).

2.1.6 Description of the virion

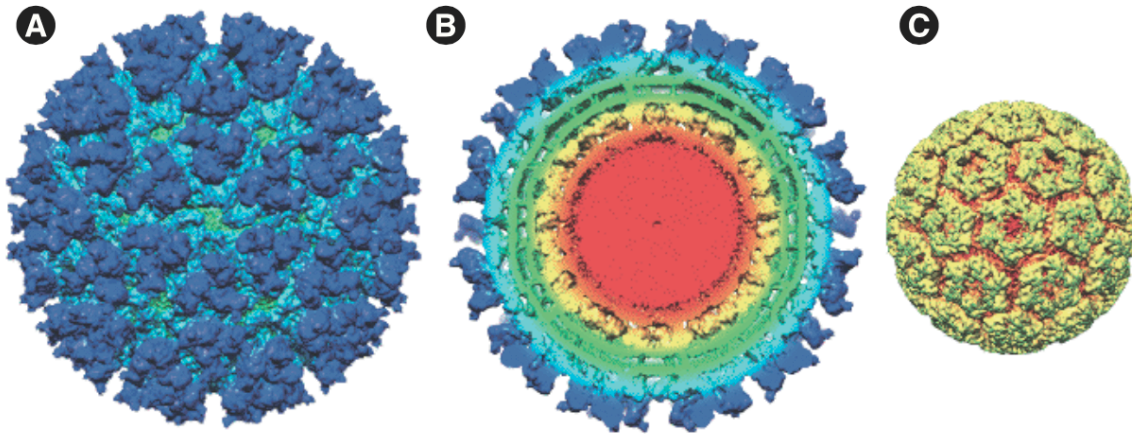


Figure 7: Chikungunya particle A) envelope B) cross section C) core (Jose et al. 2009).

The chikungunya virion is an enveloped spherical particle with a diameter of 70 nm. The core of the particle contains one single copy of (+)ssRNA packed into an icosahedral nucleocapsid (NC) shell formed by 240 capsid proteins arranged in a $T=4$ symmetry. Maintaining the symmetry of the NC the two envelope glycoproteins E1 and E2 are spanning the host-cell derived lipid bilayer in 80 trimeric spikes each composed of three heterodimers of E1 and E2. Each capsid protein interacts with one E2 moiety of the envelope protein heterodimers providing a very homogenous and compact structure of the virion (Sun et al. 2013) (Strauss & Strauss 1994).

2.1.7 Genome organisation

The genome of CHIKV is organised in a linear single-stranded positive-sense RNA of 11.8 kb. It possesses a 5' cap and a poly-adenylated tail (poly(A) tail). Therefore it contains the same structural elements as the cellular messenger RNA (mRNA) and can be directly translated by host-cell ribosomes (Gould et al. 2010).

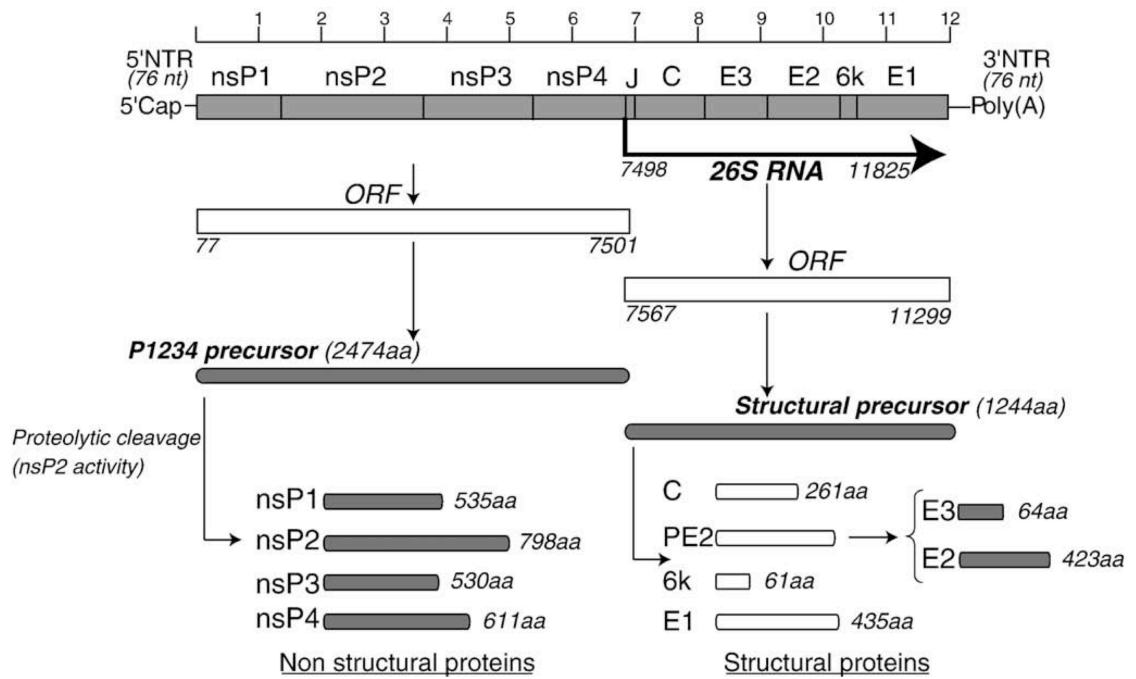


Figure 8: Genome organisation of CHIKV and RNA replication and translation (Solignat et al. 2009) reprinted with permission.

The genome is structured into two open reading frames (ORFs) containing the information for the non-structural proteins (nsPs) (ORF1) and the structural proteins (ORF2) respectively. The 5' cap and 3' poly(A) tail are not translated. The same applies to the junction region (J) separating the two ORFs. These regions are also called un-translated regions (UTR). The 5' cap is followed by the sequence of the polyprotein precursor P1234 encoding nsP1, nsP2, nsP3 and nsP4. The junction region contains the promoter for the subgenomic RNA to ensure the translation of the structural proteins from their subgenomic template, which corresponds basically to ORF2 (Strauss & Strauss 1994).

ORF2 encodes the following structural proteins: the capsid protein (C), three envelope proteins E3, E2 and E1 (E2 and E3 appear as PE2 precursor and are only cleaved late during the viral lifecycle), and a small protein 6k for which various auxiliary functions are suggested (Melton et al. 2002). The sequence of 6k is located between the envelope proteins PE2 and E1. ORF2 is only translated via the 26S subgenomic RNA that is synthesised by the replication complexes from the minus strand template after the recognition of an internal subgenomic RNA promoter within the junction region of the intermediate minus strand template (Strauss & Strauss 1994)(Solignat et al. 2009).

2.1.8 Chikungunya replication cycle

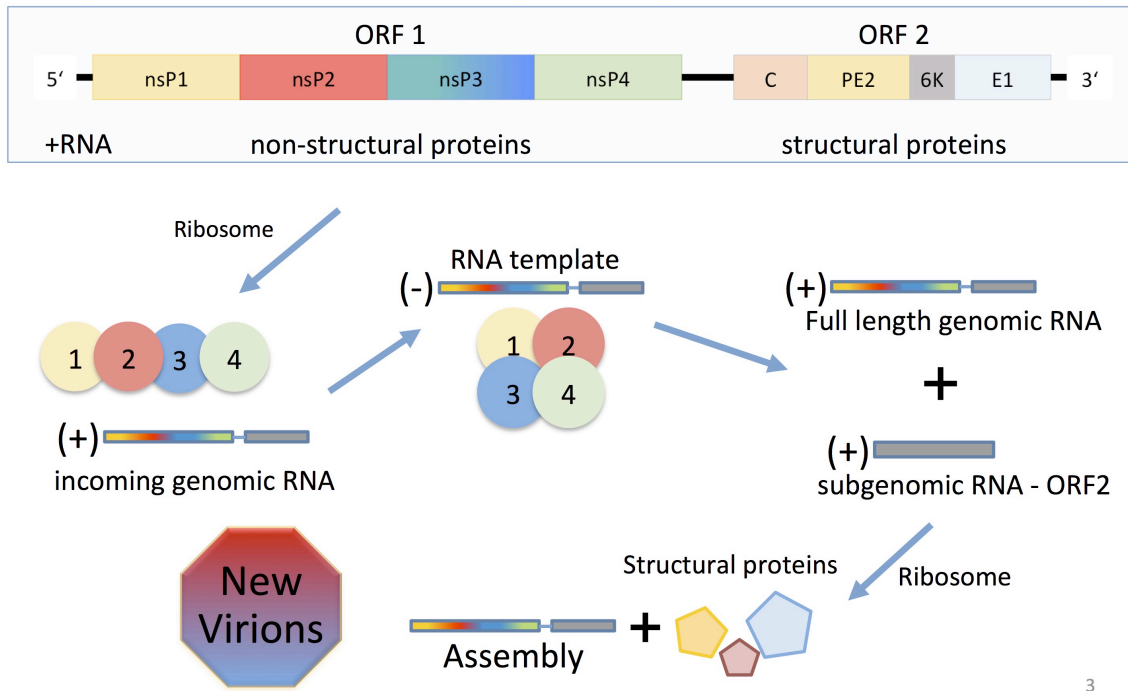


Figure 9: Schematic description of CHIKV genome replication and life cycle.

First, a CHIKV particle attaches to the cell membrane of a suitable host cell via the envelope protein E2 recognising specific receptors on the cell surface, which are not yet fully characterised/identified. This process causes curvature of the membrane leading to clathrin-mediated endocytosis of the viral particle (Hoorweg et al. 2016). After the entry the membrane of the virus fuses with the membrane of the surrounding endosome promoted by a decrease in pH in the endosome. The fusion of the membranes breaks open the virus envelope and releases the nucleocapsid core, which disassembles in the cytoplasm mainly by binding to free cytoplasmic ribosomes (Fields & Kielian 2015).

These ribosomes subsequently start to translate the non-structural proteins from ORF1 yielding either P123 + nsP4 or, by the read-through of an opal stop codon, P1234, which is quickly cleaved at the 3/4 site. P123 + nsP4 then form the early replication complex (RC), which undergoes two further cleavages. Cleavage leads to rearrangement of the nsPs within the RC. RCs early and late in infection possess different structure and produce all required viral RNA species in different stages of the infection (Strauss & Strauss 1994)(Rupp et al. 2015).

The early RCs composed of P123 and nsP4 first produce full length minus strand copies of the genomic RNA. It is likely that certain host factors are required for efficient replication, too (Rupp et al. 2015)(Schwartz & Albert 2010). The P123 precursor is then further cleaved by its own protease encoded in the nsP2 moiety resulting in intermediate RC nsP1+P23+nsP4. The cleavage of the polyprotein promotes a rearrangement of the RC. This intermediate RC is able

to synthesise both positive and negative stranded RNAs and prefers the full-length genomic RNA promoter to the subgenomic on the minus strand template. The intermediate RC can be barely detected in wild-type infections suggesting a very short half-life. First a cleavage at the 1/2 site occurs in *cis* (the protease on the same precursor cleaves its own 1/2 site) and immediately after that the cleavage of the 2/3 site in *trans* (a protease from a different precursor or a mature nsP2 protein cleaves the 2/3) takes place. The resulting late stage RCs are stable and indicate the advanced CHIKV replication by switching the RNA synthesis completely to the positive strand yielding genomic and subgenomic RNAs (Vasiljeva et al. 2003).

The subgenomic RNA is then translated by host cell ribosomes to produce structural proteins. Like the non-structural proteins they are synthesised as one polyprotein precursor. The capsid protein C is autocatalytically cleaved during the translation of the polyprotein and remains in the cytoplasm whilst a translocation signal is exposed on the remaining polyprotein leading to its translocation to the endoplasmic reticulum (ER). In the ER the precursor is then cleaved into E1, 6K and PE2. The passage through the Golgi apparatus induces glycosylation and palmitoylation. Before being presented at the cell membrane, PE2 is cleaved into E2 and E3 by the host enzyme furin (Uchime et al. 2013).

For the budding of viruses, the capsid proteins bind to one copy of full-length genomic positive sense RNA and self-assemble in the cytoplasm to form the nucleocapsid. The nucleocapsid then recognises regions in the cell membrane with a high density of envelope proteins and forms an interaction, which leads to the formation of the membrane enveloped viral particle (Strauss & Strauss 1994).

2.1.9 Viral proteins as potential targets

2.1.9.1 Structural Proteins

The structural proteins are synthesised from a positive strand subgenomic RNA transcribed from the complementary full length minus strand of the genome previously synthesised by the replication complexes. It is translated by the host cell ribosomes in the following order: 5'-C-PE2-6K-E1-3'. The structural proteins play important roles especially in the entry process and during budding.

2.1.9.1.1 Capsid protein

The capsid protein together with the genomic RNA forms the nucleocapsid core (NC), i.e. the inner part of the virion. During the transcription of the structural proteins the capsid protein is formed first and cleaves itself co-translationally and auto-catalytically from the nascent

polyprotein strand. Therefore it possesses an intrinsic protease at its C-term which resembles chymotrypsin-like serine proteases.

For CHIKV this proteolytic cleavage then stops the function of the capsid protease by a Trp residue arrested in the cleavage site (Aggarwal et al. 2015).

Furthermore the capsid protein possesses a binding site for the PE2 protein. A pocket at the C-terminal of the capsid protein is able to bind the cytoplasmic domain of E2 and connects the capsid core with the transmembrane envelope proteins (Wilkinson et al. 2005).

2.1.9.1.2 Envelope protein E1

The envelope glycoprotein E1 is located on the 3' end of the subgenomic RNA and is translated together with E3 and E2 and the small protein 6K from the structural polyprotein precursor. After the autocatalytic cleavage of the capsid protein, the remaining precursor protein is inserted to the endoplasmic reticulum for post-translational modification and for correct folding at and within the cell membrane.

E1 is classified as fusion protein and contains a so-called fusion loop to be inserted into membranes and refolds during the fusion of the viral membrane with the host cell endosome membrane during the entry of the virus. E1 is a trans-membrane protein with a short stem region on the inner side of the viral particle and a longer ectodomain on its outside. Both pre- and post fusion conformations are determined via X-ray diffraction for several alphaviruses including CHIKV (Sun et al. 2013)(Voss et al. 2010).

2.1.9.1.3 Envelope protein E2

E2 is first synthesised together with E3 in a segment of the polyprotein called PE2 or P62. This part is only cleaved shortly before the presentation of the glycoproteins at the cell membrane by a host enzyme called furin. As E1 it comprises a stem region on the inner side of the virion, a transmembrane segment and an ectodomain. E2 is responsible for binding to the capsid protein in order to form organised viral particles and to maintain the symmetry of the virion. Furthermore it is the attachment protein of the virus, recognising and binding to clathrin pits or other receptors on the cell surface in order to initiate the internalisation into the host cell. It is not directly involved in the fusion of the viral membrane with the host cell membrane but the correct interaction with E1 is required to form intact viral particles and plays a role in the budding of the virions. Dimerisation of E2 with E1 happens already in the ER and posttranslational modifications occur during the transit through the Golgi apparatus (Jose et al. 2009).

2.1.9.1.4 Envelope protein E3

The small glycoprotein E3 is formed by the 60-70 N-terminal amino acids of PE2. It is cleaved by the host cell enzyme furin previous to the presentation of the envelope glycoproteins at the cell membrane. After cleavage of the capsid protein it is responsible for folding and assembly of the envelope proteins during their maturation. After furin cleavage it is assembled into viral particles of some alphaviruses but not in the case of chikungunya (Uchime et al. 2013; Voss et al. 2010).

2.1.9.1.5 6k protein

The 6K protein is a short membrane protein located between the PE2 and the E1 protein within the precursor. It spans the membrane twice and the short linker between the transmembrane domains is palmitoylated. Due to its association with PE2 and E1 it reaches the cell membrane at the budding regions and is therefore incorporated to the virion particles but only in very small amounts. It has been shown that 6K can form pores in membranes and is suggested to be a virus encoded ion channel. Several studies have suggested different roles for 6K: regulation of glycoprotein trafficking and assembly, interactions with E2, enhancement of budding, membrane modification and alteration of the permeability of membranes, viroporin function or promotion of apoptosis due to interaction with caspase (Melton et al. 2002; Jose et al. 2009).

2.1.9.2 Non-structural Proteins (nsPs)

All non-structural proteins are transcribed by host cell ribosomes as a long polyprotein precursor. CHIKV produces two different types of precursors for the non-structural proteins. Usually P123 is produced because of the opal stop codon at the cleavage site of P3/4. But in some cases a read-through of the opal stop codon takes place yielding the complete polyprotein precursor P1234 comprising of all four non-structural proteins. If P1234 is produced, a rapid *trans*-cleavage at the 3/4 site takes place yielding P123 + nsP4. Later cleavages occur at the 1/2 site and then at the 2/3 site.

Although the non-structural proteins act together as the replication complex (RC) each of them has its specific role within the complex and some of them also exhibit separate regulatory functions i.e. interaction with host cell proteins (Rupp et al. 2015).

2.1.9.2.1 nsP1

nsP1 is the first protein of the 5' end of the genome. It is 535 amino acids long and functions as the capping enzyme of the virus. It exhibits two different enzymatic properties: the methyltransferase and the guanylyltransferase moiety. Both are required to cap the viral RNAs with a structure resembling the human RNA cap (Li et al. 2015).

The second important function of nsP1 apart from the capping process is to anchor the RCs in the host membrane. nsP1 is further suspected to play a role in membrane rearrangement during the viral infection. Two distinct sites contribute to that structural role: The amphipathic helix in the centre of nsP1 and the palmitoylation of cysteine residues in the 3' third of nsP1 (Rupp et al. 2015). Palmitoylation was shown to be responsible for the formation of filopodia in cell culture with infected cells but the importance of palmitoylation for CHIKV *in vivo* is yet to be determined (Laakkonen et al. 1998).

Within the RCs nsP1 directly interacts with nsP4 and nsP3. It is reported to play a role in the minus strand synthesis together with nsP4. The interaction with nsP3 was shown by co-immunoprecipitation but its function is unclear. It is likely that other interactions also occur but they might be less strong (Sreejith et al. 2012).

2.1.9.2.2 nsP2

The nsP2 protein is 798 aa long and can be divided into three different domains. The first one is the helicase domain, which is located at the N-terminus of the protein. Helicases unwind secondary structure elements within the RNA during the transcription process. nsP2 has also an ATP-binding site and possesses nucleoside triphosphatase activity. Therefore it is possibly interacting with the polymerase during the replication. Unfortunately there is no crystal structure available for the helicase domain of nsP2, which would facilitate rational drug design against that target (Karpe et al. 2011).

In the central part of nsP2 lies the protease domain. The catalytic residues are Cys 1013 and His 1083. Previously the nsP2 protease was assumed to be a papain-like protease and was classified as C9 peptidase (Russo et al. 2006; Vasiljeva et al. 2001). A recent study reinvestigated the CHIKV protease discovering a unique feature contradicting the papain-like protease hypothesis. Molecular dynamics revealed that Ser 1017 can be interchanged with the Cys 1013 of the catalytic dyad. Depending on the substrate either Cys or Ser form the partner in the dyad with His 1083. Further, Trp 1084 seems to play a role in substrate recognition and affinity (Saisawang et al. 2015). The viral protease is responsible for all cleavage steps of the P1234 or P123 polyprotein precursor. The cleavage sites are in positions 535/536, 1333/1334

and 1863/1864 and the protease is generally able to cleave substrates in *cis* and in *trans* depending on the site.

The third moiety of nsP2 is an inactive methyltransferase-like (MT-like) domain (Shin et al. 2012). It spans approximately the 200aa residues of the C-terminus of nsP2. Mutation analysis within that domain revealed its importance in the virus-host interaction and a possible role in minus-strand synthesis. Some of the mutants, conserved arginine and lysine residues were mutated to alanine, showed reduced minus-strand synthesis but the protease activity was not affected. Especially arginine 1141 or 1142 of CHIKV might play a role in host response inhibition, because mutations in that site caused non-cytopathic infections (Sawicki et al. 2006).

Finally nsP2 possesses a nuclear localisation signal (NLS) (aa 1182-1186), which might be responsible for translocating nsP2 into the nucleus. A free part of nsP2 was shown to migrate into the nucleus during infection and is believed to interfere with the regulation of cellular immune response towards the virus (Fros et al. 2013).

It can thus be concluded that nsP2 exhibits several different functions during the viral life cycle and that the different domains have distinct functions and work also independently from each other. The functions of the different domains of nsP2 in the replication complex and especially in the nucleus are very complex and have to be further explored.

2.1.9.2.3 nsP3

nsP3 is probably the most enigmatic of the nsPs. It is 530 aa long and can be subdivided into three domains. First the macro or X domain comprising the 160 N-terminal aa of the protein, followed by a central alphavirus unique domain (AUD) of 161 aa in length with a zinc binding site and finally the hypervariable region forming the C-terminal part of the protein.

The macro domain of CHIKV and VEEV was crystallised by (Malet et al. 2009) and several other studies were carried out to characterise the macro domain biologically. Like all macro domains it has a structurally conserved ADP-ribose binding site, which in some cases exhibits ADP-ribose phosphatase activity. It is also able to bind poly(ADP-ribose) (PAR), RNA, DNA and other adenosine containing molecules that are coordinated by Asp 10. Interestingly ADP-ribose and PAR/RNA do not extend towards the same binding pocket and mutations of Asp 10 did not completely abolish the binding of PAR and RNA to the protein. For CHIKV the ADP-ribose phosphatase activity was experimentally proved but it is not clear which role this function has in the replication of CHIKV (Malet et al. 2009).

The central part of nsP3 is characterised as AUD. This domain is not comparable to any other structural or functional domain and does not exhibit relevant sequence identity to other proteins. The crystallisation of a truncated P23 precursor of Sindbis virus (SINV) spanning the

protease and MT-like domain of nsP2 and the macro domain and AUD of nsP3 revealed a zinc-binding motif within the AUD. The zinc-ion very likely plays a role in stabilising the secondary structure of the complex (Shin et al. 2012). That structure also provides a valuable template for a homology model of the AUD of CHIKV. Several studies demonstrated the importance of the AUD for RNA synthesis and polyprotein processing and mutations in that part affected the neurovirulence of the virus in mice (Park & Griffin 2009a).

The C-terminal part of nsP3, named hypervariable region, is very difficult to characterise. Crystallisation of this domain failed so far and only sequence comparison and mutation studies give insights to potential functions of that domain. Several sequence elements are conserved but their pattern varies a lot, even among alphaviruses. The hypervariable region contains serine and threonine residues, which are phosphorylated to different extents and phosphorylation also varies during the progress of the viral life cycle. The phosphorylation is carried out by host cell enzymes and might provide a regulatory tool for the virus against the host cell (Varjak et al. 2009).

Several recent studies show important interactions of nsP3 with host enzymes responsible for the cellular stress regulation especially during viral infections. nsP3 prevents the assembly of stress granules by interaction and depletion of cellular enzymes G3BP1 and G3BP2. It is not clear if this mechanism is utilised by the virus to enhance its replication or just a simple defence against the host cell immune response (Scholte et al. 2015). It is clear that nsP3 plays an important role within the viral life cycle especially for the interaction with the host cell but further investigations are needed to clarify the exact mechanism of these interactions.

2.1.9.2.4 nsP4

The nsP4 harbours the RNA-dependent RNA-polymerase (RdRp) and is therefore solely capable of RNA synthesis. It is 611 aa long and despite the lack of a confirming crystal structure, sequence-annotation confidently matches the necessary domains of known RdRps (palm, thumb, fingers and the important GDD active site motive). The quantity of nsP4 produced from the precursor is less than that of the other nsPs due to the opal stop codon. Furthermore nsP4 is quickly degraded by the N-end rule if not bound to the RC (Bachmair et al. 1986). Although nsP4 seems to contain all necessary features of RdRp, it is not capable to synthesise RNAs without the other parts of the RC (Tomar et al. 2006).

Despite the fact that polymerases are quickly mutating it would be beneficiary for drug design purposes to gain a structure confirmation by crystallisation in order to define suitable sites for drug design and to study the interactions of all components within the RC (Rana et al. 2014).

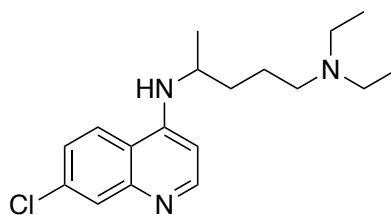
2.1.10 Current research on specific antivirals and vaccines against CHIKV

2.1.10.1 Antivirals in development against CHIKV

Structural proteins on the surface of the viral particle might be promising targets to neutralise the attachment and entry of CHIKV particles into the cells and thus prevent infection. The non-structural proteins of CHIKV are crucial to the viral replication as these proteins form the replication complex that copies the incoming viral RNA for protein production and genome replication for nascent virions. The strongly concerted action of these proteins makes each of them a promising target for antiviral compound design. Furthermore, host proteins are often utilised by the virus to facilitate entry or replication steps. These proteins might be hijacked by more than one virus, so that they can potentially serve as targets for broad-spectrum inhibitors.

Finally, many antiviral compounds are discovered by cell-based antiviral screenings and are further developed without the full understanding of the underlying mechanism of action. In the last years also repurposing screens of already marketed drugs proved to be a quick and efficient strategy to find antiviral compounds that could be quickly tested in humans because their safety was already proven (Ashburn & Thor 2004). This might be especially useful in an outbreak setting where quick interventions are urgently needed. Nonetheless, thus far no specific antiviral agent was marketed for the use against CHIKV. In the following section the recent and most promising antivirals in development for CHIKV are briefly described. This section summarizes crucial reviews that thoroughly collected data of new and established CHIKV inhibitors and prospective therapeutic approaches (Abdelnabi et al. 2015; Powers 2018; da Silva-Júnior et al. 2017; Kaur & Chu 2013).

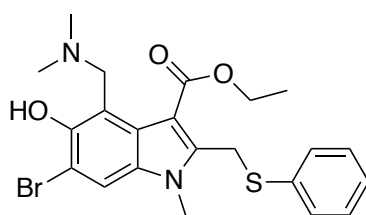
2.1.10.1.1 Entry inhibitors



Chloroquine

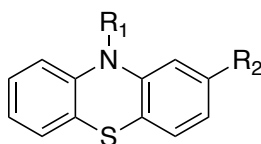
Chloroquine is an antiparasitic drug that is used to treat and prevent malaria. Furthermore, it is used in some cases of autoimmune disorders, particularly rheumatoid arthritis and lupus erythematosus where it may be used as second line treatment (DrugBank 2019). Historically, this drug was used against CHIKV as it was reported to be effective in human use in some cases. Further studies showed that chloroquine is effective *in vitro* and mode of action studies revealed a role in pH-dependent fusion of CHIKV envelope proteins with endosomal

membranes. A clinical study could not find a benefit in vivo among people infected with CHIKV. Therefore, an approval for the use against CHIKV is currently not expected (ClinicalTrials.gov ID:NCT00391313 2006).



Arbidol

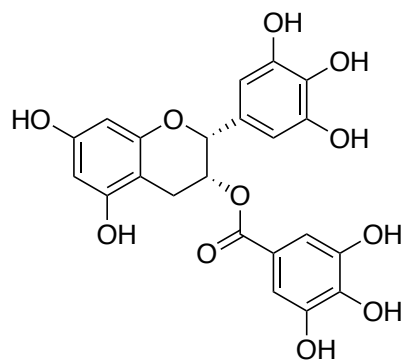
Arbidol is classified as an antiviral with a broad-spectrum activity and is approved to prevent respiratory infections and influenza in Russia and China. Delogu et al. demonstrated it to exert activity against CHIKV in cell-based assays. The mode-of-action of this compound seemed to rely on interference with attachment of the virus to the cell or by altering membrane structures that are relevant for the early steps of the viral infection (Delogu et al. 2011; Blaising et al. 2014).



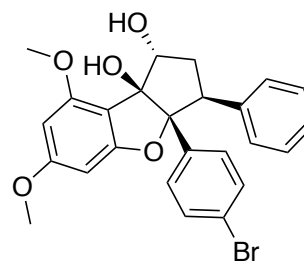
Phenothiazines

As a group of potential entry inhibitors several phenothiazine compounds (CNS active drugs) including chlorpromazine among others was tested against as entry inhibitors for its known activity to inhibit the formation of clathrin-coated pits. CHIKV is known to use clathrin-mediated endocytosis but does not uniquely rely on this mechanism to enter the host cell (Pohjala et al. 2011; van Duijl-Richter et al. 2015). Because of the expected CNS side-effects repurposing of these compounds as antivirals should be considered with caution. Compounds with activity against CHIKV are 10*H*-phenothiazines that possess an amine containing alkyl substituent (linear, branched or cyclic at its extreme) for R₁ and some of them contain lipophilic substitutions (-S-alkyl or -X) on R₂.

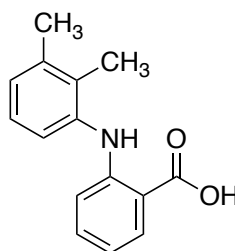
Other entry inhibitors



Epigallocatechin gallate



Flavagline FL3



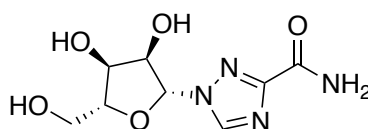
Mefenamic acid

Epigallocatechin gallate is a component of green tea and shows activity against the attachment and entry of several different viruses among them also CHIKV (Weber et al. 2015). Flavaglines are compounds that prevent the binding between CHIKV and a cell-surface receptor Prohibitin-1 (Wintachai et al. 2015). Mefenamic acid, an acidic non-steroidal anti-inflammatory drug and was found to be inhibiting CHIKV entry thus presenting an agent that can be used to treat CHIKV symptoms potentially with an additional benefit of a direct antiviral effect. Additional benefits can be seen when it is given in combination with ribavirin (Rothan et al. 2016).

Monoclonal antibodies

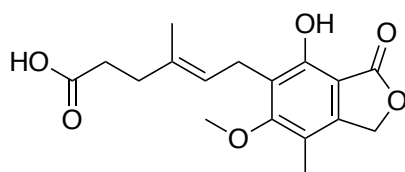
Another strategy to target entry of viral particles into cell is with antibodies. This strategy can be used in a therapeutic setting to prevent and treat viral infections efficiently. Antibodies recognizing epitopes on CHIKV envelope proteins were able to block CHIKV entry into cell as well as release of viral particles from the cells. IM-CKV063 presents an interesting candidate for further development (Jin et al. 2015).

2.1.10.1.2 Replication inhibitors



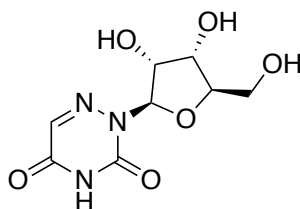
Ribavirin

The guanosine analogue ribavirin is a well-characterised broad-spectrum antiviral agent that is approved for respiratory syncytial virus in infants. Furthermore it is approved in combination with other antivirals and pegylated interferon α against chronic hepatitis C virus infections. Although exerting only moderate antiviral activity in vitro, it was found to have additional antiviral effects together with interferon $\alpha 2b$. As the mechanism-of-action of ribavirin the depletion of cellular GTP pools was proposed, which is a result of the inhibition of inosine monophosphate dehydrogenase (IMPDH). Furthermore, it might interact also with the capping mechanism of viral RNAs or be incorporated into the viral RNA strand abrogating viral RNA synthesis and replication (Franco et al. 2018).



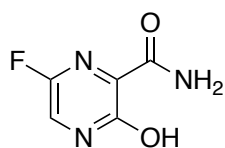
Mycophenolic acid

Like ribavirin mycophenolic acid inhibits IMPDH and depletes GTP pools, although via a different mechanism. For these properties it is approved as immunosuppressant after solid organ transplants. Unsurprisingly, also mycophenolic acid is active against CHIKV but the immunosuppressant properties might not be tolerated as a side effect in a compound marketed against viral infections (Diamond et al. 2002)



6-Azauridine

Comparably to the other two nucleoside-analogues 6-azauridine inhibits the enzyme responsible for the production of UTP. Both RNA and DNA viruses with a quick replication cycle are in need of a high amount of this nucleotide and the depletion of intracellular UTP is therefore fatal for the replication of these viruses (Rada & Dragúň 1977).



Favipiravir (T-705)

Favipiravir and its defluorinated analogue T-1105 both exhibit antiviral activity against CHIKV. Favipiravir was approved against influenza virus infections in Japan in 2014. It was studied against several other viruses and shows a broad spectrum of activity against different families of viruses, even towards Ebola virus (Oestereich et al. 2014). The intracellularly ribofuranosyl-5'-triphosphorylated active forms of favipiravir and T-1105 act on the RNA-dependent RNA-polymerase preventing the natural nucleotides ATP and GTP from being incorporated into the nascent RNA strand. To date both, chain termination and induction of lethal mutagenesis are proposed as mechanisms (Furuta et al. 2013; Delang et al. 2014).

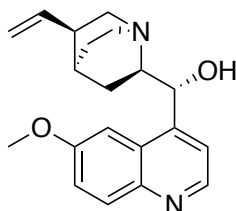
2.1.10.1.3 Protein synthesis inhibitors and compounds directly targeting CHIKV proteins

siRNAs and shRNAs

Small interfering RNAs are short stretches of single or double stranded RNA that is annealing with the target RNA strand in the for example during viral RNA replication. Against E1 protein and nsP3 small interfering RNAs were tested and reduced viral titers up to 99.6%. siRNAs that have nsP1 and E2 protein as targets were not as efficient in the reduction of virus titers. In vivo experiments with these siRNAs inhibited the replication of CHIKV when they were given 72h after infection. A limitation of siRNAs is owed to intracellular degradation (Rashad et al. 2014).

Short hairpin RNAs follow a similar principle and were developed against E1 and nsP1 and the capsid protein. The first was effective both in vitro and in vivo. Also nsP1 shRNA showed a strong inhibition of viral infection, whereas the shRNA against the capsid protein did not prove as effective. Prophylactic administration to C57BL/6 suckling mice resulted in complete protection from CHIKV and the survival rate after exposure was 100% after 15 days. Passaging in cell culture did not yield mutated shRNAs after 50 passages of the E1 shRNA (Rashad et al. 2014).

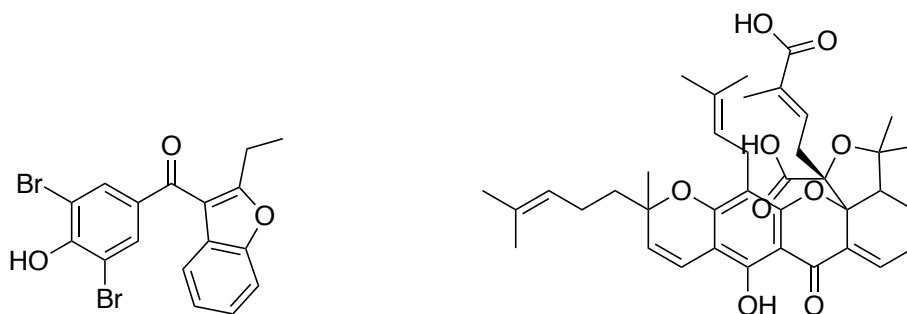
Compounds targeting nsP1



Quinine

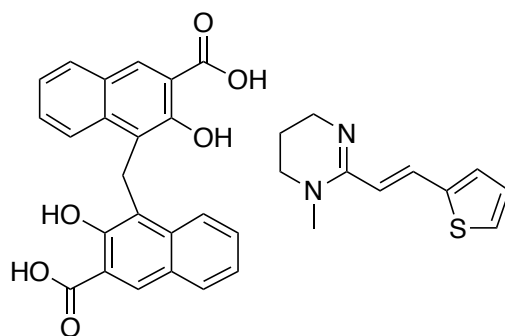
Quinine was suggested as an inhibitor of nsP1. The compound inhibits CHIKV *in vitro* and mutations arising against the compound map to nsP1 (Di Mola et al. 2014; de Lamballerie et al. 2012).

Very recently, compounds that directly target the nsP1 were identified from a screening of an FDA approved compound library. From this study, four compounds that target both dengue virus and chikungunya virus capping machinery (nsP1 for CHIKV) were identified:

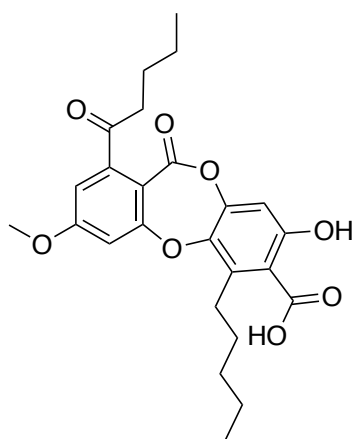


Benzbromarone

Garcinolic acid



Pyrantel pamoate



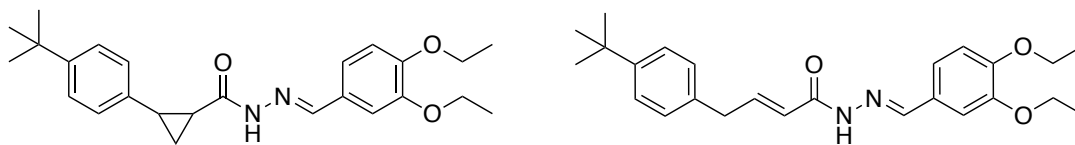
Lobaric acid

Benzbromarone, garcinolic acid, pyrantel pamoate, and lobaric acid inhibited nsP1 of CHIKV with 7.0 ± 0.6 , 14.5 ± 4.9 , 13.0 ± 0.9 and 5.0 ± 0.3 μM K_i , respectively. Lobaric acid also demonstrated activity against CHIKV *in vitro* in a low micromolar range (Feibelman et al. 2018).

Compounds targeting nsP2

A hit identified in a computer-aided structure-based virtual screening approach against nsP2 yielded a compound with low micromolar activity against CHIKV in cell-based assays. Docking

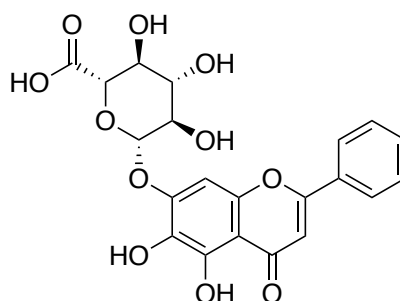
studies suggest the protease function of nsP2 as a target and synthesised analogues yielded compounds with improved activity (EC_{50} 3.2 μ M) (Bassetto et al. 2013).



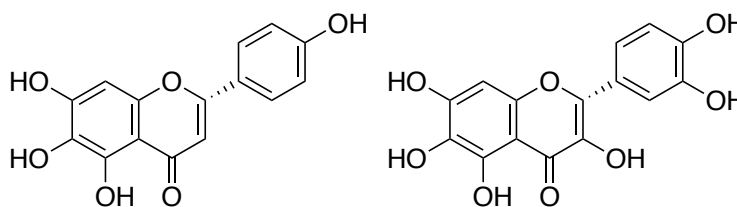
Compounds from Bassetto et al. 2013

Lucas-Hourani and co-workers identified compound ID1452-2 against nsP2 targeting the host-cell transcription shutoff with an EC_{50} of 31 μ M (Lucas-Hourani et al. 2013). Singh et al. identified potential compounds against nsP2 by *in silico* screenings but they did not test the compounds *in vitro* (Singh et al. 2012).

Compounds targeting nsP3



Baicalin



Naringenin

Quercetagetin

Flavonoids were identified as a group of compounds of which several representatives reportedly target the CHIKV nsP3 protein. In a docking study the binding affinities for Baicalin, Naringenin, and Quercetagetin were calculated based on a previous re-docking of the co-crystallised ligand ADP-ribose. Indeed, the best-ranked compound Baicalin (Seyedi et al. 2016) was found to directly bind to the macro domain of nsP3 *in vitro* (personal communication, Bruno Coutard and Ana Sofia Ferreira Ramos).

2.1.10.1.4 Host targeting compounds against CHIKV

Furin inhibitors

Furin inhibitors are compounds that inhibit the host protease furin, which is responsible for the cleavage of the envelope protein precursor into their active individual proteins. CHIKV

strictly requires this enzyme to produce mature virions. Decanoyl-RVKR-chloromethyl ketone inhibits furin irreversibly thus preventing the maturation and spread of mature CHIKV virions (Ozden et al. 2008).

Kinase modulators

Kinases play important roles in the physiology of human cells. Several compounds that act upon cellular kinases were found to inhibit the replication of viruses. Among them are the protein kinase C activators prostratin and 12-O-tetradecanoylphorbol 13-acetate (TPA) that were found to be effective against CHIKV. Furthermore, several kinase inhibitor scaffolds were reported to yield active CHIKV inhibitors. These compounds had benzofuran, pyrrolopyridine or thiazol-carboxamide scaffolds. The mechanisms involved in this process are not fully understood yet (Bourjot et al. 2012; Cruz et al. 2013; Abdelnabi et al. 2017).

HSP-90 inhibitors

The heat shock protein 90 (HSP-90) is a cellular chaperone molecule that is produced in different isoforms and interacts intracellularly with many proteins in physiological procedures. Isoform α is particularly expressed during cellular stress situations. It was found to directly interact with nsP3 and nsP4 of CHIKV, thus it was not surprising that treatment with HS-10 and SNX-2112, two HSP-90 inhibitors, resulted in inhibition of CHIKV *in vitro* and *in vivo* (Rathore et al. 2014).

Furthermore several compounds and strategies to modulate the immune system of the host were employed to treat CHIKV infection. Among them, are interferon α (also in combination with ribavirin); viperin, a host protein that is targeting virus clearance from the cells; poly I:C, a compound resembling double stranded RNA that stimulates the immune system and many more. Boosting the immune system to support viral clearance can be a successful strategy. Other specific host targets must be investigated with great care as side effects might occur more frequently and might be more severe than by targeting viral proteins (Abdelnabi et al. 2015; da Silva-Júnior et al. 2017).

2.1.10.2 Vaccines

Although CHIKV vaccine development efforts started already in the 1960s just over one decade ago the field received a boost due to the re-emergence of CHIKV from 2006 onwards. Early vaccine candidates included several inactivated preparations from different cell-cultures and different CHIKV strains (Powers 2018). The most promising at that time were the 15561 strain, a formaldehyde inactivated preparation (Harrison et al. 1971) and a live attenuated vaccine

candidate 181/25 (TSI-GSD-281), developed on its basis (Levitt et al. 1986). This live attenuated vaccine passed clinical phase I and phase II showing in 98% of the subjects neutralizing antibodies that in 85% of the cases persisted for one year (Edelman et al. 2000). In 2012, it was found that this strain is attenuated only by 10 nucleotide changes within the whole genome, five of which are synonymous and only three lie within the coding region of the envelope proteins. Thus, there were concerns of reversion back to wild type, and reversions were observed in mice when taking samples and sequencing them (Gorchakov et al. 2012). For research purposes, the investigatory new drug (IND) protocol for TSI-GSD-281 was kept until 2011 and samples were vialled and stored. In 2006, facing the threat of a CHIKV epidemic in La Reunion the French government requested material from the US Department of Defense for the further development of this candidate (Hoke et al. 2012). So far, this vaccine has not reached the market but there are reports of research and development with this vaccine strain especially in India (Powers 2018).

In the meantime, many other strategies in vaccine development were also applied to obtain a new CHIKV vaccine candidate. The approaches include virus-like particles (VLPs), subunit vaccines, vector- or chimeric vaccines and nucleic acid vaccines and live attenuated vaccines.

The most advanced candidates are two VLP vaccines PXVX0317 CHIKV-VLP by Pax Vax (ClinicalTrials.gov ID:NCT03483961 2018) and VRC-CHKVLP059-00-VP by NIAID (ClinicalTrials.gov ID:NCT02562482 2015) and one MV-CHIKV vector vaccine (developed by Themis Bioscience GmbH) that all completed phase I clinical trials successfully (ClinicalTrials.gov ID:NCT01489358 2016). The identifiers of phase II clinical trials for the former agents and study titles are listed in Table 1.

Table 1: Vaccine candidates in clinical trials phase II

Vaccine candidate	Study Title	Trial Identifier from ClinicalTrial.gov	Year
MV-CHIKV	Double Blinded, Randomized, Priorix®- and Placebo-controlled, Trial to Evaluate the Optimal Dose of MV-CHIK Vaccine (Against Chikungunya Virus) in Regard to Immunogenicity, Safety and Tolerability in Healthy Volunteers	NCT02861586	2016
	Phase 2 Study of a Live Attenuated Measles Virus-Vectored Chikungunya Vaccine in a Previously Epidemic Area	NCT03101111	2017
	Observer Blinded, Randomised Study to Investigate Safety, Tolerability and Long-term Immunogenicity of Different Dose Regimens and Formulations of MV-CHIK in Healthy Volunteers	NCT03635086	2018
	Phase 2 Study of a Live Attenuated Measles Virus-Vectored Chikungunya Vaccine in Previously Exposed Adults	NCT03807843	2019
PXVX0317 CHIKV-VLP	A Phase 2 Parallel-Group, Randomized, Double-Blind Study to Assess the Safety and Immunogenicity of PXVX0317 (Chikungunya Virus Virus-Like Particle Vaccine [CHIKV-VLP], Unadjuvanted or Alum-adjuvanted)	NCT03483961	2018
VRC-CHKVLP059-00-VP	Phase 2 Randomized, Placebo-Controlled Trial to Evaluate the	NCT02562482	2015

	Safety and Immunogenicity of a Chikungunya Virus-Like Particle Vaccine, VRC-CHKVLP059-00-VP, in Healthy Adults		
--	--	--	--

Several other candidates are currently evaluated in pre-clinical studies and phase I clinical trials. Detailed discovery and conceptualisation of these agents was reviewed by Powers (2018). A detailed status of the new vaccine candidates recently under phase I clinical investigation including their identifiers is reported in the publication of Goyal et al. (2018).

In conclusion, the vaccine development efforts for CHIKV look very promising and one would expect that at least one vaccine preparation would reach the market in the upcoming years. This would mark a big milestone not only towards the protection of naïve populations and travellers but also towards the reduction of import of CHIKV into new territories. Most importantly, it would be a step to reduce the impact of CHIKV related illness on the socioeconomic status of affected individuals and countries.

2.2 Aims

Although several active compounds with diverse mode of actions against CHIKV were discovered in the past, none of the direct-acting antivirals has reached the market with an approval for CHIKV as an indication, yet. It was therefore one of the goals of this study to discover, improve and characterise new antiviral compounds with a good activity against CHIKV and low cytotoxicity in the cell-based assays. The prerequisites were the crystal structures of parts of nsP2 and nsP3, of which primarily nsP3 was chosen as the main target.

In the first part the complete binding site of nsP3 macro domain was targeted with a structure-based pharmacophore screening approach. The hits would be tested in cell-based antiviral assays and active compounds would be used to explore the chemical features necessary for the antiviral activity of the compounds. In collaboration with the Aix-Marseille Université binding assays on the CHIKV macro domain could be performed to confirm the mode-of-action of the designed compounds.

In a second pharmacophore screening, only the part occupied by the distal ribose of ADP-ribose was targeted. The promising hit compounds would again be tested in cell-based antiviral assays.

Furthermore, a crystal structure of parts of the P23 precursor protein was available for Sindbis virus, a closely related Alphavirus. Homology modelling would be used to obtain the P23 precursor for CHIKV. The CHIKV P23 precursor would be investigated for potential new ligand

binding sites, which would then be targeted using a high-throughput virtual screening approach to screen a large compound library in order to identify potential hit compounds on a completely new binding site.

Active compounds would be evaluated in further experiments for their activity on other viruses and for the confirmation of the target and mode-of-action.

2.3 Targeting CHIKV nsP3 macro domain

One aim of this thesis is to discover new antiviral small-molecules against CHIKV. In the introduction about CHIKV the different structural and non-structural proteins were described. The non-structural protein nsP3 is the main target protein in this work. NsP3 can be divided into three distinct parts: the N-terminal globular macro domain, the central alphavirus unique domain (AUD), and the C-terminal unstructured region that varies in lengths for different alphaviruses. The main focus of this chapter lies on the CHIKV macro domain.

2.3.1 Macro domains

Very generally speaking, macro domains are protein modules that are conserved among all kingdoms of life and they are mainly involved in the binding and processing of ADP-ribose (ADPR). They were discovered as conserved proteins of unknown function among several positive-sense single-stranded RNA, human pathogenic viruses, such as severe acute respiratory syndrome-related coronavirus (SARS-CoV) and Middle East respiratory syndrome-related coronavirus (MERS-CoV), Rubella virus, Hepatitis E virus (HEV), and alphaviruses. All of them possess a conserved macro domain within their non-structural proteins, formerly named X domain (Koonin et al. 1992). Only later when a homologous domain in the human histone H2A was discovered the name macro domain was established (Allen et al. 2003). The human genome encodes for seventeen macro domain-containing proteins, acting in several different regulatory pathways, such as DNA-damage sensing and repair, remodelling of chromatin, proliferation, transcription and cell signalling (Fehr et al. 2018; Barkauskaite et al. 2015).

Macro domains are often coupled with other structural building blocks in multi-domain proteins and, depending on their arrangement in the protein sequence and to which other structural elements they are coupled, they fulfil different purposes. Macro domain is not the only structural element involved in the regulation of ADP-ribose as post-translational modification (PTM): among others, the WWE domain and the PBZ protein should be mentioned, which are degrading PAR-chains at different points resulting in cleavage products that differ from ADP-ribose (Barkauskaite et al. 2015; Palazzo et al. 2017). The processes of addition and removal of ADP-ribose onto/from target proteins is depicted in Figure 10.

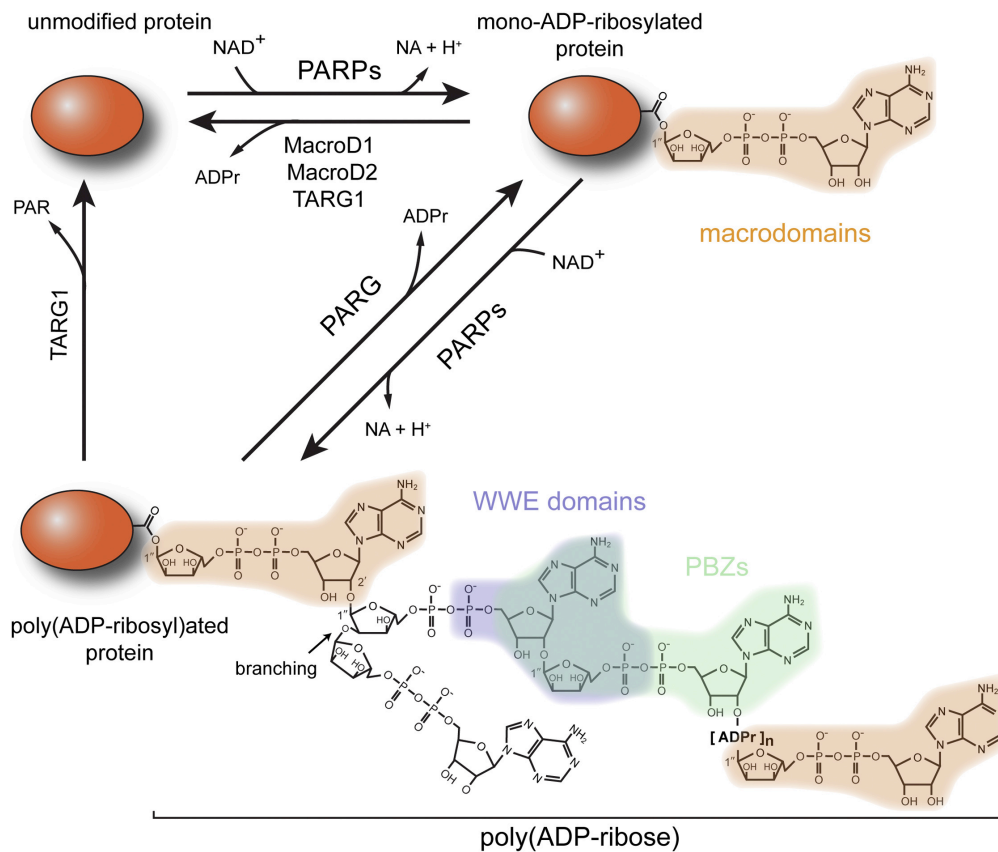


Figure 10: Cycle of ADP-ribosylation

NAD⁺ acts as a co-factor for the ADP-ribosylation of proteins by PARPs. PARG and TARG1 remove ADP-ribose from proteins releasing either ADP-ribose or PAR. WWE domains and PBZs cleave PAR at different points compared to macro domains. Figure with permission of Cell Press (Barkauskaite et al. 2015).

Macro domains can be classified into different groups according to their structure, enzymatic activity, and their function. Most macro domains are members of the MacroD group, which includes also viral macro domains, but there are three other groups, namely ALC1-like macro domains, macro H2A-like, and PARG proteins. They differ in the substrate specificity and in the reactions they are able to catalyse (Fehr et al. 2018).

In human cells, macro domain proteins are classified into a group of proteins that are called ADP-ribose transferases (ART) or polyADP-ribose polymerases (PARP). Generally, the nomenclature of proteins involved in ADP-ribose-handling as posttranslational modifications is inconsistent, reflecting mostly the context of their discovery, and should be unified according to experts in the field (Hottiger et al. 2010). Some macro domains are able to synthesise polyADP-ribose (PAR) chains linking them to specific residues of proteins (polymerases), others are only capable of transferring monoADP-ribose (MAR) generated from NAD⁺ onto their target protein (APD-ribose transferases). The third group possesses enzymatic activity to remove either mono- or polyADP-ribose from proteins (like polyADP-ribose glycohydrolase (PARG), MacroD1 and D2 or terminal ADP-ribose glycohydrolase (TARG1)). Other macro domain containing proteins are only binding MAR or PAR as PTM, which may lead to an assembly of protein complexes – macro domain containing proteins can serve as hubs for

protein-protein interactions – or conformational changes that affect an effector domain, which transmits information via this pathway. Thus, macro domain containing proteins can be readers, writers or erasers of ADP-ribose post-translational modifications. The PTM ADP-ribose and its derivatives is involved in a vast amount of cellular regulatory pathways (Palazzo et al. 2017).

MonoADP-ribosylation (MARylation) and polyADP-ribosylation (PARylation) play also a role when pathogens and host cells interact. Certain PARPs are up-regulated by transcription factors, which results in an increased ADP-ribosylation of cellular proteins. The fact that viruses encode a protein domain that can interact specifically with ADP-ribosylation makes it conceivable that this protein counteracts the cellular reaction to viral infection. The known viral macro domains are recognising ADP-ribosylated proteins and could remove MAR or PAR from the protein in question. Both viral and host-cell macro domains might therefore be interesting targets to investigate in the context of infection, to first understand the cellular context, and to make use of this knowledge to combat the virus. The slowly evolving and highly conserved structure and the various functions of this ubiquitous structural module may play an interesting role, especially at the interplay between a pathogen and its host (Vivelo & Leung 2015; Feijs et al. 2013; Rosenthal et al. 2013; Karlberg et al. 2013; Schreiber et al. 2006).

2.3.2 Properties of viral macro domains

Viral macro domains were first discovered through sequence analysis and comparison revealing this new conserved structural domain with previously unknown functions (Koonin et al. 1992). Meanwhile, the structure of several viral macro domains was resolved and well characterized, and their enzymatic function was investigated in several biochemical studies. The viral macro domains of alphaviruses such as CHIKV, O'nyong'nyong virus (ONNV), SINV, and Venezuelan Equine encephalitis Virus (VEEV), orthohepevirus Hepatitis E virus (HEV) and severe acute respiratory syndrome (SARS) and middle east respiratory syndrome (MERS), both coronaviruses, have been studied in greater detail to characterize their biochemical properties, also in comparison with human macro domains (reviewed in Fehr et al. 2018). Viral macro domains exhibit ADP-ribose-1''-phosphate monophosphatase activity, first discovered for a homologous protein in yeast (Kumaran et al. 2005). ADP-ribose-1''-monophosphate is a small molecule that results from tRNA splicing, a process predominately found in yeast, but it is not clear why viruses should devote a gene for this purpose (Fehr et al. 2018; Palazzo et al. 2017). Later it was discovered that viral macro domains also bind mono- and poly-ADP-ribose, a PTM often attached to proteins via acidic amino acids like glutamate and aspartate. Therefore, the observed phosphatase activity might be a coincidental effect but not the main enzymatic activity, which seems to be the cleavage of the ester bond between the 1'' distal ribose and

acidic amino acids on the MAR or PARylated protein. The binding cleft of viral macro domains is also able to accommodate other adenine containing molecules such as 2'-5'-oligoadenylate or small fragments of RNA (Malet et al. 2009; Ecker et al. 2017; Li et al. 2016).

Viral macro domains are essential for viral infection. The most important residues for substrate specificity and enzymatic function were determined by mutation studies on the different viral macro domains and the effects are summarized in Table 2 (Leung et al. 2018).

Table 2: Macro domain mutations and their effects (adapted from Leung et al. 2018)
Colour code matches the corresponding residues of CHIKV macro domain in Figure 11.

	Virus	Macro domain mutant	Phenotypes	References
Alphaviruses	CHIKV	D10A, G32E, G112E	Mutants abolished ADP-ribosylhydrolase activity totally or partially in BHK21 cells and revert to wt. G32E also reverts in C6/36 mosquito cells	(McPherson et al. 2017)
	CHIKV	G32S, G32A, T111A, Y114A	Virulence in mice and replication depends on ADP-ribosylhydrolase activity	(McPherson et al. 2017)
	SINV	N10A/N24A	Slower replication in neurons compared to BHK21 cells, revert to wt in neurons. Mature neurons have reduced cell death and reduces SINV RNA synthesis, not observed for immature neurons. In 2-week-old mice mutant showed attenuated virulence, not observed in 5-day-old mice.	(Park & Griffin 2009b)
Coronaviruses	SARS-CoV	N1040A	Slower growth, while reaching similar titers eventually compared to wt. IFN α and IFN γ sensitivity enhanced compared to wt.	(Kuri et al. 2011)
	SARS-CoV	D1022, N1040A, H1045A, G1130V	Less virulent and viral load decreased in vivo, mutants replicate like wt in culture. N1040A shows enhanced IFN and cytokine response in early infection in vivo. Co-infection with N1040 and a lethal SARS-CoV protects mammalian hosts.	(Fehr et al. 2016)
	Human CoV 229E	N1305A	Slower growth and heightened IFN α response with respect to wt.	(Kuri et al. 2011)
	Human CoV 229E	N1305A	Similar growth in culture, no differences in RNA synthesis	(Putics et al. 2005)

			observed.	
	Mouse hepatitis virus strain JHM	N1348A	Only slightly reduced replication in vitro. Reduced virulence, titers, cytokine/chemokine expression in mice.	(Fehr et al. 2014)
	Mouse hepatitis virus strain A59	N1348A	No acute hepatitis in mice, replicates like wt, induces less TNF α and IL-6.	(Eriksson et al. 2008)
Orthohepevirus	HEV	G48V	Tested in HuH-7/S10-3 cells using GFP-replicon constructs. Viability observed.	(Parvez 2015)
	HEV	N38A, G48A	GFP-based replicon constructs, replication slower than wt in HuH-7/S10-3 cells	(Parvez 2015)
	HEV	N42A, H45L, G49A, G49V, G50A, G50V	GFP-based replicon constructs, not viable	(Parvez 2015)
	HEV	N42A, G50A, G48S/G49S, G48S/G49S/G50A, G123A	ADP-ribosylhydrolase activity reduced, lower replication rate in HuH-7 cells, luciferase-based replicon constructs.	(Li et al. 2016)

Mutations have reduced infectivity in certain cell types. In vivo infection with mutated macro domain-containing virus reduced the virulence of the infection in mice. Examples show that viral macro domains mutants affect virus replication as well as interferon-response and cytokine/chemokine expression. Furthermore, they may counteract human PARPs of which several are reported to contribute to antiviral defence mechanisms. MAR and PAR also serve as an attachment point for proteins that are only “readers” of this PTM, and viruses potentially exploit certain cellular proteins that they can bind, thus serving as a hub for protein-protein interactions. This function might be exploited in stress granules where ADP-ribosylated proteins might be utilised to enhance viral replication. This topic was very comprehensively reviewed by Fehr and co-workers and Leung and colleagues (Fehr et al. 2018; Leung et al. 2018).

2.3.3 Structure of CHIKV macro domain

The macro domain of CHIKV (PDB: 5GPO and 5GPQ) was co-crystallised with ADP-ribose, with a short piece of RNA, and in the apo-form (5GPG) (Malet et al. 2009). Another structure was released in the PDB containing 2'-5'-Oligoadenylate as a co-crystallised ligand (4TU0) (Morin et al. 2014) but it will not be further discussed in this thesis.

CHIKV macro domain presents an $\alpha/\beta/\alpha$ composition. A central six-stranded mixed and twisted β -sheet is surrounded by three α -helices on one and one α -helix on the other side. Usually macro domains contain at least five α -helices, but due to a deletion in the region of residue 48, the fifth helix is not present in the CHIKV structure, when compared to SARS-CoV.

CHIKV macro domain was biochemically analysed for its ability to bind different adenine containing molecules, and results were verified by mutations to propose a structural conclusion for the mechanism of binding and enzymatic activity, pinpointing the contributions of single residues. Moieties tested in a thermal shift assay (TSA) were ADP-ribose, ADP, ATP and NAD⁺ as well as ADP-glucose. ADP-containing molecules were binding successfully to the protein and increasing the melting temperature of the protein. The binding of adenine containing molecules is directly related to aa D10, which is coordinating the N6 of adenine. This was further confirmed by mutation studies.

The diphosphate perfectly fills the crevice between the catalytic loop and the phosphate-coordinating loop and a mono- or triphosphate in the same position was not optimal. Finally, the ribose is the ideal sugar on the distal end, as it can be positioned perfectly between the tyrosine and the catalytic loop.

Malet and co-workers conclude that the protein exhibits a higher specificity towards adenine as a base in comparison with guanine, fits two phosphate groups best, and prefers a ribose to a glucose at the distal position. Therefore, ADP-ribose exhibits highest thermal shift values and a strong binding in the ITC assay (Malet et al. 2009).

The coordination of ADP-ribose and a small fragment of RNA in the binding site is depicted in Figure 11. First, the residues coordinating the adenine base of the molecule are D10 with N6 of adenine, R144 (omitted in Figure 11) and G32, which is part of the catalytic loop. The proximal ribose is further H-bonding with T111. In this area several water molecules are present in the crystal structure, which form a network of hydrogen bonds that is fixing the distal part of the molecule in its presented conformation. The phosphate-binding loop is forming several interactions with the ADP-ribose molecule, mainly involving the backbone NH-groups of G112, V113 and Y114 in hydrogen bonds that, with their positive charge, accommodate the negatively charged phosphate groups readily in this area of the crevice. The distal ribose lies between D31, N24, which form protein backbone interactions with the OH-groups of the ribose, and Y114 that holds the distal ribose in place from the top (Malet et al. 2009).

Superimposition of 3GPO and 3GPQ

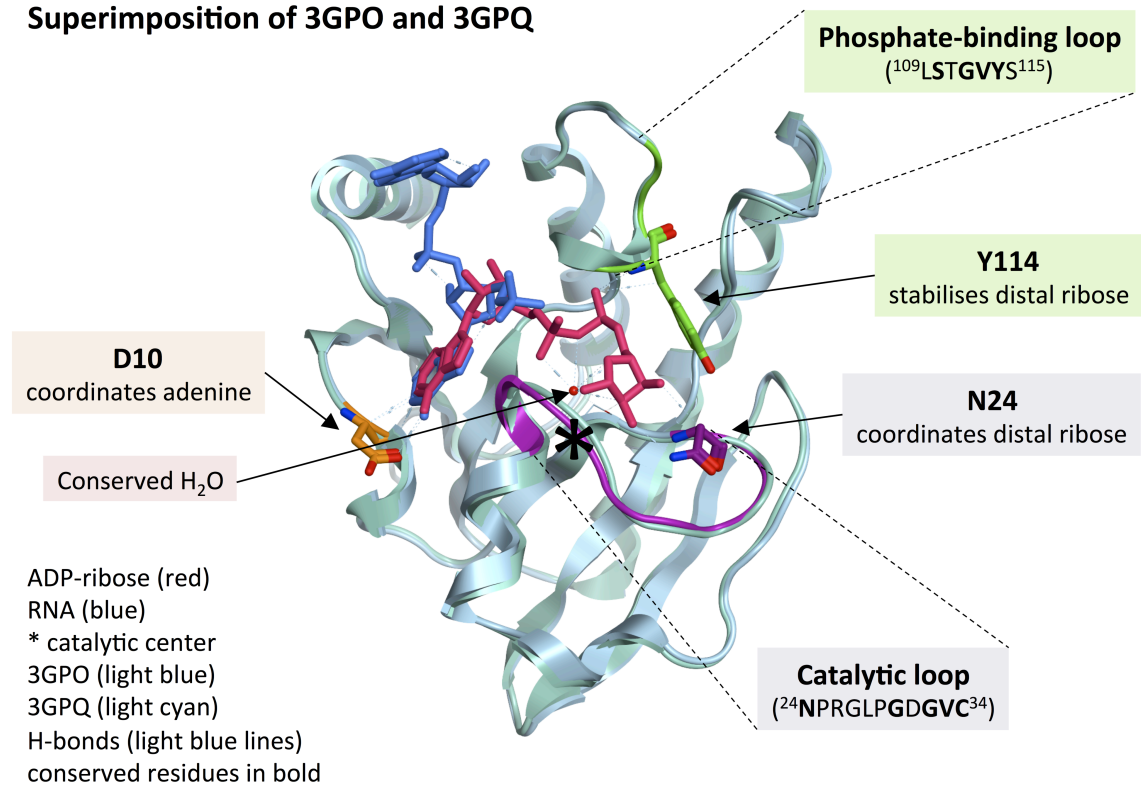


Figure 11: CHIKV macro domains 3GPO and 3GPQ superimposed

ADP-ribose in red, small RNA fragment in blue, D10 orange, catalytic loop and N24 purple, phosphate-binding loop and Y114 green, * demarks the catalytic centre, red dot represents a conserved structural water molecule, light blue dashed lines are H-bonds, 3GPO coloured in light blue and 3GPQ in light cyan.

2.3.4 Binding and enzymatic functions of CHIKV macro domain

In order to validate their findings, Malet et al. first assessed the binding of different adenosine-containing molecules and some other compounds that might fit into the binding site of CHIKV macro domain with TSA and ITC. The compounds are binding with descending strength assessed by an increase in the melting temperature of the protein: ADP-ribose, ADP, ATP, NAD, SAH, AMP, GDP, MgCl₂, ADPG, MnCl₂. To evaluate which residues were responsible for the binding, mutations were introduced in the conserved positions known as being important in other macro domains as well. Above all, D10 is important for ADP-ribose binding (Malet et al. 2009).

In the next step the authors wanted to evaluate the APD-ribose 1''-phosphate phosphatase activity that was first found in Poa1p in yeast (Martzen et al. 1999) and also detected for other macro domains in SARS-CoV and HEV (Egloff et al. 2006). The activity of CHIKV macro domain to cleave the phosphate group off the distal ribose was comparable to the one found in yeast Poa1p, whereas the macro domains of SARS-CoV and HEV had a much slower turnover. The D10 mutation did not abolish, but decreased the activity and N24 and Y114 mutants were inactive. N24 was also proposed to be the crucial residue involved in the cleavage of ADP-

ribose from proteins that are linked via an aspartic or glutamic acid ester to the distal ribose in 1'' position. Involved is also a structurally conserved water molecule that is activated by the proximal phosphate group of ADP-ribose to perform a nucleophilic attack on the C1'' of the distal ribose (Li et al. 2016). A conclusive mechanism still needs to be determined for CHIKV macro domain. Comparing the O-acetyl cleavage demonstrated by Kumaran et al. using H₂¹⁸O and the mechanism reported for Macro D2 by Jankevicius et al., using a similar method, the catalytic mechanism for CHIKV macro domain might be the one postulated by Li et al., but experiments to confirm that are still needed (Kumaran et al. 2005; Jankevicius et al. 2013; Li et al. 2016).

2.3.5 Aims

No specific antiviral compounds against CHIKV have reached the market so far. Non-structural proteins are well-established targets for the development of new antiviral compounds against viral infections. In the introduction of this chapter the information about CHIKV and other macro domain is summarised. This information was used to target CHIKV nsP3 macro domain in order to disrupt CHIKV replication. The influence of nsP3 during replication is absolutely critical for viral replication. The detailed mode-of-action and other auxiliary functions are not fully elucidated, yet. Therefore, new antivirals targeting nsP3 might also contribute to the understanding of the proteins functions.

In this chapter the whole ADP-ribose binding site was investigated. Pharmacophore searches were used to select molecules for a docking and consensus scoring approach. The compounds selected with the computational screening were then tested in cell-based antiviral assays. The resulting active compounds served as the starting point for further chemical optimisation and characterisation of the antiviral activity with biochemical and virological assays, with the goal to find new antiviral compounds and to gain insights into their mode of action.

2.3.6 Results and Discussion

2.3.6.1 Screening on CHIKV macro domain for new small molecule inhibitors

The main aim of the present project was to find small molecules against CHIKV in order to hinder the virus replication and to prevent the spread of the virus. For the target nsP3 macro domain the workflow to pursue that aim is depicted in Figure 12.

Several crystal structures are available in the PDB for the nsP3 macro domain of CHIKV, with and without ligands. The structures used for this project are the apo-protein 3GPG, 3GPO and 3GPQ with ADP-ribose, and a small RNA fragment co-crystallized, respectively (Malet et al.

2009). They were all retrieved from the PDB using Molecular Operating Environment (MOE) 2015.10 (Chemical Computing Group Inc. 2016) and prepared for further use.

A consensus pharmacophore model was built based on the two ligands ADP-ribose and RNA and subsequently used to screen the SPECS library (<http://www.specs.net>) of commercial compounds. The resulting compounds were then explored for their possible conformations and these conformations were docked into the original binding pocket. The docking poses were evaluated by rescoring them with 3 different scoring functions (GlideXP, FlexX and PLANTS) and for each scoring function the first quartile was calculated. The molecules falling into the best 25% of at least 2 of the 3 scoring functions were chosen for a fingerprint clustering in order to reduce the number of compounds. The remaining compounds were visually inspected for their interactions with the protein and finally 26 molecules were selected. These molecules were purchased and sent to our collaborators at the Rega Institute for Medical Research at KU Leuven (BE) for a first biological evaluation.

The individual steps of the general workflow depicted in Figure 12 are described in detail in the corresponding sections of this chapter.

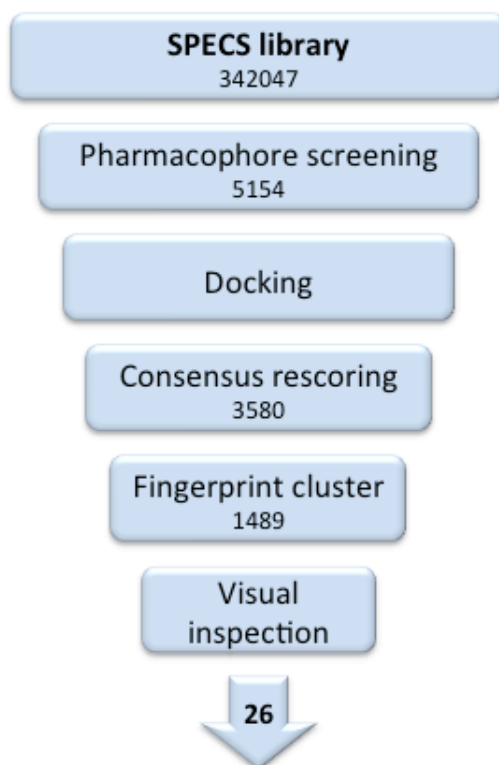


Figure 12: Screening workflow pursued within this project to select compounds for biological evaluation

2.3.6.2 Building the Pharmacophore model for nsP3 macro domain

For the pharmacophore model, the crystal structures 3GPO and 3GPQ of nsP3 macro domain were retrieved from the PDB and prepared for subsequent use using Protonate 3D in MOE (Labute 2008a). They were aligned and superposed. Figure 13 shows the two crystal structures 3GPO and 3GPQ in teal with their ligands ADP-ribose in red and RNA in blue.

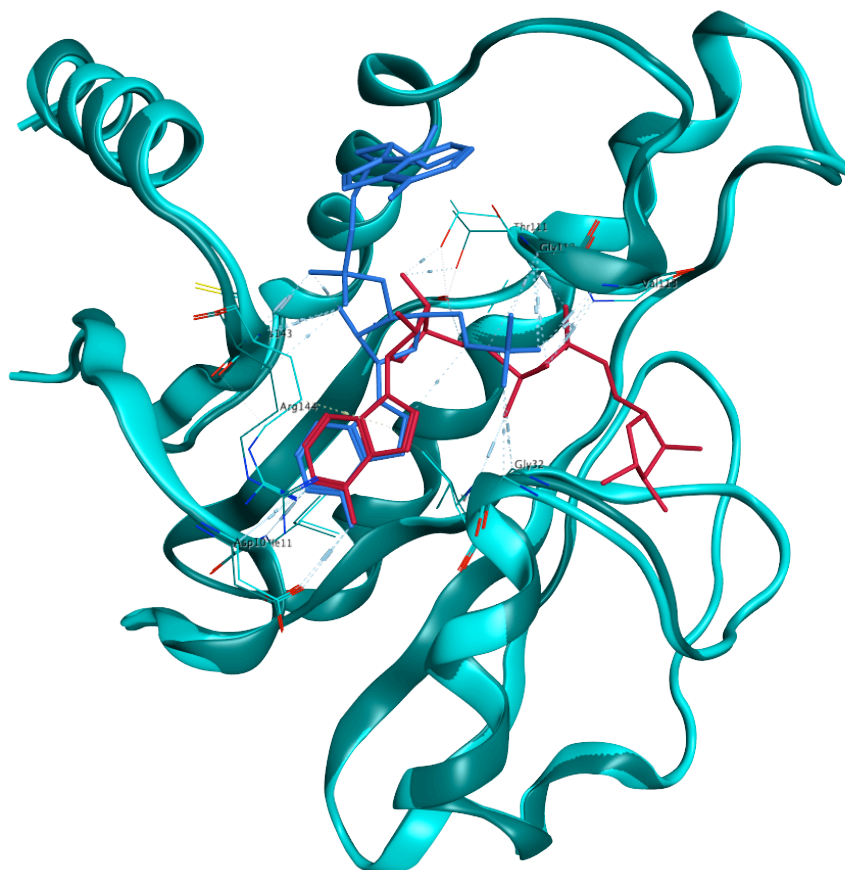


Figure 13: Crystal structures 3GPO and 3GPQ with their ligands ADP-ribose (red) and a small fragment of RNA (blue)

The interactions of the two molecules with their receptor depicted as dashed lines were investigated. Several interactions were reported in the literature in Malet et al. 2009 and could be visualised in MOE. If the two structures were superposed, the position of the adenine overlapped, with the common moiety between the two ligands being an AMP molecule.

In order to create a pharmacophore model that contains the important interactions with the receptor, the main interactions of both ADP-ribose and RNA were used to build the model. The most important interaction reported for all molecules binding within that receptor pocket is the one of the aromatic amino group of adenosine with aspartic acid D10. It is addressed in feature F1 depicted as a purple sphere in the pharmacophore model (Figure 14). Feature F2

represents the backbone interaction of isoleucine I11 with the nitrogen N1 of adenine. It is depicted in the model as blue hydrogen bond acceptor sphere adjacent to the purple sphere of F1. The orange sphere F3 encircles the 5-membered ring of adenine and stands for an aromatic moiety possibly interacting with both the sidechains of arginine R144 and valine V33 via pi stacking. The fourth feature F4 only addresses the part of the pocket occupied by RNA. It is chosen on the basis of a possible interaction of cysteine C143 with the 2' OH-group of the proximal ribose and is represented by a peach coloured hydrogen bond donor/acceptor sphere. The choice of the stronger interaction of the same cysteine with the phosphate group of the RNA-fragment would be too restrictive and its position would be too distant to the rest of the features and was therefore neglected. Feature F5 extends towards the diphosphate pocket of ADP-ribose. The 5' phosphate moiety of the first nucleotide of RNA points towards this position and the phosphate groups engage in several side-chain interactions (Malet et al. 2009). ADP-ribose spans the tunnel-shaped pocket with the diphosphate part of the molecule but for RNA this pocket is too narrow. Therefore a big anionic hydrogen acceptor feature was placed there, represented by the big blue sphere. The sixth feature, a hydrogen bond acceptor group, was positioned in the location of the ring-oxygen of the proximal ribose. Like all acceptor features, it is represented by a blue sphere. Finally, the shape of the pocket was restricted, adding an exclusion volume to the query, preventing molecules matching the query from extending towards areas occupied by receptor atoms.

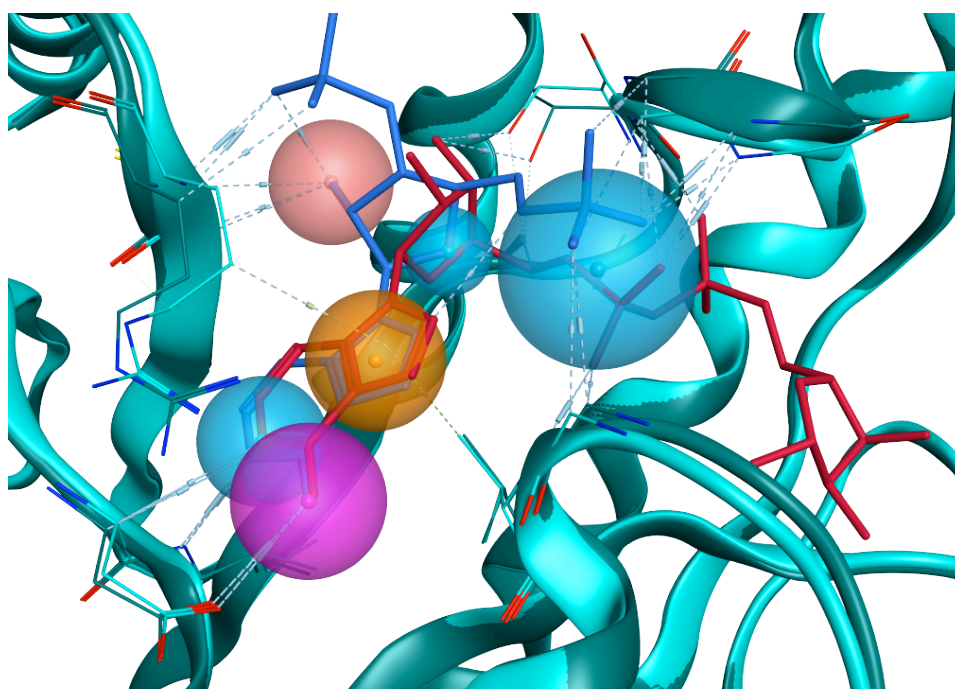


Figure 14: Pharmacophore model for 3GPO and 3GPQ

For the query, the features printed in bold in Table 3 were set as essential. Partial match of the query was enabled but the minimum of features was set to three. The size of the spheres in the picture shows the permitted region for the feature in a matching molecule. The sizes of the spheres in the present query are ranging from 1 Å radius for F6 to 1.4 Å radius for F4. Smaller spheres render the query more restrictive.

Table 3: Features of the pharmacophore model

Features in bold were set essential for the query; Don = hydrogen bond donor feature, Acc = hydrogen bond acceptor feature, Aro = aromatic feature, Ani = anionic feature

	Feature	Interacting amino acid	Molecule	Moiety
F1	Don >0.6	Asp 10	ADP-ribose RNA	Adenine
F2	Acc >0.8	Ile 11	ADP-ribose RNA	Adenine
F3	Aro	Val 33	ADP-ribose RNA	Adenine
F4	Don >0.6 Acc >0.8	Cys 143	RNA	proximal ribose RNA nt1 phosphate
F5	Ani + Acc >0.8	Gly 32 Gly 112	ADP-ribose RNA	PP ADP-ribose RNA nt1 phosphate
F6	Acc >0.8	-	ADP-ribose RNA	Both proximal ribose
V1	Exclusion volume			Receptor atoms

2.3.6.3 Running the Pharmacophore query

The query was subsequently used to screen the SPECS library, which had previously been downloaded and prepared for screening purposes in our laboratory. It contained 342,047 compounds in total and the final query described below yielded 5,154 unique matching molecules. This was a reasonable amount of molecules to continue with, in the docking and scoring step.

2.3.6.4 Docking

In the next step the compounds resulting from the pharmacophore screening were further evaluated for their fit and position within the binding pocket of the target, using a docking approach. First, the 5,154 molecules were prepared with the software Maestro using the Ligand Preparation tool in Schrödinger Release 2016-1 (Schrödinger LLC 2016). For each molecule up to 3 conformations were generated yielding 16,626 conformations in total. Then the protein 3GPO was prepared with the Protein Preparation Wizard in Maestro. Hydrogens were added to the structure and any structural errors detected in the crystal structure of the target protein could be corrected at that point. For docking, solvent molecules such as water were deleted.

When using the docking software Glide, Schrödinger Release 2016-1, Glide (Schrödinger LLC 2016), a grid has to be generated for the specification of the position that the docked molecules are allowed to occupy. For this purpose, the prepared 3GPO protein was used in .pdb format including the ligand ADP-ribose. This structure was chosen because of the distribution of the pharmacophore features that are all falling into the region of the ADP-ribose ligand area. ADP-ribose was kept as the reference ligand to specify the position of the grid. The grid size was set to 12 Å to accommodate docked ligands with a reasonable size.

The compounds were docked with Glide in standard precision mode (Glide SP). Up to five docking poses were generated for each conformation resulting in 76,317 poses.

2.3.6.5 Consensus scoring

The aim of the docking with Glide SP was to position the diverse molecules resulting from the pharmacophore screening within the binding pocket of the macro domain of CHIKV and obtain low energy conformations that fit best into the pocket. In order to assess the quality of the poses, three different scoring functions were used: Glide XP (extra precision mode – score in place), FlexX (BioSolveIT GmbH 2006) and PLANTS (Korb et al. 2009).

Each scoring function was used to assess all the poses generated with Glide SP. Only poses positively evaluated by two or three scoring functions were further considered in the selection process. For each scoring function the first quartile was calculated. Every docking pose receives a value for the calculated binding energy of the molecule within the binding pocket. These numerical values are then listed from the lowest (most negative binding energy) (best pose) to the highest value (worst pose). Then the SIGN function was applied, which gives each pose either a +1, if it falls into the first quartile or a -1 if it doesn't. The SIGN function was applied to all poses of all the three scoring functions. A general equation referred to as the consensus scoring function for a docking pose (X) is shown below and this approach to find a consensus score is referred to as rank-by-vote strategy (Wang & Wang 2002).

$$\Sigma \text{SIGN}(X) = \text{SIGN}(\text{fquartA}-X_A) + \text{SIGN}(\text{fquartB}-X_B) + \text{SIGN}(\text{fquartC}-X_C)$$

Equation 3: Consensus scoring function sum of SIGN value for each scoring function (A, B or C) calculated as the quartile minus the score of the pose X of each scoring function.

For the selection of the compounds only entities with a Σ SIGN of +3 and +1 were chosen. Mathematically +2 and -2 values only occur if the pose scores exactly the value of the quartile of one scoring and is then ranked positively or negatively by two of the other scoring functions.

From the 76,317 poses only 903 poses got a +3 score, which equals to 370 unique molecules. 10,422 poses got a +1 score corresponding to 4,436 unique molecules. Curiously there was one molecule scoring +2. In order not to be too restrictive prior to the visual inspection, structural similarity fingerprint clustering was performed in MOE. The parameter for the similarity was set to 83 and the structural overlap to 55 (empirical parameters previously discovered and used in our laboratory) resulting in 1,489 clusters, which were inspected in the next step.

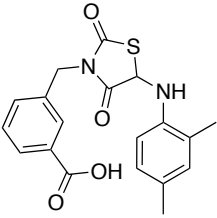
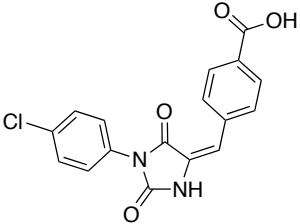
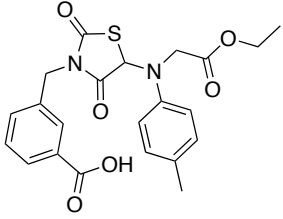
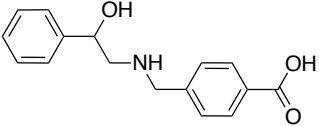
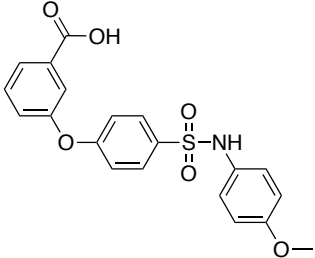
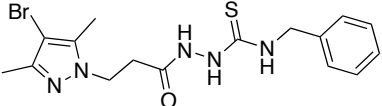
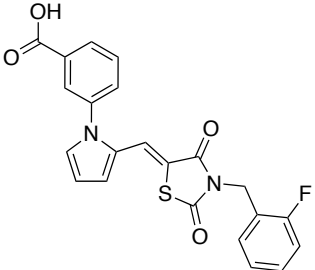
2.3.6.6 Visual inspection

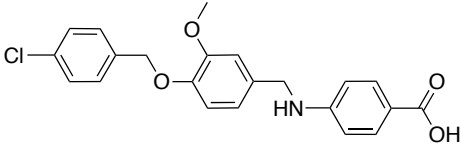
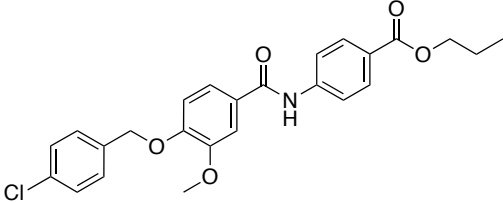
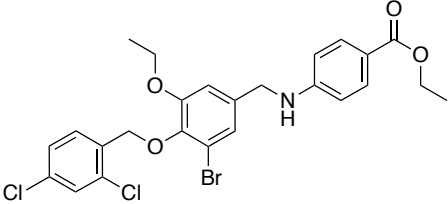
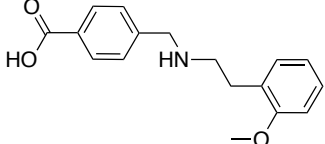
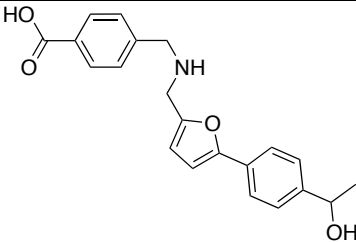
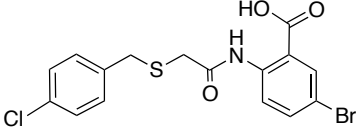
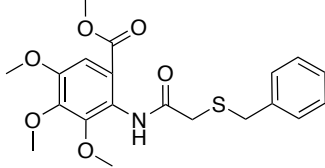
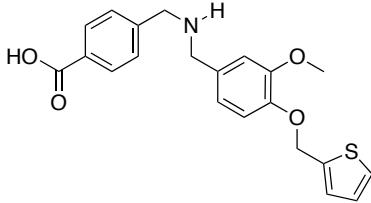
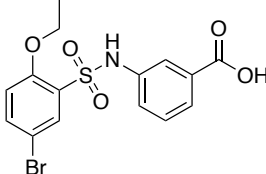
Prior to the visual inspection of the molecules Lipinski parameters (Lipinski et al. 2012) such as molecular weight and SlogP (calculated logarithm of the octanol/water partition coefficient including implicit hydrogens) (Wildman & Crippen 1999) were calculated. Then the 1,489 fingerprint clusters were inspected and a first selection resulted in 142 molecules. In iterative re-examinations the number of compounds was finally reduced to 25.

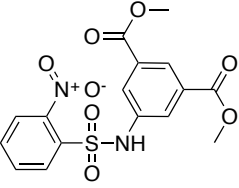
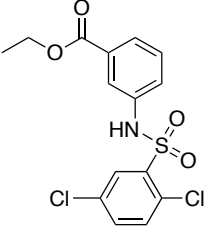
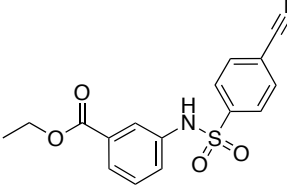
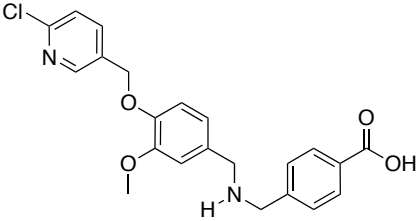
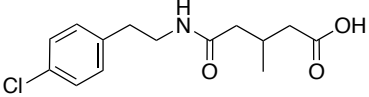
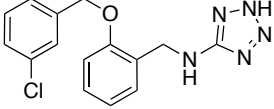
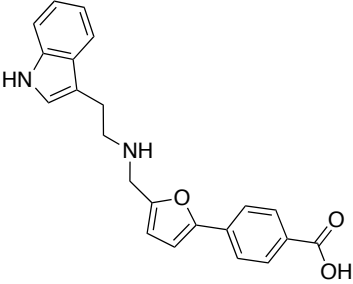
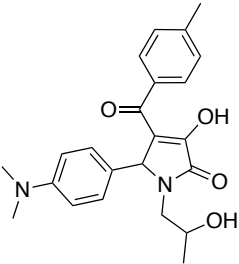
The commercial availability of the compounds was checked and three of the compounds were not available from the provider SPECS (www.specs.net). Two of the compounds were already present in our laboratory from previous studies and had already been tested against CHIKV resulting to be inactive in the cell-based assays (internal confidentiality).

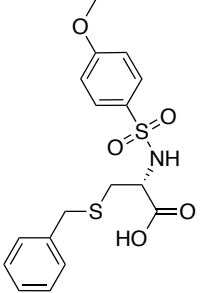
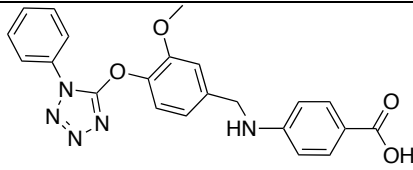
Many of the structures contain a free carboxylic acid group and in some cases esters were available. For some of the molecules the esters were additionally purchased leading to a final number of 26 compounds. The structures of the compounds can be found in Table 4.

Table 4: Final compound selection purchased from SPECS

Nr	Structure	SPECS code
1		AP-893/40972491
2		AH-487/42146104
3		AP-893/40872500
4		AN-465/43411239
5		AQ-390/43238255
6		AK-968/41924648
7		AN-648/41664994

Nr	Structure	SPECS code
8		AN-465/41988493
9		AM-879/15041349
10		AN-465/40934203 AB-1244
11		AN-465/43410994
12		AN-465/43421895
13		AG-205/12680036
14		AO-080/43378337
15		AN-465/43029020
16		AP-263/43371355

Nr	Structure	SPECS code
17		AG-690/11632212
18		AN-652/41151310
19		AO-854/43462375
20		AN-465/42896751
21		AS-662/43412827
22		AN-465/43421736
23		AN-465/43411318
24		AQ-149/43285043

Nr	Structure	SPECS code
25		AP-906/42853565
26		AN-465/43384104

2.3.6.7 Biological evaluation of the selected compounds

The compounds selected through the computational selection process and especially in the last step by visual inspection were sent to our collaborators at the Rega Institute for Medical Research, KU Leuven, Belgium. In a first screening the compounds were evaluated in the cytopathic effect (CPE) reduction assay to assess their general antiviral activity.

The principle of the CPE reduction assay relies on the fact that CHIKV is able to induce a cytopathic effect in susceptible cells, apparent as morphological changes of the cells or changes in metabolic rate, eventually leading to cell-death. Two read-outs are commonly used to determine the status of the cells. Microscopic investigation and evaluation uses a scale from 0-5 to describe the severity of the morphological changes observed in the cells. The MTS/PMS method is a colorimetric assay that measures the cell viability. MTS undergoes a reaction in metabolically active cells, the coloured product can be measured at 490nm with a standard plate reader. The assay is also performed on non-infected cells, which then serve as a standard (Promega Corporation 2006).

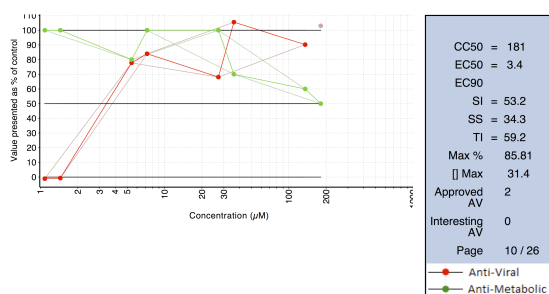
Four compounds showed antiviral activity in the CPE screen. Compound **10** and **22** are able to inhibit the viral replication up to 100% at a non-cytotoxic concentration with an EC_{50} of 3.4 μ M and 51.4 μ M, respectively. Compound **19** inhibits the viral replication up to 60% with an EC_{50} of 54.3 μ M and compound **26** up to 80% with an EC_{50} of 23.4 μ M at non-cytotoxic concentrations. The inhibition profiles of the compounds are shown in the image below.

Compound **22** shows a questionable cytotoxicity profile at a concentration just above the maximum inhibitory concentration and therefore compound **10** was put forward as the best hit compound. A batch of the compound was synthesized in our laboratory to confirm the identity with the compound from SPECS and then resent for further tests.

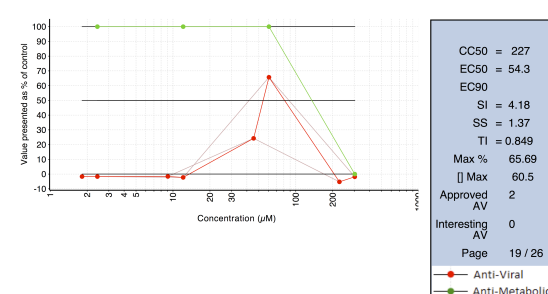
Compound	Virus	Cell	AV Method	AM Method
Primary code	Species	Type	Type	Type
AB_12 44	Chikungunya virus	No Type	899	Vero
				A
			Absorbance	MTS - 498nm
				Microscopy
				Toxicity scoring

Compound	Virus	Cell	AV Method	AM Method
Primary code	Species	Type	Type	Type
AB_12 53	Chikungunya virus	No Type	899	Vero
				A
			Absorbance	MTS - 498nm
				Microscopy
				Toxicity scoring

Needs more data.



[Export chart data to CSV](#)



[Export chart data to CSV](#)

Summary values			
Statistic	CC50	EC50	EC90
Median	= 181	= 3.43	= 72.5
Med.Abs.Dev.		= 0.356	= 61.2
Mean	= 181	= 3.43	= 72.5
Stdev.		= 0.504	= 86.5

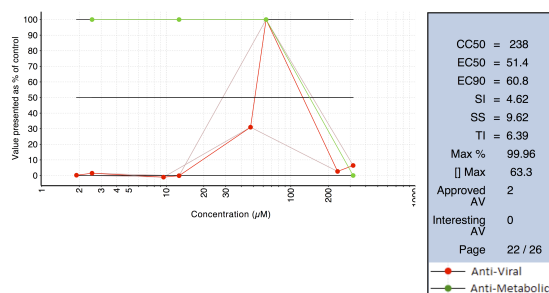
Summary values			
Statistic	CC50	EC50	EC90
Median	= 112	= 41.8	
Med.Abs.Dev.		= 23.7	
Mean	= 112	= 41.8	
Stdev.		= 33.5	

Figure 15 The plots show the antiviral activity and the cytotoxicity in the cells of compounds 10 and 19

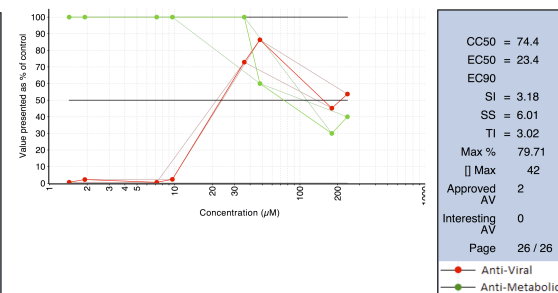
Compound	Virus	Cell	AV Method	AM Method
Primary code	Species	Type	Type	Type
AB_12 56	Chikungunya virus	No Type	899	Vero
				A
			Absorbance	MTS - 498nm
				Microscopy
				Toxicity scoring

Needs more data.

Compound	Virus	Cell	AV Method	AM Method
Primary code	Species	Type	Type	Type
AB_12 60	Chikungunya virus	No Type	899	Vero
				A
			Absorbance	MTS - 498nm
				Microscopy
				Toxicity scoring



[Export chart data to CSV](#)



[Export chart data to CSV](#)

Summary values			
Statistic	CC50	EC50	EC90
Median	= 146	= 28.4	= 54
Med.Abs.Dev.		= 4	
Mean	= 146	= 28.4	= 54
Stdev.		= 5.66	

Summary values			
Statistic	CC50	EC50	EC90
Median	= 113	= 22.7	
Med.Abs.Dev.		= 6.31	
Mean	= 119	= 22.7	
Stdev.		= 15.1	

Figure 16 The plots show the antiviral activity and the cytotoxicity in the cells of compounds 22 and 26

2.3.6.8 Synthesis and biological evaluation of 10 and its derivatives

Compound **10** was re-synthesized in a greater amount and a small library of derivatives was designed and synthesized in order to draw conclusions on the importance of the functional groups of the hit molecule and the possible impact of varied substituents on the activity profile of this series of compounds. For the first round of biological evaluation, **10** and fourteen other compounds were prepared and shipped. Biological activity of compounds synthesised in round 1 are listed in Table 9. Considering the results of this round of analogues, some new molecules were synthesised harbouring combinations of features that seemed to contribute to the antiviral effect of the compounds in round 1. Synthesis of these compounds is found in section 2.3.6.8.2. The third round of design and synthesis was conceptualised for the improvement of

physicochemical parameters especially the solubility of the compounds and is described in section 2.3.6.8.3. Together with some missing compounds of the initial design, the molecules designed in round 2 and 3 were sent for testing. The data for their biological evaluation can be found in Table 10, Table 13 and Table 18.

2.3.6.8.1 First modifications

The chemical structure of compound **10** can be dissected into three parts also referred to as building blocks or rings A, B and C as depicted in Figure 17.

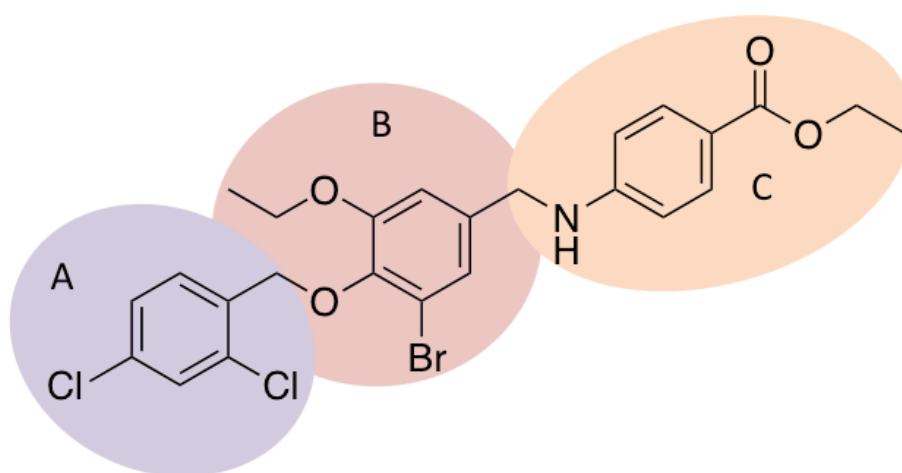


Figure 17: Compound **10** and its building blocks for synthesis

First, compound **10** was synthesised to not only to obtain a stock of the compound for future testing but also to evaluate the feasibility and the conditions of synthesis for analogues of the hit compound. The general synthetic strategy for **10** and its analogues is a two-step procedure schematically depicted in Figure 18 and Figure 19.

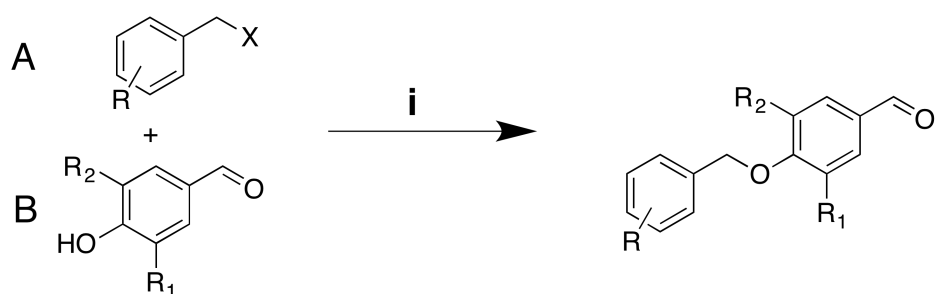


Figure 18: General reaction scheme for the nucleophilic substitution (Zhang et al. 2015)

i) *p*-Hydroxy benzaldehyde derivative (1 eq.) and K₂CO₃ (1.5 eq.) in dry DMF, 30', 25°C; then substituted benzyl halide (X= -Cl, or-Br) (1 eq.) 25°C-40°C, 4-8h.

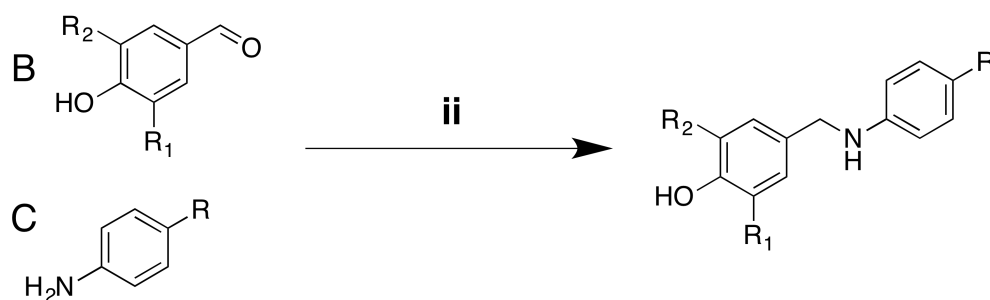


Figure 19: General scheme for the reductive amination (Abdel-Magid et al. 1996)

ii) *p*-Hydroxy benzaldehyde derivative (1 eq.), substituted aniline (1 eq.) and CH₃COOH (1.5 eq.) in DCE, 12h, 25°C then NaBH(OAc)₃ (1-2 eq.), 25°C, 4-8h.

To link building block A with building block B a nucleophilic substitution was performed and to join B with C a reductive amination was applied. For the second step the same procedures were used, but with the intermediates as starting materials, in order to link A to BC or AB to C. For compounds with modifications in ring B, the molecule needed to be synthesised entirely. The route A+B and then +C proved to be best, as the intermediate AB can be obtained pure and as a solid by precipitation in water in most of the cases.

10 was synthesised first linking A+B and then +C as depicted in Figure 20.

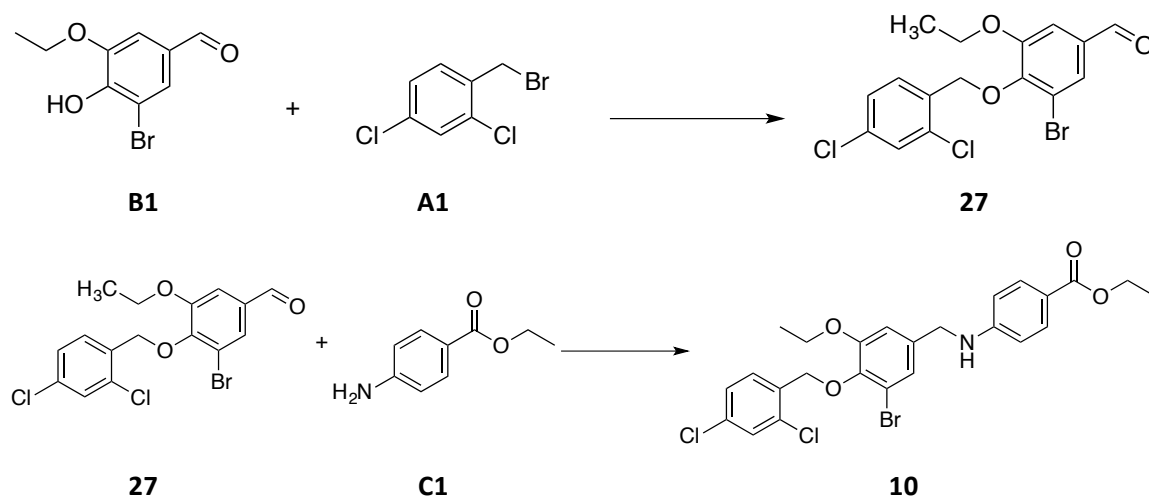


Figure 20: Synthesis of intermediate 27 and compound 10

Modifications on ring A were synthesised using the common intermediate (**28**) consisting of rings B and C, which remain unchanged at this stage. To compound **28** the ring A was added via nucleophilic substitution using the reagents depicted in the table below.

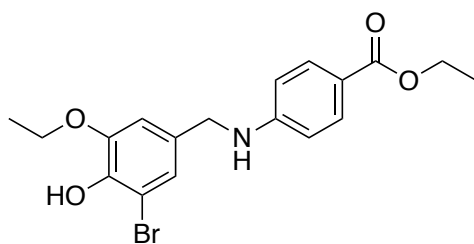


Figure 21: Compound 28

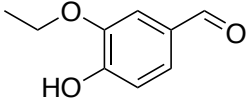
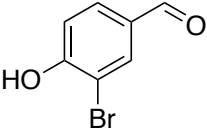
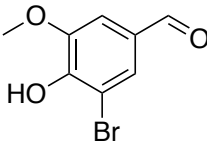
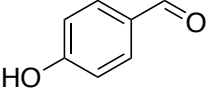
Table 5: Modifications of ring A

Compound	Reagent		Eq. A	Eq. BC	Eq. K ₂ CO ₃	Yield %
29	A2		1	1	1.5	29.5
30	A3		1	1	1.5	43.5
31	A4		1	1	1.5	63.3
32	A5		1	1	1.5	39.9
33	A6		1	1	1.5	12.7
34	A7		1	1	1.5	54

The modifications in building block A were introduced to understand the importance of the substituents in this ring, focussing on nature and position of the substitutions introduced. The two Chloro-groups in ring A contribute significantly to the molecular weight of the compound. **29** was synthesised to test whether any substitution is needed in this ring or if the lipophilicity of the benzene ring was sufficient in order to exhibit antiviral effect in presence of the other substitutions on ring B and C. Compounds **31**, **32** and **33** were testing the optimum position of a mono-substitution on ring A. In **34** a methyl-group was used instead of the chloro-substituent in *para* reducing the molecular weight and increasing the electron density in the aromatic ring. For the nitrile-group in **30**, reduced logP and improved interaction potential was reported in literature investigating nitrile-groups in drugs and drug candidates. Hence, this substitution was investigated as an alternative to the chloro-group and as a substituent with more interaction potential (Fleming et al. 2011).

For the modifications on ring B the molecules had to be synthesised completely and each of them required a different intermediate. The starting materials for building block B are depicted in the table below. Both the equivalents and yields for the nucleophilic substitution to obtain the intermediates are reported in Table 6. The equivalents and yields for the reductive aminations resulting in the final products are reported in Table 7.

Table 6: Synthesis of the intermediates with modified ring B

Compound	SM ring B	Eq. A	Eq. B	Eq. K ₂ CO ₃	Yield %
35	B2 	1	1	1.5	100
36	B3 	1	1	1.5	66
37	B4 	1	1	1.5	92
38	B5 	1	1	1.5	80

The compounds were obtained by precipitation in water. They were washed and dried and used for the next step without further purification.

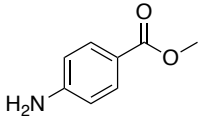
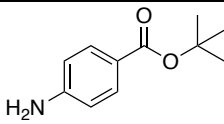
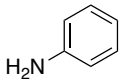
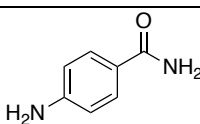
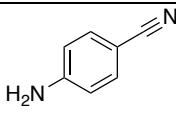
Table 7: Compounds with modified ring B

Compound	Eq. AB	Eq. C1	CH ₃ COOH	NaBH(OAc) ₃	Yield %
39	1 35	1	1	2	33
40	1 36	1	1.5	2	52
41	1 37	1	2	2	46
42	1 38	1	1.5	2	66

The modifications in ring B were made to see if the bromine is really necessary and if the ethoxy-group has an influence on the antiviral activity. Therefore, the compound with the ethoxy-group only (**39**), the bromine-group only (**40**), the methyl-ether of the original hit (**41**), and the unsubstituted ring B (**42**) was synthesised.

Modifications on ring C were synthesised using the intermediate **27** consisting of rings A and B and adding the ring C by reductive amination. In the table below the compounds synthesised via this pathway are depicted including the equivalents of the reagents used in the reaction and reporting the respective yields.

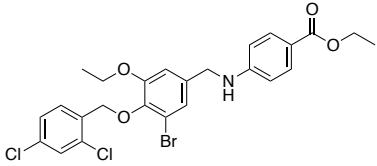
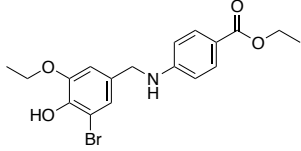
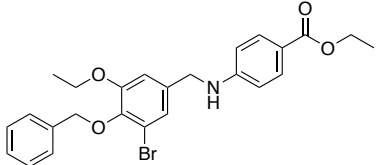
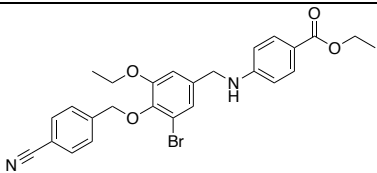
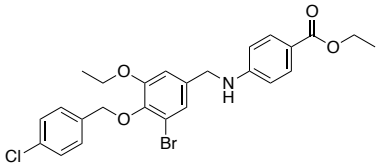
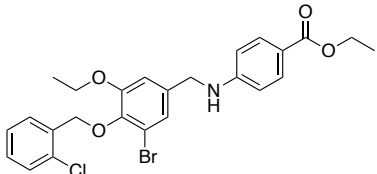
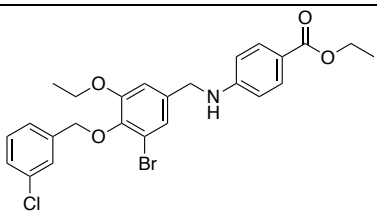
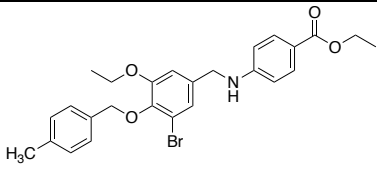
Table 8: Modifications of ring C

Compound	Reagent	Eq. AB	Eq. C	CH ₂ COOH	NaBH(OAc) ₃	Yield %
43	C2 	1	1	1	2	45
44	C3 	1	1	1	2	65
45	C4 	1	1	1	2	63
46	C5 	1	1	1	2	7.6
47	C6 	1	1	1	2	57

In ring C the different modifications were first, esters of different size on the carboxyl-group (**43**, **44**), then a compound without substitutions on the aniline ring was synthesised (**45**). The two other modifications on ring C were the amide substitution (**46**) instead of the carboxylate and the nitrile substituted aniline in *para*-position (**47**). Compound **48** has a free carboxylic acid function on ring C in *para*-position and was obtained by hydrolysis of the ester of **10**. The procedure is described in the experimental part.

The resulting compounds were a first set of analogues to assess the different contributions of the different substituents to the antiviral activity. A classical structure-activity relationship (SAR) study is usually conducted on a purified protein, but only the cell-based assay was available at that stage of the project. It is important to keep in mind that factors like lipophilicity might change the permeation of the compound into the cells and therefore antiviral activity cannot directly be compared, and too strong conclusions about the importance of the substituent should not be drawn.

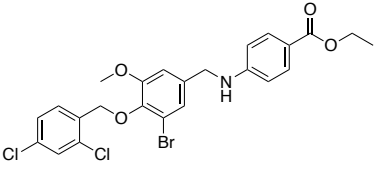
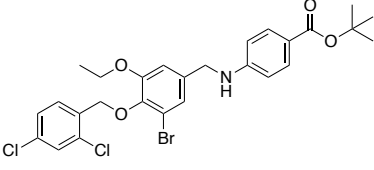
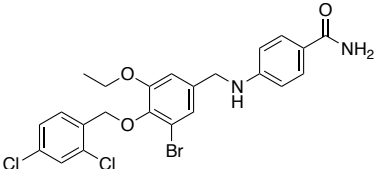
Table 9: Biological evaluation of synthesised compounds

Compound	Structure	Assay	EC ₅₀ (μM)	CC ₅₀ (μM)
10		1	3.3	200
28		1	66	69
29		1	inactive	
30		1	1.65	34.4
		2	19	35
31		1	inactive	
32		1	inactive	
33		1	13	53
34		1	inactive	

Compound	Structure	Assay	EC ₅₀ (μM)	CC ₅₀ (μM)
39		1	2.67	20.5
		2	24 +/- 30	21
40		1	<1.2	13
		2	inactive	
42		1	7.6	99
43		1	toxic	3.1
45		1	1.7	18
		2	Data were not reliable	
47		1	1.72	17.3
		2	Toxic	3.4
48		1	>3.8	3.8
		2	3.6	6.6

Compounds **41**, **44** and **46** were of the initial series but the synthesis was not completed for the first shipment and therefore tested with the second compound shipment.

Table 10: Series 1 of compound 10 analogues tested in shipment 2

Series	Compound	Structure	Assay	EC ₅₀ (μ M)	CC ₅₀ (μ M)
2	41		1	6.1497	>60
			2	>15	>50
2	44		1	>20	>60
2	46		1	2.3978	49.886

2.3.6.8.2 Second round of modifications

The second round of modifications was designed after the first results from the biological evaluation were obtained. Compounds **30**, **39**, **40**, **42** and **45** seemed promising and the analogues **52** to **56** were synthesised. The compounds with the modified building blocks, synthesis routes and yields for the respective analogues can be found in Table 11 and Table 12. Building block A for this series of compounds was always **A3** and both ring B and C were modified.

Table 11: Nucleophilic substitution for intermediates 49-51 for compounds 52-56

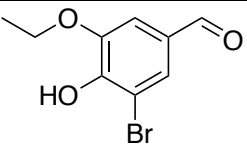
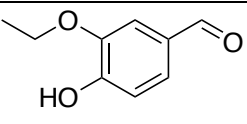
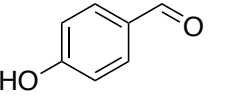
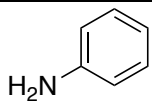
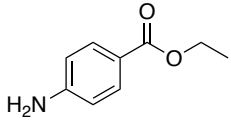
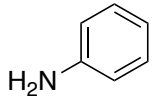
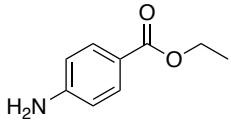
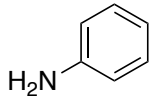
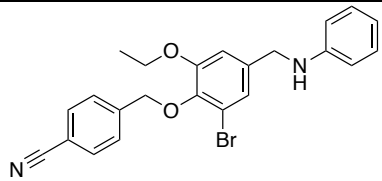
Compound	SM ring B		Eq. A	Eq. B	Eq. K ₂ CO ₃	Yield %
49	B1		1	1	1.5	48
50	B2		1	1	1.5	82
51	B5		1	1	1.5	78

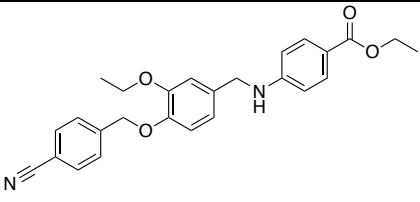
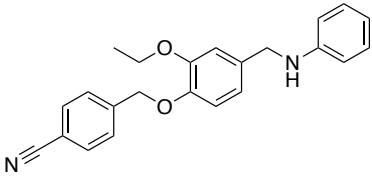
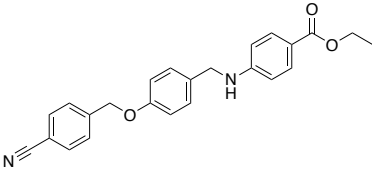
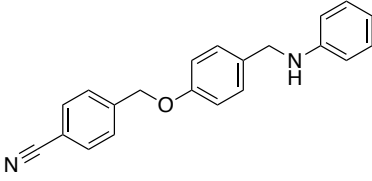
Table 12 Reductive amination for compounds 52-56

Compound	Reagent	Eq. AB	Eq. C	Eq. CH ₂ COOH	Eq. NaBH(OAc) ₃	Yield %
52	C4 	1 49	1	1.5	2	37.3
53	C1 	1 50	1	1.5	2	25.5
54	C4 	1 50	1	1.5	2	60.3
55	C1 	1 51	1	1.5	2	55.0
56	C4 	1 51	1	1.5	2	73.4

The biological evaluation of compounds **52-56** is listed in Table 13. As it is represented in Table 9 and Table 10 a second assay was performed by our collaborators at the Rega Institute of Medical Research due to problems with the cell-culture medium. The data of the first assay on which the modifications of series 2 was based are mostly not valid. After a second biological evaluation of these compounds the initially promising analogues did not show the same activity anymore and mostly had a narrow window of antiviral effect before they became cytotoxic. Furthermore, many of the tested compounds showed crystallisation in the cell-culture medium, so that the reported concentrations, especially for cytotoxicity are likely not exact. Nevertheless, **52** showed antiviral activity in the same range as **10**. The data are listed in the table below.

Table 13: Biological activity of derivatives of series 2

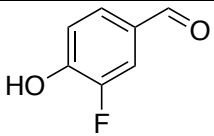
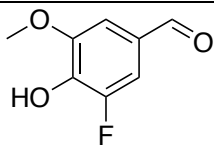
Compound	Structure	Assay	EC ₅₀ (μM)	CC ₅₀ (μM)
52		1	4.46	>60

53		1	NA	>60
54		1	NA	>60
55		1	NA	>60
56		1	NA	>60

2.3.6.8.3 Third round of modifications

The third and last round of analogues of hit **10** was first exploring the different substituents on ring B, as the bromine group might contribute to the solubility problems and an unfavourable logP of the compounds. Substituting it with different smaller halogens like fluoro- or chloro-groups might improve the solubility of the compounds without changing their main properties. Due to inconclusive contributions of the ethyl-ether group on ring B different halogen containing building block that also have a modified second substituent on ring B were purchased. Building block A was 2,4-dichloro-1-(chloromethyl)benzene (**A8**).

Table 14: Synthesis of the intermediates 57-60 for compounds 61-64

Compound	SM ring B	Eq. A	Eq. B	Eq. K ₂ CO ₃	Yield %
57	B6 	1	1	1.5	72.7
58	B7 	1	1	1.5	66.5

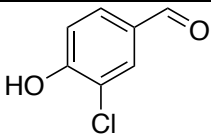
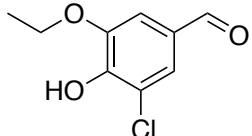
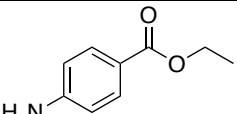
59	B8		1	1	1.5	53.1
60	B9		1	1	1.5	19.2

Table 15: Conditions and yields to obtain compounds 61-64

Compound	Reagent	Eq. AB	Eq. C	Eq. CH₂COOH	Eq. NaBH(OAc)₃	Yield %
61	 C1	1 57	1	1	2	58
62		1 58	1	1	2	12
63		1 59	1	1	2	16
64		1 60	1	1	2	20

Then different heterocyclic building blocks were introduced in rings A and C keeping the substituents of the original hit **10**. Concerning the synthetic procedure, the reductive amination, linking building block C to B or to the intermediate of AB (**27**), had to be adapted as heterocyclic and electron-deficient aromatic amines did not form an imine that could then not be reduced to the amine in the second step of the process. Gutierrez et al. reported the use of Lewis acids such as $\text{TiCl}(\text{O}^i\text{Pr})_3$ to support the imine formation whereas in general the conditions for the reaction did not significantly change (Gutierrez et al. 2005).

Table 16: Compounds 65-67 with heterocycles in building block C

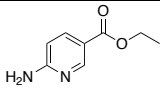
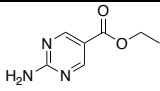
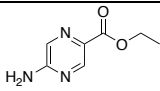
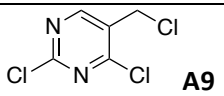
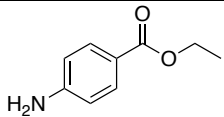
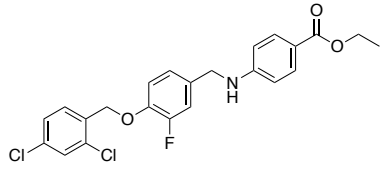
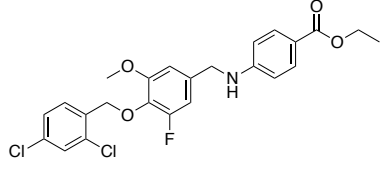
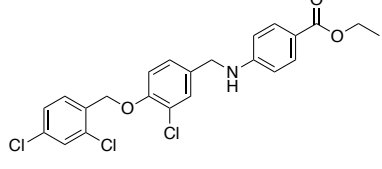
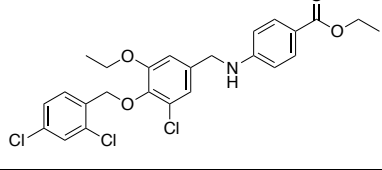
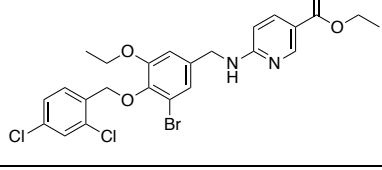
Comp.	Reagent	Eq. 27	Eq. C	Eq. TiCl(OⁱPr)₃	Eq. NaBH(OAc)₃	Eq. CH₃COOH	Yield %
65		1	1.22 C7	2.7	2	3 gtt	3.0
66		1	1.1 C8	2.7	2	3 gtt	3.3
67		1	1.1 C9	2.7	2	3 gtt	2.5

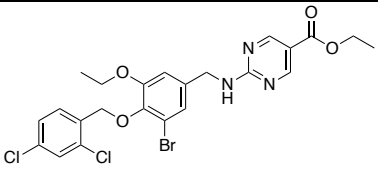
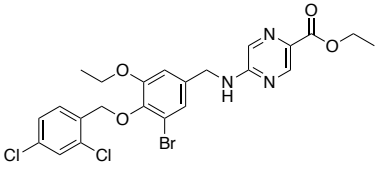
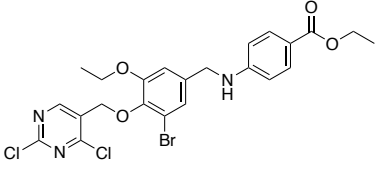
Table 17: Synthesis of 69 with heterocycle in building block A

Compound	SM ring A	Eq. A	Eq. B	Eq. K ₂ CO ₃		Yield %
68	 A9	1	1	1	2	72.0
Compound	SM ring C	Eq. 68	Eq. C	Eq. TFA	Eq. NaBH(OAc) ₃	
69	 C1	1.1	1	2	1.2	6.3

Below the compounds of round three of synthesis and the data from the biological evaluation are listed. Compounds were tested in the CPE-reduction assay as the compounds selected from the virtual screening.

Table 18: Biological evaluation of the compounds from the third round of modifications

Compound	Structure	Assay	EC ₅₀ (μM)	CC ₅₀ (μM)
61		1	5.7938	>20
62		1	2.4482	18.346
63		1	4.6029	>6.666
64		1	4.144	>60
65		1	3.8357	>20

66		1	>20	57.299
67		1	4.3608	>60
69		1	16.953	18.972

Most of the compounds show some antiviral activity in the assay that is in a similar range as the hit compound. None of the compounds shows an improved activity versus cytotoxicity range and further optimization would be needed to obtain compounds with both, better antiviral activity and physicochemical parameters. The collaborators from the Rega Institute for Medical Research communicated, that some compounds classified as cytotoxic might obtain this score because there is extensive precipitation of the compounds in the medium, for which there is no separate score when microscopically assessing the cell-viability.

2.3.6.9 Compound 10 – confirmation of a promising hit compound

Compound **10** was sent for testing again at the Rega Institute for Medical Research in Leuven, Belgium. Several different assays were performed in order to confirm the antiviral activity of this compound. The EC₅₀ of the compound was determined with the CPE and the virus yield assay. The time of addition assay and the CHIKV pseudo particles assay might give insight into the mode of action of the compound.

2.3.6.10 Determination of the EC₅₀ of compound 10

2.3.6.10.1 CPE inhibition assay

The CPE inhibition assay, was used for the first evaluation and for the determination of the EC₅₀ of the compound.

The second batch of compound **10** was evaluated again with the CPE reduction assay. The curve shows the percentage of inhibition of the CPE when the compound is added. The EC₅₀ is therefore the concentration of the compound that is capable to protect 50% of the cells from virus-induced cytopathic effects. The EC₅₀ value measured in the CPE assay was 5 µM.

The CC_{50} was evaluated previously and is above 200 μM , which is sufficiently high to consider the compound as safe and to be sure that the antiviral activity does not result from a cytotoxic effect of the compound.

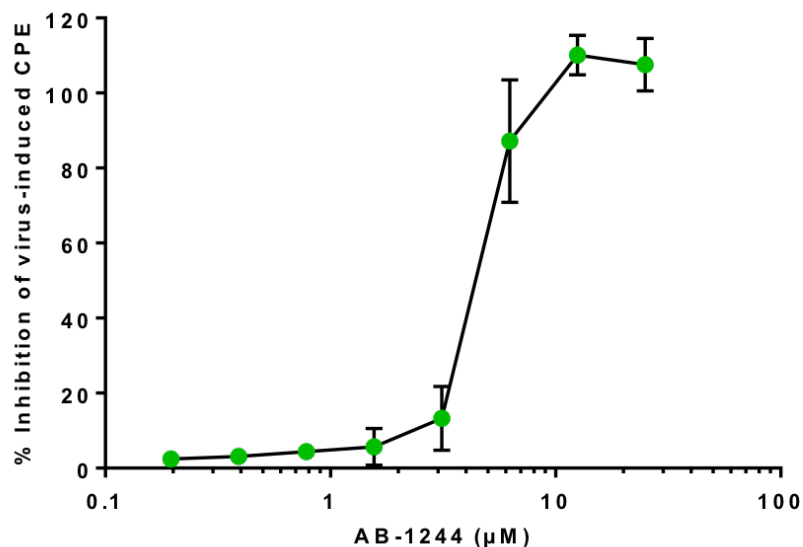


Figure 22: Concentration dependent antiviral activity of 10 (AB-1244) as % inhibition of the virus induced cytopathic effects

2.3.6.10.2 Virus yield assay

The amount of virus was then determined in two different ways with the virus yield assay. After the incubation of the cells with the compounds and the infection with the virus the viral RNA in the supernatant and the amount of infectious virus was determined by dilution and endpoint titration. The corresponding graphs are depicted below.

The EC_{50} values determined in the two different measurements correspond very well and are in the same range as the ones found in the first test and the ones determined in the CPE reduction assay. EC_{50} for the determination of viral RNA compared to untreated cells is 1.4 μM and the amount of infectious virus is reduced by 50% at a concentration of 1.3 μM compared to the untreated infected cells.

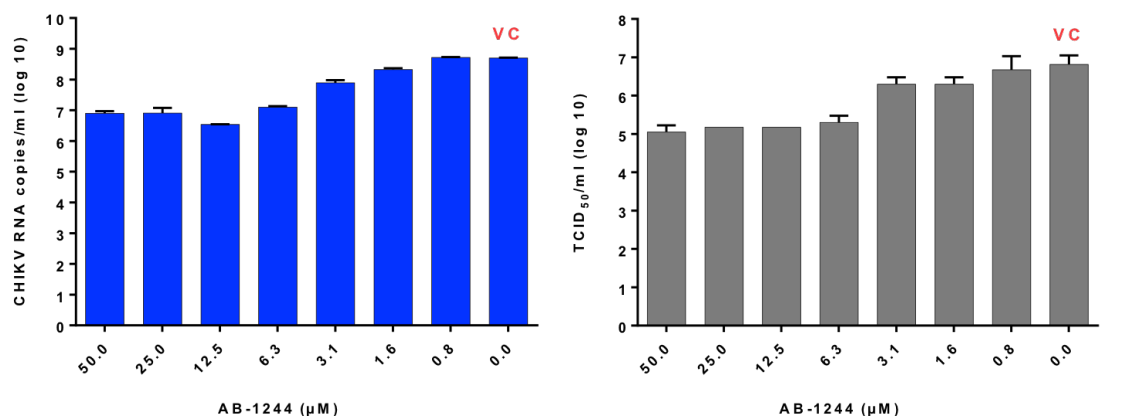


Figure 23: Results of the virus yield assay (quantified viral RNA in blue and infectious virus progeny in grey)

Table 19: EC₅₀ summary of the assays performed for compound 10

Virus	CC ₅₀ (μM)	EC ₅₀ (μM)		
		CPE	Viral RNA	Infectious virus
CHIKV899	>200	5 ± 1	1.4 ± 0.1	1.3 ± 0.12

Compound **10** has a dose-dependent inhibitory effect on viral RNA replication and infectious virus yield (maximum inhibition ≈2 log, EC₅₀ 1.4 and 1.3 μM, respectively).

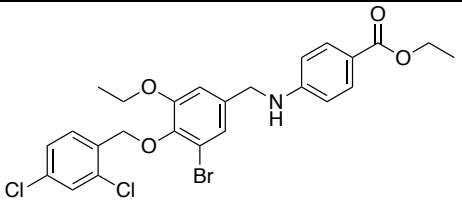
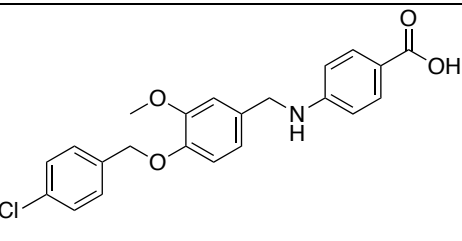
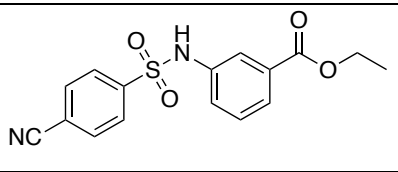
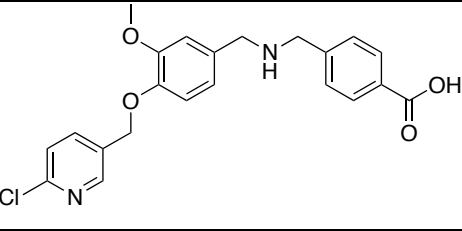
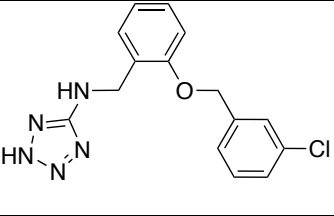
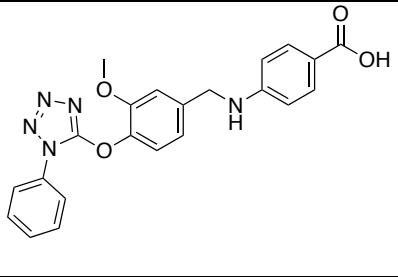
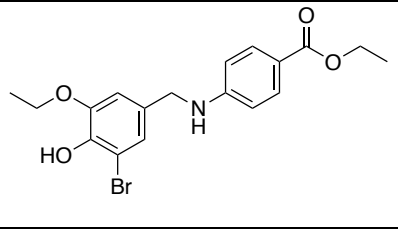
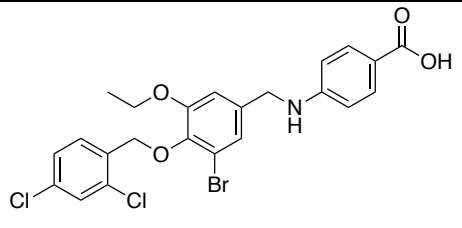
2.3.6.11 Assays on the purified macro domain of CHIKV

2.3.6.11.1 Thermal Shift Assay

The Thermal Shift Assay (TSA) is a technique to measure the stability of a protein in presence or absence of a binding ligand. The protein is incubated with a fluorescent dye that strongly binds to the hydrophobic regions of the protein that are usually not solvent-exposed. Then the protein, with or without a potential binder, is subjected to increasing temperature and eventually undergoes “melting” or denaturation. Thereby, it gets disordered and the fluorescent dye gets exposed. The measured relative fluorescence units are increasing steadily before reaching a peak and drop, when the protein agglomerates. A standard curve for the protein is determined that shows the temperature at the maximum fluorescence. If a ligand binds strongly to the protein this time point is shifted to higher degrees and the protein unfolds slightly later than it would without a ligand. The shift in melting temperature therefore indicates the binding of a potential ligand.

Some of the initial screening compounds (**8**, **19**, **20**, **22**, **26**) and the best hit compound **10** together with selected analogues from the synthesis (**28**, **48**) were assessed for their binding to the purified macro domain of CHIKV (Table 20) by Thermal Shift Assay (TSA). Unlike for ADP-ribose, the thermal shift assay did not reveal any stabilisation of the protein in presence of the tested compounds in concentrations up to 600 μ M, suggesting no or low direct binding of the compounds to the protein.

Table 20: Compounds analysed in the thermal shift assay

10	
8	
19	
20	
22	
26	
28	
48	

2.3.6.11.2 Co-crystallisation

The co-crystallisation was attempted only on **10**. The molecule contains one bromine and two chlorine atoms in the structure and should therefore be detected by X-Ray diffraction at 0.92 Å wavelength, a strategy that can be used to select brominated fragments for drug design (Tiefenbrunn et al. 2014). No anomalous signal for the bromine could be detected in the density map obtained by X-Ray diffraction, suggesting the absence of the compound within the crystals. In conclusion, the compounds were not present in the crystal. Together with the TSA results, no evidence suggests a binding to the purified macro domain and the antiviral activity of **10** could thus result from a different mode of action.

2.3.6.12 Cell-based mode of action studies for **10**

2.3.6.12.1 Time of addition assay

In the time of addition assay the compound was added at different time points respective to the infection with the virus, which is time point 0. When the cells are pre-treated with the compound at -2h (two hours before the infection with the virus), or treated at the same time of the infection, the inhibitory effect on the virus is highest. That suggests that **10** acts most likely at an early stage of the viral infection. This effect has been tested at a concentration of 50 µM of the compound. The effect is clearly visible but is not as strong as the one of the entry inhibitor chloroquine, which served as a control. The second control was T-705 (favipiravir) which is replication inhibitor.

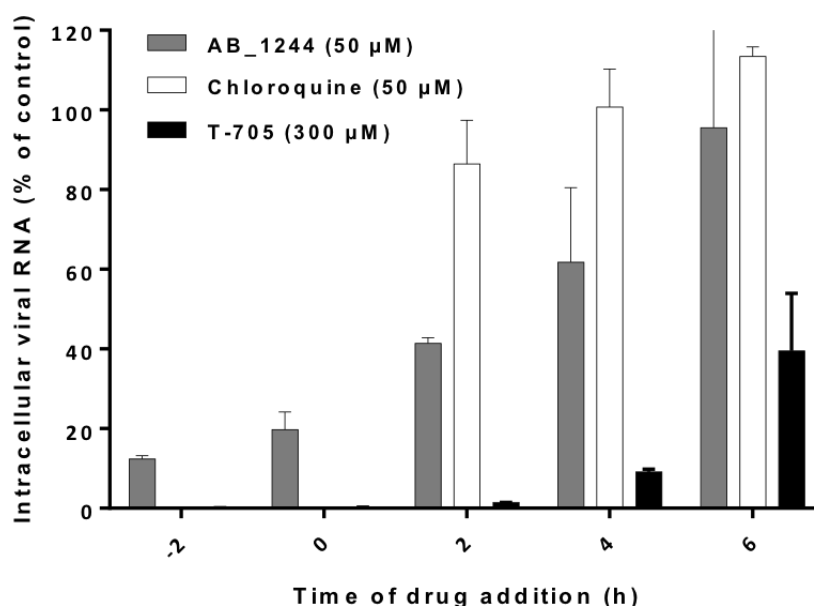


Figure 24: Time of addition of the antiviral compounds **10 (AB-1244) in grey, Chloroquine in white and T-705 (favipiravir) in black**

2.3.6.12.2 Chikungunya virus pseudo particle entry assay

Another way to test if the compound acts as an entry inhibitor or after the entry step is to test the entry of CHIKV pseudo particles (CHIKVpp) into BGM cells. CHIKVpp contain the envelope of CHIKV and the core of a retrovirus coding for a luciferase signal. If the particles are incorporated into the cells and the retroviral RNA is released, the luciferase gene can be integrated into the host cell genome and the cells glow. A compound that interacts with the envelope proteins or a cell surface receptor utilized by the virus will block the incorporation of the virus and can therefore be termed entry inhibitor.

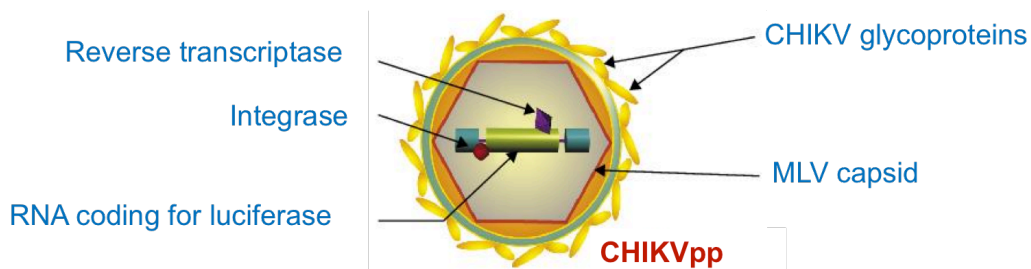


Figure 25: Structure of a CHIKV pseudo particle

10 was tested in comparison with a known entry inhibitor chloroquine in different concentrations.

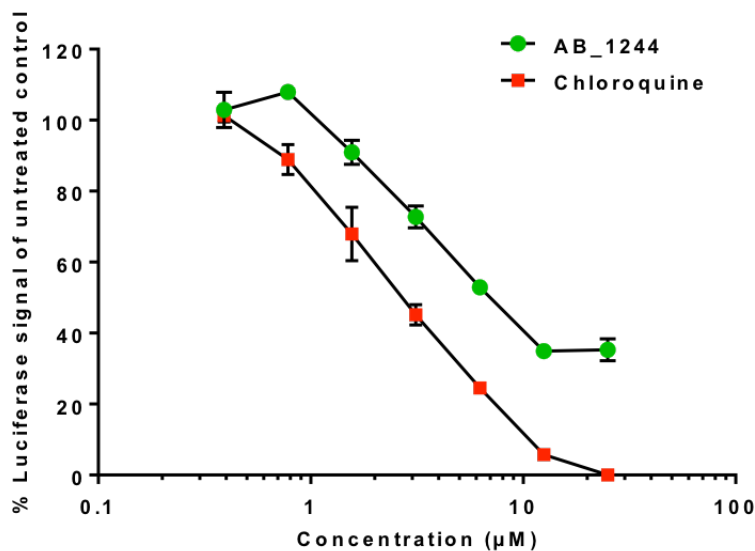


Figure 26: Dose dependent inhibition of the CHIKVpp entry into BGM cells for **10** (green) compared to chloroquine (red)

10 is able to inhibit the entry of CHIKVpp into BGM cells in a dose dependent manner, but less efficiently than chloroquine.

2.3.6.13 Tests against other viruses

After evaluating the activity of **10** on CHIKV, the aim was also to investigate the activity on other related and unrelated viruses. In a personal communication with Bruno Coutard, he offered to contact Gilles Querat, a colleague at Aix-Marseille Université, for the testing of **10** and **48** on a panel of viruses.

The compounds were tested against CHIK OPY strain, O’Nyong Nyong virus (closely related to CHIKV but replicates only in primates), Ross River Virus (a distantly related Old World alphavirus) and VEEV (a New World alphavirus, very distant from CHIKV). Furthermore, the compounds were tested against West Nile Virus (WNV), a flavivirus. The compounds were initially designed as macro domain inhibitors and WNV does not encode a macro domain. Therefore this virus serves as an unrelated negative control.

For adequate assay controls the cytotoxicity of the compounds was addressed first. **10** shows a flat cytotoxicity above 25-50 μM , which is due to a very limited solubility in the cell culture medium. Indeed plenty of crystals were observed at 200 μM and diminishing with the concentration, but were still visible at 25 μM . An estimate of the maximum solubility of the compound in aqueous medium would be in the range of 20-30 μM . This puts a limit on the CC_{50} measurements because the sufficient concentrations cannot be reached. **48** shows a more classical cytotoxicity curve and whereas cytotoxic concentrations were not precisely determined during the setup of the experiment, it looks as if the CC_{50} might be around 200-300 μM .

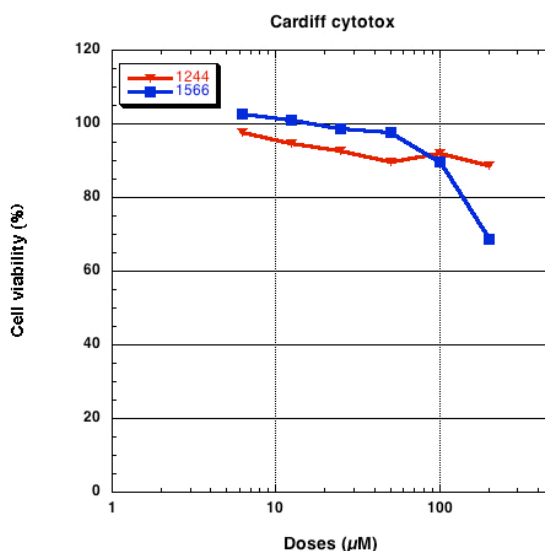


Figure 27: Cytotoxicity curves for **10** (red) and **48** (blue)

Then the antiviral activity of the two compounds was determined. **10** exhibits a low but significant inhibition of virus yield upon addition of the compound, but this inhibition is not specific as it can also be demonstrated with the negative control, WNV, where the inhibition is even higher. The virus yield reduction is not very pronounced and because the CC_{50} and EC_{50}

could not be accurately determined in this screening setup, no clear conclusion can be drawn at that point. The results might be interpreted such as there is low and unspecific inhibition of compound **10** in the present assay, which can be related to impaired metabolism of the cells. A slight reduction in cell viability can be observed at around 40 μM . As for **48**, no virus yield reduction can be observed up to 40 μM , it can be concluded that this compound is inactive in cell culture.

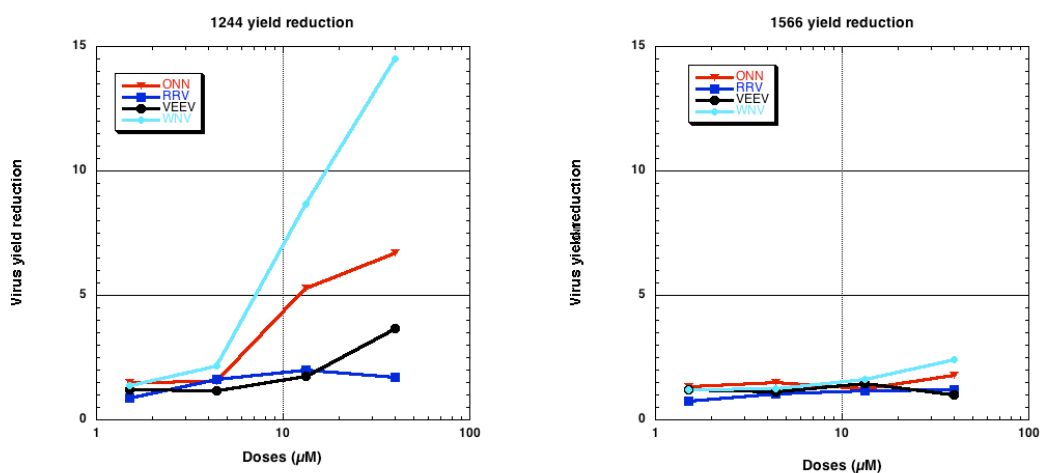


Figure 28: Virus yield assays for 10 and 48

To conclude, the two compounds are non-specific inhibitors of alphaviruses in cell culture in the present assay. However, this does not negate that **10** might be an inhibitor of virus replication, although unspecific. The compound might be blocking a cell factor, which is necessary for virus entry or replication of both alphaviruses and WNV.

The Rega Institute for Medical Research also tested **10** against enterovirus A71 (EV-A71), Zika virus (ZIKV) and yellow fever virus (YFV), but it was not active in any of the screenings.

2.3.7 Conclusions

Computational studies on the CHIKV macro domain have yielded 26 compounds that were sent to our collaborators in Belgium to be tested in a first cell-based assay. Four of the compounds were active against CHIKV but only **10** showed a promising antiviral activity and toxicity profile.

10 was therefore further explored: in a first step the compound was synthesised again in house to gain more material for further testing. Subsequently also the derivatisation of the compound was planned to better understand which of the functional groups of the hit were responsible for its activity. The derivatives were designed via a classical medicinal chemistry approach. Mainly two different pathways were used to synthesize the majority of the compounds in two steps as described before. 29 novel final compounds and 14 intermediates were synthesized.

While the synthesis of the analogues was ongoing, a second batch of **10** was evaluated by our collaborators in greater details. A second CPE reduction assay was performed as well as the virus yield assay with two different results: The determination of the EC₅₀ referring to the viral RNA, and to the number of infectious viral progeny. All three assays result in an EC₅₀ within the low micromolar range. The cytotoxicity of the compound was evaluated in the first screen and lies above the 200 µM, so the compound can be considered as safe in the applied concentration and does not exhibit its antiviral activity via cytotoxic effects.

In order to investigate the mode of action of the compound the time of addition assay was performed. The results show activity mostly at an early stage. Two controls were used: the entry inhibitor chloroquine and the RNA-dependent RNA-polymerase inhibitor T-705 (favipiravir). An early stage inhibitor is either acting on the entry process or on the early events in viral replication. The latter would confirm the hypothesis of **10** to target the CHIKV macro domain.

In addition to the time of addition assay, an entry assay was performed. In the CHIKVpp entry assay the compound was compared to chloroquine a well-characterised entry inhibitor. The results from this assay strongly suggest a role as entry inhibitor for **10**. The compound acts in a dose-dependent manner and yields a result in the low micro molar range.

In the meantime several compounds were tested in a thermal shift assay using purified CHIKV macro domain but unfortunately none of them showed a stabilisation of the protein that can be considered as direct interaction or binding. Subsequently, also co-crystallisation for **10** was attempted, but no extra electron-density could be observed in the well-characterised macro domain of CHIKV. Drawing conclusions from this biochemical evaluation the effect of **10** and the other tested compounds listed in Table 20 cannot be attributed to the direct interaction with CHIKV macro domain.

10 was then tested against a panel of different viruses at the Rega Institute of Medical Research in Belgium and at the Aix-Marseille Université by Gilles Querat. The viruses used were EV71, YFV and ZIKV at the Rega Institute and CHIKV OPY, VEEV, RRV, ONN and WNV by Gilles Querat. Interestingly, some activity could be observed towards WNV, which does not belong to the group of alphaviruses but to the flaviviruses and does not encode a macro domain in its genome. It is therefore intriguing to believe that WNV and CHIKV might use a common host factor that accounts for the activity of the compound in both viruses, but this is so far only a speculation.

It is important to note that **10** was the only compound that was evaluated in such depth and that synthetic optimisation attempts did not yield a more active compound with improved physicochemical parameters. In several assays, **10** precipitated in aqueous medium at higher concentrations. Gilles Querat perhaps best characterised the solubility limit in his assay

conditions. Leen Delang and Rana Abdelnabi in Leuven attempted to raise resistance mutations in CHIKV using **10** but treatment with high concentrations of compound was not possible due to solubility limits. Unfortunately the compounds that did not show any visible precipitation did not show antiviral activity. Even the last round of derivatives harbouring a heterocycle in ring A or ring C did not improve the solubility and were reported to show crystals or precipitation at high concentration in the CPE assays.

In conclusion, this thesis presents the discovery of a new antiviral hit against CHIKV in cell-based assays and a series of analogues thereof. Several steps towards the elucidation of the mode of action of respective hit compound were made and this project has not yet reached a final conclusion, but it can take up to years to identify and characterise the mode of action of a new compound towards a previously unknown target. The toolbox of computational and medicinal chemistry to contribute to these efforts is rather limited to be applied to such a complex and wide-spanning research question. Thus, such an endeavour requires a well-concerted interdisciplinary cooperation of chemists, structural biologists and virologists. The results of this project could serve as a starting point to improve the active hit compounds, as well as it would be worthwhile investigating the mode of action of this new antiviral scaffold in greater depth.

2.3.8 Experimental

2.3.8.1 Chemistry

2.3.8.1.1 Reagents

All reagents were purchased at Sigma Aldrich, Alfa Aesar, Acros, Apollo Scientific, Fluorochem, TCI and used as received. Dry solvents were purchased from Sigma Aldrich.

2.3.8.1.2 Thin layer chromatography - TLC

Silica gel plates (Merck Kieselgel 60F₂₅₄) were used. All plates have been developed by the ascending method. The visualisation of the spots was carried out under UV light at the wavelength of 254 nm after evaporation of the solvents.

2.3.8.1.3 Column Chromatography

Glass columns were packed with Woelm silica (32-63 mm) in the appropriate eluent. The samples were applied in the same eluent as a concentrated solution. Fractions containing the product were identified by TLC, then combined and their solvent was evaporated under reduced pressure.

2.3.8.1.4 Automated flash column chromatography

Column chromatography purifications were carried out automatic flash column chromatography on a Biotage Isolera One. Fractions containing the product were identified by TLC and pooled, and the solvent was removed *in vacuo*.

2.3.8.1.5 UPLC-MS

UPLC-MS analysis was conducted on a Waters UPLC system with both Diode Array detection and Electrospray (+ve and -ve ion) MS detection. The stationary phase was a Waters Acquity UPLC BEH C18 1.7µm 2.1x50mm column. The mobile phase was H₂O containing 0.1% Formic acid (A) and MeCN containing 0.1% Formic acid (B). Column temperature: 40°C. Sample diluent: acetonitrile. Sample concentration 10 µg/mL. Injection volume 2 µL.

Two methods were used:

Linear gradient standard method (A): 90% A (0.1 min), 90%-0% A (2.1 min), 0% A (0.8 min), 90% A (0.1 min); flow rate 0.5 mL/min.

Linear gradient standard method (B): 90% A (0.1 min), 90%-0% A (1.5 min), 0% A (1.4 min), 90% A (0.1 min); flow rate 0.5 mL/min.

2.3.8.1.6 NMR Spectra

^1H , ^{13}C and ^{19}F NMRs were performed on a Bruker AVANCE 500 spectrometer with 500 MHz, 125 MHz and 470 MHz respectively. Chemical shifts (δ) are reported in units of parts per million (ppm) downfield from tetramethylsilane (TMS) $\text{Si}(\text{CH}_3)_4$ for the ^1H proton NMR and relative to the respective solvent's peak for the ^{13}C carbon NMR. Multiplicities are reported as s (singlet), bs (broad singlet), d (doublet), t (triplet), q (quartet), dd (doublet of doublet) td (triplet of doublets) dt (doublet of triplets) or m (multiplet). The ^1H - ^1H coupling constants J are reported in Hertz (Hz).

2.3.8.1.7 Mass Spectra

The mass spectra were recorded on the UPLC-MS system.

For the masses where the ionisation was not sufficient with the UPLC-MS instrument mass spectra were recorded on a Bruker Daltonics microTof-LC system (atmospheric pressure ionisation, electron spray mass spectroscopy) in positive mode.

Masses for the compounds are reported as $[\text{M}+\text{H}]^+$ or $[\text{M}+\text{Na}]^+$, $[\text{M}-\text{H}]^-$ or as [Fragment] m/z .

2.3.8.1.8 General Procedures Synthesis

2.3.8.1.8.1 General procedure 1: Nucleophilic substitution

Potassium carbonate (1.5 eq.) was added to a solution of 1 equivalent of the appropriately substituted benzylhalogenide (A) in dry DMF and stirred for 30 minutes under argon atmosphere. Then the substituted *p*-hydroxybenzaldehyde (B) (1 eq.) was added to the reaction and the mixture was stirred until the reaction reaches completion, which was monitored by TLC. The reaction mixture was poured into water and the precipitate was isolated by vacuum filtration. If no precipitate was formed the reaction mixture was extracted with ethyl acetate or dichloromethane, washed with brine and the organic solvents were evaporated under reduced pressure (Zhang et al. 2015).

2.3.8.1.8.2 General Procedure 2: Reductive amination

One equivalent of building block B (the substituted *p*-hydroxybenzaldehyde containing the connective aldehyde function), 1 equivalent of building block C and 1.5 equivalents of glacial

acetic acid were stirred over night in the appropriate amount of dichloroethane at room temperature. Then 2 equivalents of sodium triacetoxyborohydride (STAB) were added to the reaction mixture and stirred until the reaction reaches completion. The reaction is then quenched with sodium hydrogencarbonate, extracted with dichloromethane and the organic layer is washed with brine and dried over sodium sulphate or magnesium sulphate. The organic solvent is evaporated under reduced pressure and the resulting product is dried under vacuum (Abdel-Magid et al. 1996).

2.3.8.1.8.3 General procedure 3: Reductive amination

The procedure for the reductive amination had to be adapted for electron-deficient anilines and in heterocyclic amines, when the general procedure 2 only led to the reduction of the carbonyl component resulting the corresponding alcohol (Figure 29). An example is shown for the common starting material for all modifications on ring C. The protocol was adapted according the Gutierrez et al. (Gutierrez et al. 2005).

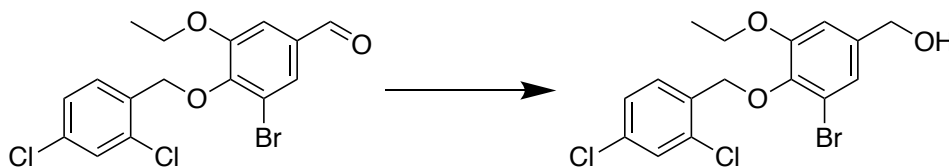


Figure 29: Reduction of aldehyde in general procedure 2 when attempting a coupling with electron-deficient anilines and heterocyclic amines

To prevent this event it is necessary to stabilize the imine formation previous to the reduction. The compounds were synthesized as follows: To a stirring solution of 1 eq. of the aldehyde moiety and 1.1 eq. of the aromatic nitrogen-containing building block C in anhydrous DCM 2.2 eq. of $\text{TiCl}(\text{O}^i\text{Pr})_3$ are added under argon. After 5 minutes stirring at room temperature 5 eq. of $\text{NaBH}(\text{OAc})_3$ are added portionwise. Three drops of glacial acetic acid are added and the mixture is left stirring for 6h. Then the mixture is poured into NaHCO_3 and extracted 4 times with DCM. The organic phase is washed with brine and dried over MgSO_4 . The organic solvent is then evaporated on the rotavapor.

2.3.8.1.9 Purification by automated flash column chromatography

If necessary, purification by column chromatography was performed and fractions containing the product were combined, evaporated under reduced pressure and dried under high vacuum.

2.3.8.2 Synthesis of the compounds

3-Bromo-4-((2,4-dichlorobenzyl)oxy)-5-ethoxybenzaldehyde (27)

Chemical formula: C₁₆H₁₃BrCl₂O₃; MW: 404.08; CAS-Nr: 345980-30-5 (Zhang et al. 2015)

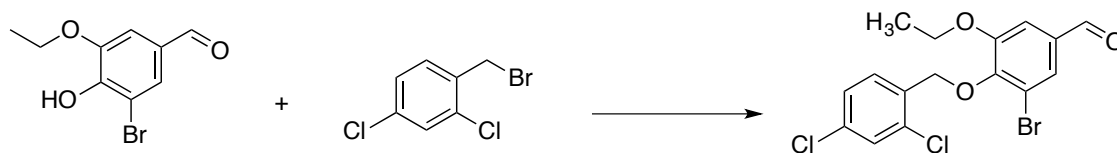


Table 21: Chemicals for the preparation of 27

Name	Eq.	mmol	MW	g	(d)	V
3-bromo-5-ethoxy-4-hydroxybenzaldehyde	1	0.834	245.07	0.204		
2,4-dichlorobenzylbromide	1	0.834	239.92	0.200		
K ₂ CO ₃	1.5	1.251	138.2	0.173		
DMF anhydrous			73.09		944 kg/m ³	2 mL

Procedure: General procedure 1

TLC system: EtOAc:Hexane 20:80 v/v, R_f = 0.7

Purification: Further purification was performed using manual flash column chromatography (EtOAc:Hexane 10:90 v/v stepwise increased to 50:50 v/v, at which the product was eluted.

Outcome: Yield: 249mg, 73,9%, white powder

Analysis:

¹H NMR (500 MHz, CDCl₃) 1.46 (t, *J* = 6.9 Hz, 3H), 4.17 (q, *J* = 6.9 Hz, 2H), 5.28 (s, 2H) 7.34 (dd, *J* = 2.0 Hz, *J* = 8.3 Hz, 1H), 7.42 (d, *J* = 1.6 Hz, 1H), 7.43 (d, *J* = 2.0 Hz, 1H), 7.68 (d, *J* = 1.5 Hz, 1H), 7.75 (d, *J* = 8.3 Hz, 1H) 9.87 ppm (s, 1H)

¹³C NMR (125 MHz, CDCl₃) 14.63 (CH₃), 64.96 (CH₂, aliphatic), 71.13 (CH₂, aliphatic), 110.93 (CH, aromatic), 118.15 (C-Br, aromatic), 127.23, 128.58, 129.05, 130.58 (CH, aromatic), 133.27, 133.35 (C-C, aromatic), 133.47, 134.33 (C-Cl, aromatic), 150.39 (C-O, aromatic), 153.36 (C-O, aromatic), 189.90 (C=O)

Ethyl 4-((3-bromo-4-((2,4-dichlorobenzyl)oxy)-5-ethoxybenzyl)amino)benzoate (10)

Chemical formula: C₂₅H₂₄BrCl₂NO₄; MW: 553.27; CAS-Nr: 445414-28-8

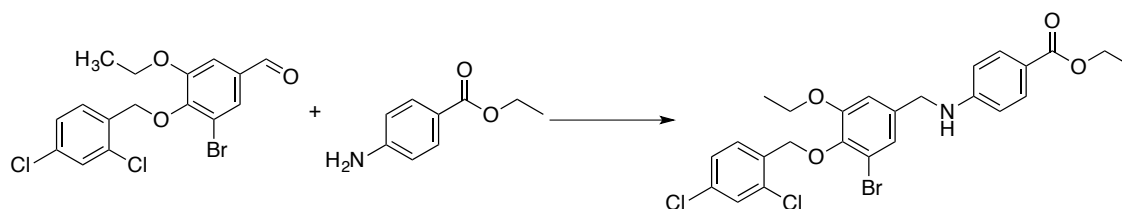


Table 22: Chemicals for the preparation of 10

Name	Eq.	mmol	MW	g	(d) g/cm ³	V (mL)
3-bromo-4-((2,4-dichlorobenzyl)oxy)-5-ethoxybenzaldehyde	1	0.247	404.08	0.113		
Ethy 4-aminobenzoate	1	0.247	165.19	0.0452		
CH ₃ COOH	1	0.247	60.05	0.016	1.049	0.013
DCE			98.95			2 mL
NaBH(OAc) ₃	1.5 + 0.5	0.371	211.94	0.079 +0.03		

Procedure: General procedure 2

Work up: Extraction with DCM

Purification: The residual components were purified with column chromatography using a gradient of 0 to 50% EtOAc in n-hexane (v/v) and the fractions containing the product were identified by TLC in 254 nm UV light. The relevant fractions were evaporated under reduced pressure.

Outcome: Yield: 61 mg, 51%, white solid

Analysis:

¹H NMR (500 MHz, CDCl₃) 1.36 (t, *J* = 7.1 Hz, 3H), 1.38 (t, *J* = 6.9 Hz, 3H), 4.02 (q, *J* = 6.9 Hz, 2H), 4.31 (q, *J* = 7.1 Hz, 2H), 4.32 (s, 2H), 4.51 (s, 1H), 5.11 (s, 2H), 6.58 (dd, *J* = 8.9 Hz, *J* = 2.1 Hz, 2H), 6.83 (d, *J* = 1.9 Hz, 1H), 7.12 (d, *J* = 1.9 Hz, 1H), 7.29 (dd, *J* = 8.3 Hz, *J* = 2.1 Hz, 1H), 7.39 (d, *J* = 2.1 Hz, 1H), 7.75 (d, *J* = 8.3 Hz, 1H), 7.87 ppm (dd, *J* = 8.9 Hz, *J* = 1.9 Hz, 2H)

¹³C NMR (125 MHz, CDCl₃) 14.63, 14.90 (CH₃), 47.28, 60.47, 64.83, 71.02 (CH₂, aliphatic), 111.64, 111.92 (CH, aromatic), 118.20, 119.65 (C-C, aromatic), 123.3, 127.28, 129.03, 130.73, 131.69 (CH, aromatic), 133.35 (C-C, aromatic), 134.13, 134.15 (C-Cl, aromatic), 136.18, 144.57, 153.37 (C-C, aromatic), 166.92 (C=O).

UPLC-MS: Method B

RT: 2.512 min, Purity: 100%

MS: 158.9 [C₇H₅Cl₂]⁺

Ethyl 4-((3-bromo-4-((2,4-dichlorobenzyl)oxy)-5-ethoxybenzyl)amino)benzoate (10)Chemical formula: C₂₅H₂₄BrCl₂NO₄; MW: 553.27**Table 23: Second batch of compound 10**

Name	Eq.	mmol	MW	g	(d) g/cm ³	V (mL)
3-bromo-4-((2,4-dichlorobenzyl)oxy)-5-ethoxybenzaldehyde	1	0.346	404.08	0.140		
Ethy 4-aminobenzoate	1	0.346	165.19	0.0572		
CH ₃ COOH	1	0.346	60.05	0.021	1.049	0.02
DCE			98.95			5 mL
NaBH(OAc) ₃	2	0.692	211.94	0.147		

Procedure: The reaction was repeated on the remaining amount of 45 (0.140 g). The amounts of the other reagents are adapted according to the general procedure 2. As a solvent 5 mL of DCE were used from the beginning and 2 eq. of the reductive agent NaBH(OAc)₃.

Purification: The work up was performed as described for 10. Purification with column chromatography was performed on the automated column system Interchim using an automatically generated method from the TLC measurements. The separation was obtained with a gradient of 0-100% EtOAc in n-hexane (v/v) over 10 column volumes after previous equilibration. The fractions containing the products were identified by TLC and were united and evaporated under reduced pressure.

Outcome: Yield: 84 mg, 44%; white solid

Analysis:

¹H NMR (500 MHz, CDCl₃) 1.37 (dt, *J* = 8.6 Hz, *J* = 7.1 Hz, 6H), 4.02 (q, *J* = 7.0 Hz, 2H), 4.31 (q, *J* = 7.1 Hz, 4H), 4.51 (s, 1H), 5.12 (s, 2H), 6.58 (d, *J* = 8.9 Hz, 2H), 6.84 (d, *J* = 1.9 Hz, 1H), 7.12 (d, *J* = 1.9 Hz, 1H), 7.30 (dd, *J* = 8.3 Hz, *J* = 2.1 Hz, 1H), 7.39 (d, *J* = 2.1 Hz, 1H), 7.75 (d, *J* = 8.4 Hz, 1H), 7.88 ppm (d, *J* = 8.9 Hz, 2H)

Analyses for ¹³C NMR, UPLC and MS were performed together with the first batch.

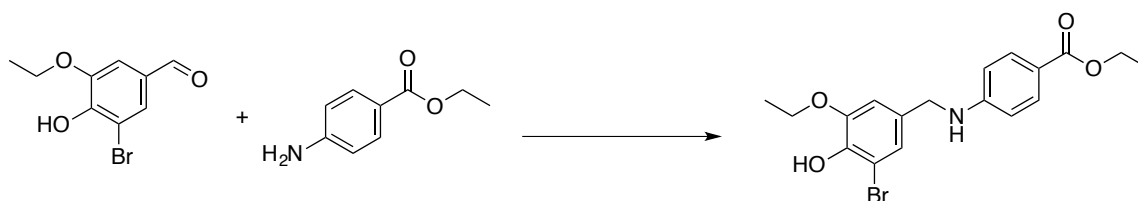
Ethyl 4-((3-bromo-5-ethoxy-4-hydroxybenzyl)amino)benzoate (28)Chemical formula: C₁₈H₂₀BrNO₄; MW: 394.27

Table 24: Chemicals for the preparation of 28

Name	Eq.	mmol	MW	g	(d) g/cm ³	V (mL)
3-bromo-5-ethoxy-4-hydroxybenzaldehyde	1	0.408	245.07	0.100		
Ethyl 4-aminobenzoate	1	0.408	165.19	0.067		
CH ₃ COOH	1	0.408	60.05	0.0403	1.049	0.038
DCE (solvent)			98.95			5
NaBH(OAc) ₃	2	0.816	211.94	0.173		

Procedure: General Procedure 2

Work up: The compound was purified with column chromatography using a gradient of 0 to 50% EtOAc in n-hexane (v/v) and the fractions containing the product were identified by TLC in 254 nm UV light ($R_f = 0.16$). The relevant fractions were evaporated under reduced pressure.

The product was confirmed by NMR.

TLC: Hexane-EtOAc 8:2 (v/v), $R_f = 0.16$

Outcome: Yield: 100 mg, 62.17%, off-white solid

Analysis:

¹H NMR (500 MHz, CDCl₃) 1.34 (t, $J = 7.13$ Hz, 3H), 1.40 (t, $J = 6.99$ Hz, 3H), 4.05 (q, $J = 7.0$ Hz 2H), 4.24 (s, 2H), 4.29 (q, $J = 7.13$ Hz 2H), 4.62 (s, NH), 6.09 (s, OH), 6.56 (d, $J = 8.9$ Hz, 2H), 6.76 (d, $J = 1.79$ Hz, 1H), 7.05 (d, $J = 1.83$ Hz, 1H), 7.85 ppm (d, $J = 8.88$ Hz, 2H)

The reaction was repeated and scaled up:

Table 25: Chemicals for the preparation of compound 28

Name	Eq.	mmol	MW	g	(d) g/cm ³	V (mL)
3-bromo-5-ethoxy-4-hydroxybenzaldehyde (43)	1	4.08	245.07	1.000		
Ethyl 4-aminobenzoate (46)	1	4.08	165.19	0.670		
CH ₃ COOH	1	4.08	60.05	0.403	1.049	0.38
DCE (solvent)			98.95			10
NaBH(OAc) ₃	2	8.16	211.94	1.73		

Procedure: General procedure 2

Work up: The residue was purified with column chromatography using a gradient of 0 to 50% EtOAc in n-hexane (v/v) and the fractions containing the product were identified by TLC in 254

nm UV light ($R_f = 0.16$). More of the final compound could be isolated by recrystallization from impure fractions.

Analysis:

^1H NMR (500 MHz, CDCl_3) 1.36 (t, $J = 7.13$ Hz, 3H), 1.44 (t, $J = 7.00$ Hz, 3H), 4.1 (q, $J = 7.00$ Hz, 2H), 4.27 (s, 2H), 4.31 (q, $J = 7.12$ Hz, 2H) 4.42 (s, 1H), 5.93 (s, 1H), 6.57 (dd, $J = 8.88$ Hz, 2H), 6.78 (d, $J = 1.79$ Hz, 1H), 7.08 (d, $J = 1.85$ Hz, 1H), 7.87 ppm (dd, $J = 8.89$ Hz, 2H).

^{13}C NMR (125 MHz, CDCl_3) 14.46, 14.77 (CH_3), 47.20, 60.28, 65.09 (CH_2), 108.18 (C), 109.93 (CH), 111.73 (2 CH), 119.36 (C), 123.39 (CH), 131.08 (C), 131.52 (2 CH), 142.61, 146.66, 151.44, 166.79 ppm (C)

After confirmation of the product the two batches were united to perform the following reactions and the characterisations.

UPLC-MS: Method B

RT: 1.974 min, Purity: 100%

MS: 392.24, 394.22 $[\text{M}-\text{H}]^-$ neg. mode

Ethyl 4-((4-(benzyloxy)-3-bromo-5-ethoxybenzyl)amino)benzoate (29)

Chemical formula: $\text{C}_{25}\text{H}_{26}\text{BrNO}_4$; MW: 484.39; CAS-Nr: 1243623-96-2

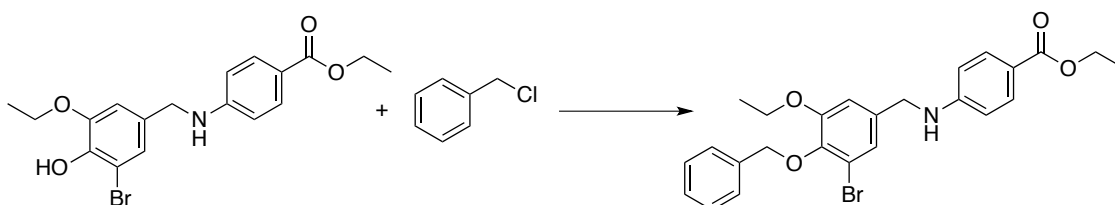


Table 26: Chemicals for the preparation of 29

Name	Eq.	mmol	MW	g	(d) g/cm^3	V mL
Ethyl 4-((3-bromo-5-ethoxy-4-hydroxybenzyl)amino)benzoate	1	0.254	394.27	0.100		
Benzylchloride	1	0.254	126.58	0.032	1.1	0.03
K_2CO_3	1.5	0.381	138.2	0.053		
DMF anhydrous					944 kg/m^3	5 mL

Procedure: General procedure 1

TLC: EtOAc-Hexane 20:80 v/v. $R_f = 0.58$

Purification: Interchim automated column chromatography (Cartidge: 25 g, 50 μ m)
EtOAc:Hexane gradient using the automated methods generated by TLC measurements.
Fractions 10-14 contained the product and were united and solvents were evaporated.

Outcome: Yield: 36.3 mg, 29.5 %, light orange solid

Analysis:

^1H NMR (500 MHz, CDCl_3) 1.36 (t, $J = 7.1$ Hz, 3H), 1.43 (t, $J = 7.0$ Hz, 3H) 4.03 (q, $J = 7.0$ Hz, 2H), 4.28 (s, 2H), 4.31 (q, $J = 7.1$ Hz, 2H), 4.56 (s, 1H), 5.02 (s, 2H), 6.57 (d, $J = 8.7$ Hz, 2H), 6.82 (s, 1H), 7.11 (s, 1H), 7.36 (m, 3H), 7.56 (d, $J = 7.4$ Hz, 1H), 7.87 ppm (d, $J = 8.7$ Hz, 2H)

^{13}C NMR (125 MHz, CDCl_3) 14.56, 14.93 (CH_3), 47.16, 60.40, 64.74, 74.81 (CH_2), 108.76, 111.68, 111.81, 113.86 (CH, aromatic), 118.34, 119.36 (C), 123.25, 128.23, 128.43, 128.67, 130.03, 131.60 (CH, aromatic), 135.81, 137.24, 144.70, 151.51, 153.43, 166.91 (C), 189.90 ppm (CH).

UPLC-MS: Method B

RT: 2.28 min; Purity 88%

The impurity ethyl-4 aminobenzoate comes from **28** (intermediate) and is a starting material of the intermediate reaction.

MS: (ESI) $^+$: 485.2 [M+H] $^+$

Ethyl 4-((3-bromo-4-((4-cyanobenzyl)oxy)-5-ethoxybenzyl) amino)benzoate (**30**)

Chemical formula: $\text{C}_{26}\text{H}_{25}\text{BrN}_2\text{O}_4$; MW: 509.4

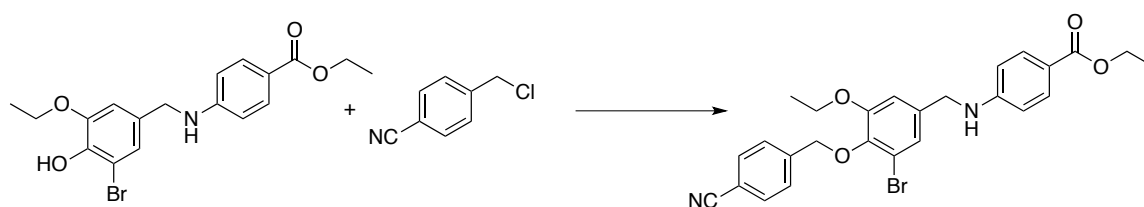


Table 27: Chemicals for the preparation of 30

Name	Eq.	mmol	MW	g	(d)	V
28	1	0.254	394.27	0.100		
4-Chloromethyl benzonitrile	1	0.254	151.60	0.0385		
K_2CO_3	1.5	0.381	138.2	0.053		
DMF anhydrous					944 kg/m 3	5 mL

Purification: Biotage Isolera One automated flash column chromatography

Cartridge: SNAP KP-SIL 25g

Elution system: n-hexane – EtOAc (3 CV 100:0 (v/v); 10 CV linear gradient reaching 40:60 (v/v); 3 CV 40:60 (v/v)); Eluted at 11-12 CV.

TLC: $R_f = 0.34$ in n-hexane - EtOAc 7:3 (v/v). Fraction 4 contained a relevant portion of the product that could be recrystallized from the column chromatography and yielded other 18 mg of white solid.

Outcome: Yield: 56.3 mg, 43.5%; off-white solid

Analysis:

^1H NMR (500 MHz, CDCl_3) 1.36 (t, $J = 7.1$ Hz, 3H), 1.42 (t, $J = 7.0$ Hz, 3H) 4.06 (q, $J = 7.0$ Hz, 2H), 4.34 (q, $J = 7.1$ Hz, 2H), 4.35 (m, 2H), 4.51 (s, 1H), 5.10 (s, 2H), 6.61 (dd, $J = 8.8$ Hz, $J = 1.9$ Hz, 2H), 6.87 (d, $J = 1.8$ Hz, 1H), 7.16 (d, $J = 1.9$ Hz 1H), 7.7 (m, 4H), 7.9 ppm (dd, $J = 8.8$ Hz, $J = 1.9$ Hz, 2H)

^{13}C NMR (125 MHz, CDCl_3) 14.45, 14.77 (CH_3), 47.11, 60.30, 64.69, 73.50 (CH_2), 111.54 (CH, aromatic), 111.71 (C), 111.78 (2 CH, aromatic), 118.02, 118.85, 119.60 (C), 123.21 (CH, aromatic), 128.41, 131.53, 132.16, (2 CH aromatic each), 136.21, 142.72, 144.32, 151.30, 153.13, 166.72 ppm (C).

UPLC-MS: Method B

RT: 2.18; Purity: 97%

MS: (ESI) $^+$ 532.9 $[\text{M}+\text{Na}]^+$

Ethyl 4-((3-bromo-4-((4-chlorobenzyl)oxy)-5-ethoxybenzyl) amino)benzoate (31)

Chemical formula: $\text{C}_{25}\text{H}_{25}\text{BrClNO}_4$; MW: 518.83; CAS-Nr: 1243789-36-7

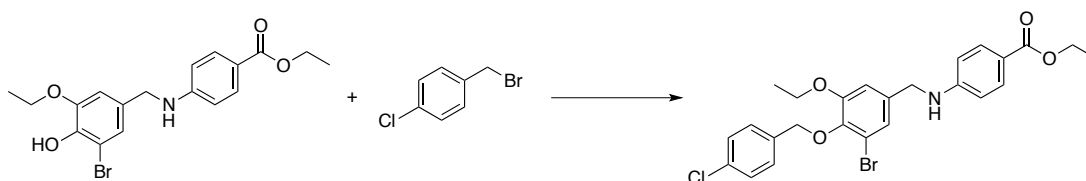


Table 28: Chemicals for the preparation of 31

Name	Eq.	mmol	MW	g	(d)	V
28	1	0.254	394.27	0.100		
4-Chlorobenzyl bromide	1	0.254	205.48	0.052		
K_2CO_3	1.5	0.381	138.2	0.053		
DMF anhydrous					944 kg/m ³	5 mL

Procedure: General procedure 1

Purification: Biotage Isolera One

Cartridge: SNAP KP-SIL 25g

Elution system: *n*-hexane – EtOAc (3 CV 100:0 (v/v); 10 CV linear gradient reaching 40:60 (v/v); 3 CV 40:60 (v/v)); Eluted at 9.3-10.3 CV.

TLC: *n*-hexane-Ethylacetate 7:3 (v/v) $R_f = 0.47$.

Outcome: Yield: 83.4 mg, 63.3%, off-white solid

Analysis:

^1H NMR (500 MHz, CDCl_3) 1.36 (t, $J = 7.1$ Hz, 3H), 1.42 (t, $J = 6.9$ Hz, 3H) 4.03 (q, $J = 6.9$ Hz, 2H), 4.32 (m, 4H), 4.48 (s, 1H), 4.99 (s, 2H), 6.58 (dd, $J = 8.8$ Hz, $J = 1.7$ Hz, 2H), 6.83 (d, $J = 1.7$ Hz, 1H), 7.12 (d, $J = 1.7$ Hz 1H), 7.34 (dd, $J = 8.36$ Hz, $J = 1.8$ Hz, 2H), 7.48 (d, $J = 8.3$ Hz, 2H), 7.9 ppm (dd, $J = 8.7$ Hz, $J = 1.8$ Hz, 2H)

^{13}C NMR (125 MHz, CDCl_3) 14.46, 14.81 (CH_3), 47.14, 60.29, 64.70, 73.84 (CH_2), 111.65, 111.76 (CH, aromatic), 118.21, 119.52 (C), 123.26, 128.48, 129.83, 131.53 (CH, aromatic), (C), 133.91, 135.73, 135.85, 144.51, 151.35, 153.27, 166.75 ppm (C).

UPLC-MS: Method B

RT: 2.35 min, Purity 100%

MS: 125.0 [$\text{C}_7\text{H}_7\text{Cl}$] $^-$

Ethyl 4-((3-bromo-4-((2-chlorobenzyl)oxy)-5-ethoxybenzyl) amino)benzoate (32)

Chemical formula: $\text{C}_{25}\text{H}_{25}\text{BrClNO}_4$; MW: 518.83; CAS-Nr: 1243732-78-6

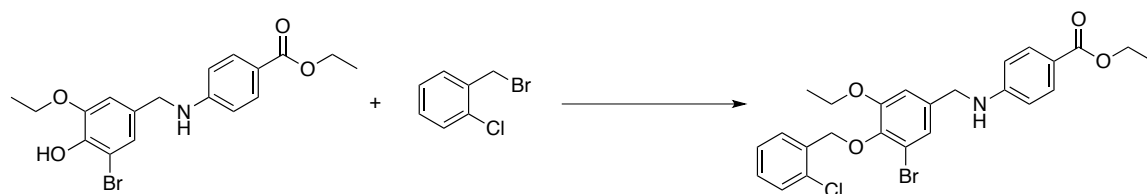


Table 29: Chemicals for the preparation of 32

Name	Eq.	mmol	MW	g	(d)	V
28	1	0.254	394.27	0.100		
2-Chlorobenzyl bromide	1	0.254	205.48	0.052		
K_2CO_3	1.5	0.381	138.2	0.053		
DMF anhydrous					944 kg/m ³	5 mL

Procedure: General procedure 1

Purification: Biotage Isolera One

Cartridge: SNAP KP-SIL 10 g

Elution system: *n*-hexane – EtOAc (3 CV 100:0 (v/v); 10 CV linear gradient reaching 40:60 (v/v); 3 CV 40:60 (v/v)); Eluted at 9 CV.

TLC: *n*-hexane-Ethylacetate 7:3 (v/v) $R_f = 0.41$.

Outcome: Yield: 52.6 mg, 39.9%; white solid

Analysis:

^1H NMR (500 MHz, CDCl_3) 1.29 (dt, $J = 9.5, J = 7.1$ Hz, 6H), 3.94 (q, $J = 7.0$ Hz, 2H), 4.22 – 4.26 (m, 4H), 4.42 (s, 1H), 5.09 (s, 2H), 6.5 (d, $J = 8.8$, 2H), 6.76 (d, $J = 1.8$ Hz, 1H), 7.05 (d, $J = 1.9$ Hz, 1H), 7.17 – 7.2 (m, 1H), 7.24 (td, $J = 7.5$ Hz, $J = 1.3$ Hz, 1H), 7.3 (dd, $J = 7.8$ Hz, $J = 1.2$ Hz, 1H), 7.73 (dd, $J = 7.6$ Hz, $J = 1.45$ Hz, 1H), 7.8 ppm (d, $J = 8.8$ Hz, 2H)

^{13}C NMR (125 MHz, CDCl_3) 14.46, 14.73 (CH_3), 47.15, 60.28, 64.73, 71.44 (CH_2), 111.65, 111.76 (CH), 118.14, 119.48 (C), 123.23, 126.79, 128.90, 129.09, 129.79, 131.53 (CH), 132.68, 135.32, 135.85, 144.72, 151.38, 153.33, 166.76 ppm (C)

UPLC-MS: Method B

RT: 2.35 min, Purity: 100%

MS: 125.0 [$\text{C}_7\text{H}_7\text{Cl}$] $^-$

Ethyl 4-((3-bromo-4-((3-chlorobenzyl)oxy)-5-ethoxybenzyl)amino)benzoate (33)

Chemical formula: $\text{C}_{25}\text{H}_{25}\text{BrClNO}_4$; MW: 518.83; CAS-Nr: 1243805-45-9

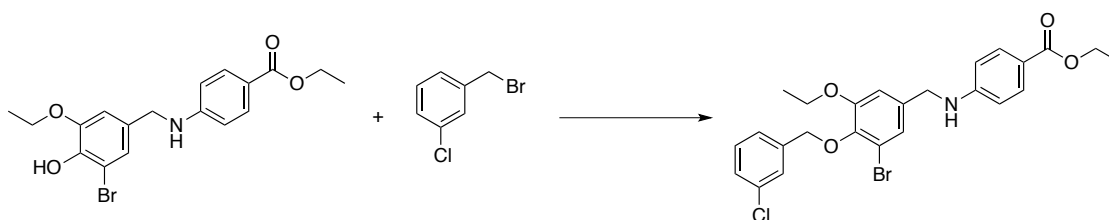


Table 30: Chemicals for the preparation of 33

Name	Eq.	mmol	MW	g	(d)	V
28	1	0.254	394.27	0.100		
2-Chlorobenzyl bromide (31)	1	0.254	205.48	0.052		
K_2CO_3	1.5	0.381	138.2	0.053		
DMF anhydrous					944 kg/m ³	5 mL

Procedure: General procedure 1

Purification: Biotage Isolera One

Cartridge: SNAP KP-SIL 10 g

Elution system: *n*-hexane – EtOAc (5 CV 100:0 (v/v); 10 CV linear gradient reaching 40:60 (v/v); 3 CV 40:60 (v/v)); Eluted at 10 CV.

TLC: *n*-hexane-Ethylacetate 7:3 (v/v) $R_f = 0.46$.

Outcome: Yield: 16.8 mg, 12.7%, yellow wax

Analysis:

^1H NMR (500 MHz, CDCl_3) 1.24 (t, $J = 7.1$ Hz, 3H), 1.29 (t, $J = 6.9$ Hz, 3H), 3.89 (q, $J = 6.9$ Hz, 2H), 4.15 (d, $J = 5.39$ Hz, 2H), 4.19 (q, $J = 7.1$ Hz, 2H), 4.59 (t, $J = 5.1$ Hz, 1H), 4.87 (s, 2H), 6.44 (d, $J = 8.7$ Hz, 2H), 6.7 (s, 1H), 6.97 (s, 1H), 7.18 (m, 2H), 7.29 (t, $J = 3.54$ Hz, 1H), 7.48 (s, 1H), 7.75 ppm (d, $J = 8.67$ Hz, 2H)

^{13}C NMR (125 MHz, CDCl_3) 14.52, 14.84 (CH_3) 47.02, 60.32, 64.68, 73.87 (CH_2) 111.54 (CH), 111.77 (2 CH), 118.9, 119.22 (C), 123.01, 126.4, 128.2, 128.52, 129.65 (CH), 131.51 (2 CH), 134.2, 136.16, 139.33, 144.36, 151.56, 153.23, 166.86 ppm (C)

UPLC-MS: Method B

RT: 2.59 min, Purity: 100%

MS: 124.8 [$\text{C}_7\text{H}_7\text{Cl}$]⁻

Ethyl 4-((3-bromo-5-ethoxy-4-((4-methylbenzyl)oxy)benzyl)amino)benzoate (34)

Chemical Formula: $\text{C}_{26}\text{H}_{28}\text{BrNO}_4$; MW: 498.42; CAS-Nr: 1243773-10-5

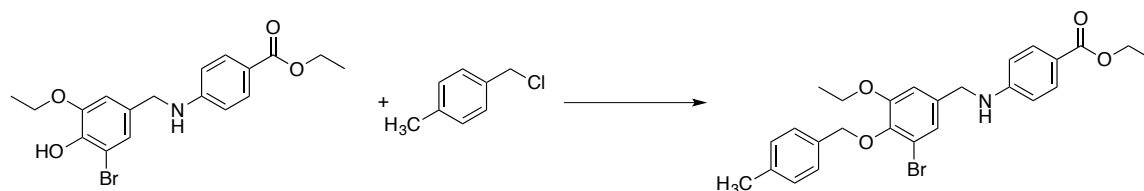


Table 31: Chemicals for the preparation of 34

Name	Eq.	mmol	MW	g	(d)	V
28	1	0.254	394.27	0.100		
1-(chloromethyl)-4-methylbenzene	1	0.254	140.61	0.0357	1.062	0.034 mL
K_2CO_3	1.5	0.381	138.2	0.053		
DMF anhydrous					944 kg/m ³	5 mL

Procedure: General procedure 1

Purification: Biotage Isolera One

Cartridge: SNAP KP-SIL 25 g

Elution system: *n*-hexane – EtOAc (5 CV 100:0 (v/v); 10 CV linear gradient reaching 40:60 (v/v); 3 CV 40:60 (v/v)); Eluted at 10-12 CV.

TLC: *n*-hexane-Ethylacetate 7:3 (v/v) $R_f = 0.43$.

Outcome: Yield: 68.4 mg, 54%, light yellow solid

Analysis:

^1H NMR (500 MHz, CDCl_3) 1.36 (t, $J = 7.1$ Hz, 3H), 1.44 (t, $J = 7.0$ Hz, 3H), 2.36 (s, 3H), 4.04 (q, $J = 7.0$, 2H), 4.28-4.34 (m, 4H), 4.45 (bs, 1H), 5.00 (s, 2H), 6.58 (d, $J = 8.7$ Hz, 2H), 6.83 (s, 1H), 7.12 (s, 1H), 7.18 (d, $J = 7.8$ Hz, 2H), 7.43 (d, $J = 7.8$ Hz, 2H), 7.88 ppm (d, $J = 8.7$ Hz, 2H)

^{13}C NMR (125 MHz, CDCl_3) 14.46, 14.84, 21.26 (CH_3), 47.16, 60.27, 64.74, 74.63 (CH_2), 111.74, 111.81 (3 CH), 118.34, 119.44 (C), 123.30, 128.67, 129.00, 131.53 (6 CH), 134.18, 135.55, 137.87, 144.85, 151.41, 153.41, 166.77 ppm (C)

UPLC-MS: Method B

RT: 2.32 min, Purity 100%

MS: 105.1 [C_8H_9] $^+$

4-((2,4-Dichlorobenzyl)oxy)-3-ethoxybenzaldehyde (35)

Chemical Formula: $\text{C}_{16}\text{H}_{14}\text{Cl}_2\text{O}_3$; MW: 325.19; CAS-Nr: 331464-01-8 (Zhen et al. 2006)

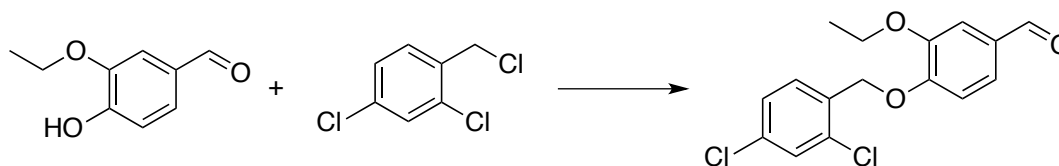


Table 32: Chemicals for the preparation of 35

Name	Eq.	mmol	MW	g	(d)	V
3-ethoxy-4-hydroxy benzaldehyde	1	0.602	166.18	0.100		
2,4-dichlorobenzylchloride	1	0.602	195.47	0.118	1.407	0.084 mL
K_2CO_3	1.5	0.903	138.2	0.124		
DMF anhydrous					944 kg/m ³	3 mL

Procedure: General procedure 1

Purification: The compound was forming a solid when the reaction was quenched and was therefore filtrated under vacuum to yield the clean product.

TLC: *n*-hexane-Ethyacetate 7:3 (v/v) $R_f = 0.47$.

Outcome: Yield: 193 mg, 99%, white fluffy solid

Analysis:

^1H NMR (500 MHz, CDCl_3) 1.43 (t, $J = 6.98$ Hz, 3H), 4.12 (q, $J = 6.98$ Hz, 2H), 5.2 (s, 2H), 6.9 (d, $J = 8.2$ Hz, 1H), 7.20 – 7.23 (m, 1H), 7.32 – 7.37 (m, 3H), 7.46 (d, $J = 8.34$ Hz, 1H), 9.78 ppm (s, 1H)

^{13}C NMR (125 MHz, CDCl_3) 14.73 (CH_3), 64.63, 67.46 (CH_2), 110.97, 112.99, 126.28, 127.47, 129.23, 129.26 (CH), 130.86, 132.76, 132.8, 134.29, 149.55, 153.24 (C), 190.9 ppm (CH)

3-Bromo-4-((2,4-dichlorobenzyl)oxy)benzaldehyde (36)

Chemical Formula: $\text{C}_{14}\text{H}_9\text{BrCl}_2\text{O}_2$; MW: 360.03; CAS-Nr: 443292-21-5

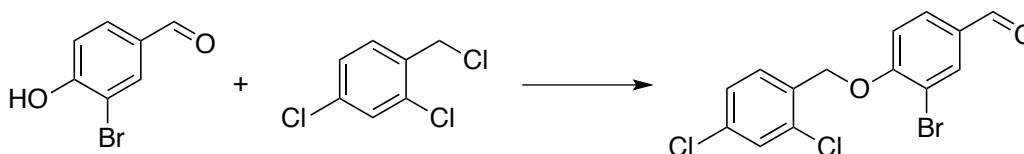


Table 33: Chemicals for the preparation of 36

Name	Eq.	mmol	MW	g	(d)	V
3-bromo-4-hydroxy benzaldehyde	1	0.497	201.02	0.100		
2,4-dichlorobenzylchloride	1	0.497	195.47	0.097	1.407	0.068 mL
K_2CO_3	1.5	0.745	138.2	0.103		
DMF anhydrous					944 kg/m ³	3 mL

Procedure: General procedure 1

Purification: The compound was forming a solid when the reaction was quenched and was therefore filtrated under vacuum to yield the clean product.

TLC: *n*-hexane-Ethyacetate 7:3 (v/v) $R_f = 0.6$

Outcome: Yield: 119.2 mg, 66%, white solid

Analysis:

¹H NMR (500 MHz, CDCl₃) 1.54 (s, 2H), 5.28 (s, 2H), 7.06 (d, *J* = 8.5 Hz, 1H), 7.33 (dd, *J* = 8.5 Hz, *J* = 1.6 Hz, 1H), 7.45 (d, *J* = 1.8 Hz, 1H), 7.64 (d, *J* = 8.4 Hz, 1H), 7.82 (dd, *J* = 8.5 Hz, *J* = 1.8 Hz, 1H), 8.13 (d, *J* = 1.8 Hz, 1H), 9.87 ppm (s, 1H)

¹³C NMR (125 MHz, CDCl₃) 67.57 (CH₂), 112.87 (CH), 113.24 (C), 127.64, 129.26, 129.29, 131.11 (CH), 131.30, 131.93, 132.71, 134.60 (C), 134.74 (CH), 159.15 (C), 189.44 ppm (CH)

3-Bromo-4-((2,4-dichlorobenzyl)oxy)-5-methoxybenzaldehyde (37)

Chemical Formula: C₁₅H₁₁BrCl₂O₃; MW: 390.05; CAS-Nr: 345980-28-1

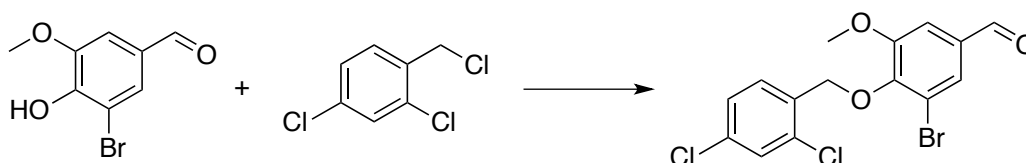


Table 34: Chemicals for the preparation of 37

Name	Eq.	mmol	MW	g	(d)	V
3-bromo-4-hydroxy-5-methoxybenzaldehyde	1	0.433	231.05	0.100		
2,4-dichlorobenzylchloride	1	0.433	195.47	0.085	1.407	0.06 mL
K ₂ CO ₃	1.5	0.650	138.2	0.0898		
DMF anhydrous					944 kg/m ³	3 mL

Procedure: General procedure 1

Purification: The compound was forming a solid when the reaction was quenched and was therefore filtrated under vacuum to yield the clean product.

TLC: *n*-hexane-Ethyacetate 7:3 (v/v) R_f = 0.59

Outcome: Yield: 155.8 mg, 92%, white solid

Analysis:

¹H NMR (500 MHz, CDCl₃) 3.95 (s, 3H), 5.25 (s, 2H), 7.34 (dd, *J* = 8.3 Hz, *J* = 2.1 Hz, 1H), 7.44 (dd, *J* = 1.8 Hz, *J* = 0.8 Hz, 2H), 7.7 (d, *J* = 1.8 Hz, 1H), 7.72 (d, *J* = 8.3 Hz, 1H), 9.89 ppm (s, 1H)

¹³C NMR (125 MHz, (CD₃)₂SO) 56.91 (CH₃), 71.13 (CH₂), 112.02 (CH), 117.82 (C), 127.63, 127.97, 129.25, 132.50 (CH), 133.75, 133.84, 134.19, 134.26, 149.49, 154.20 (C), 191.47 ppm (CH)

4-((2,4-Dichlorobenzyl)oxy)benzaldehyde (38)

Chemical Formula: C₁₄H₁₀Cl₂O₂; MW: 281.13; CAS-Nr: 70627-17-7 (Marrapu et al. 2011)

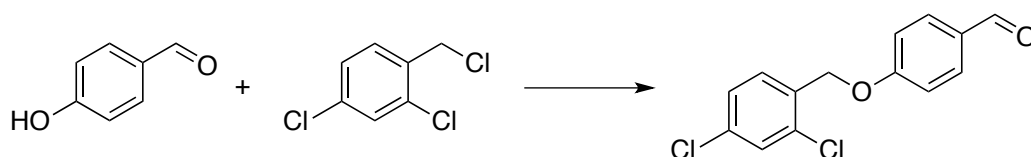


Table 35: Chemicals for the preparation of 38

Name	Eq.	mmol	MW	g	(d)	V
4-hydroxybenzaldehyde	1	0.819	122.12	0.100		
2,4-dichlorobenzylchloride	1	0.819	195.47	0.160	1.407	0.113 mL
K ₂ CO ₃	1.5	1.228	138.2	0.169		
DMF anhydrous					944 kg/m ³	3 mL

Procedure: General procedure 1

Purification: The compound was forming a solid when the reaction was quenched and was therefore filtrated under vacuum to yield the clean product.

TLC: *n*-hexane-Ethylacetate 7:3 (v/v) R_f = 0.63

Outcome: Yield: 183.6 mg, 80%, white solid

Analysis:

¹H NMR (500 MHz, CDCl₃) 5.21 (s, 2H), 7.08 (d, *J* = 8.7 Hz, 2H), 7.29 (dd, *J* = 8.3 Hz, *J* = 2.05 Hz, 1H), 7.45 (d, *J* = 2.05 Hz, 1H), 7.47 (d, *J* = 8.3 Hz, 1H), 7.86 (d, *J* = 8.8 Hz, 2H), 9.9 ppm (s, 1H)

¹³C NMR (125 MHz, CDCl₃) 66.81 (CH₂), 115.12 (2 CH), 129.42, 129.64 (CH), 130.54 (C), 132.06 (2 CH), 132.38, 133.35, 134.63, 163.14 (C), 190.70 ppm (CH)

Ethyl 4-((4-((2,4-dichlorobenzyl)oxy)-3-ethoxybenzyl)amino)benzoate (39)

Chemical Formula: C₂₅H₂₅Cl₂NO₄; MW: 474.38; CAS-Nr: 1243813-76-4

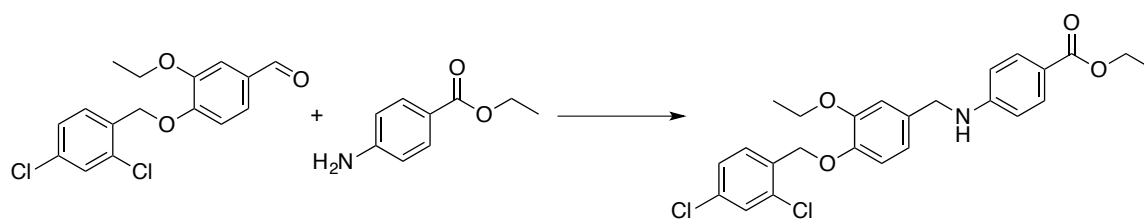


Table 36: Chemicals for the preparation of 39

Name	Eq.	mmol	MW	g	(d) g/cm ³	V (mL)
4-((2,4-dichlorobenzyl)oxy)-3-ethoxybenzaldehyde	1	0.154	325.19	0.05		

Ethyl 4-aminobenzoate	1	0.154	165.19	0.0254		
CH ₃ COOH	1	0.154	60.05	0.0092	1.049	0.0087
DCE (solvent)			98.95			5
NaBH(OAc) ₃	2	0.308	211.94	0.0653		

Procedure: General procedure 2

Purification: Biotage Isolera One

Cartridge: SNAP KP-SIL 10 g

Elution system: *n*-hexane – EtOAc (5 CV 100:0 (v/v); 10 CV linear gradient reaching 40:60 (v/v); 3 CV 40:60 (v/v)); Eluted at 11-12 CV.

TLC: *n*-hexane-Ethylacetate 7:3 (v/v) R_f = 0.61.

Outcome: Yield: 24 mg, 33%, white solid

Analysis:

¹H NMR (500 MHz, CDCl₃) 1.28 (t, *J* = 7.1 Hz, 3H), 1.37 (t, *J* = 7.0 Hz, 3H), 4.01 (q, *J* = 7.0 Hz, 2H), 4.24 (q, *J* = 7.1 Hz, 4H), 4.34 (t, *J* = 5.1 Hz, 1H), 5.1 (s, 2H), 6.50-6.53 (m, 2H), 6.77 (s, 2H), 6.83 (s, 1H), 7.18-7.20 (m, 2H), 7.33 (d, *J* = 2.1 Hz, 1H), 7.48 (d, *J* = 8.3 Hz, 1H), 7.78-7.81 ppm (m, 2H)

¹³C NMR (125 MHz, CDCl₃) 14.46, 14.89 (CH₃), 47.55, 60.22, 64.63, 68.04 (CH₂), 111.65, 112.99, 115.21 (CH), 119.15 (C), 119.80, 127.28, 129.06, 129.49, 131.51 (CH), 131.51, 132.31, 132.81, 133.79, 133.89, 147.42, 149.54, 151.65, 166.81 ppm (C)

UPLC-MS: Method B

RT: 2.35 min, Purity 100%

MS: 309.1 [C₁₆H₁₅Cl₂O₂]⁺ 158.9 [C₇H₅Cl₂]⁺ (two main fragments corresponding to building block AB and A)

Ethyl 4-((3-bromo-4-((2,4-dichlorobenzyl)oxy)benzyl)amino)benzoate (40)

Chemical Formula: C₂₃H₂₀BrCl₂NO₃; MW: 509.22

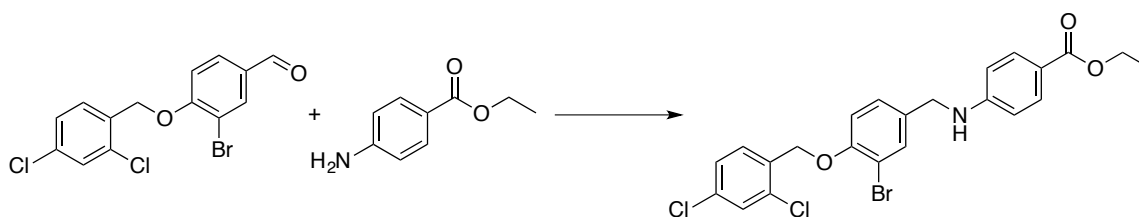


Table 37: Chemicals for the preparation of 40

Name	Eq.	mmol	MW	g	(d) g/cm ³	V (mL)
------	-----	------	----	---	-----------------------	--------

3-bromo-4-((2,4-dichlorobenzyl)oxy)benzaldehyde	1	0.194	360.03	0.07		
Ethyl 4-aminobenzoate	1	0.194	165.19	0.032		
CH ₃ COOH	1.5	0.292	60.05	0.0175	1.049	0.0167
DCE (solvent)			98.95			5
NaBH(OAc) ₃	2	0.389	211.94	0.0825		

Procedure: General procedure 2

The amount of acetic acid has been adapted to 1.5 eq.

Purification: Biotage Isolera One

Cartridge: SNAP KP-SIL 10 g

Elution system: *n*-hexane – EtOAc (5 CV 100:0 (v/v); 10 CV linear gradient reaching 40:60 (v/v); 3 CV 40:60 (v/v)); Eluted at 11 CV.

TLC: *n*-hexane-Ethylacetate 7:3 (v/v) R_f = 0.58.

Outcome: Yield: 51 mg, 52%, light yellow fluffy crystals

Analysis:

¹H NMR (500 MHz, CDCl₃) 1.28 (t, *J* = 7.1 Hz, 3H), 4.22-4.26 (m, 4H), 4.38 (s, 1H), 5.12 (s, 2H), 6.5 (d, *J* = 8.8 Hz, 2H), 6.83 (d, *J* = 8.4 Hz, 1H), 7.16 (dd, *J* = 8.4 Hz, *J* = 2.1 Hz, 1H), 7.24 (dd, *J* = 8.4 Hz, *J* = 2.0 Hz, 1H), 7.35 (d, *J* = 2.0 Hz, 1H), 7.5 (d, *J* = 2.1 Hz, 1H), 7.58 (d, *J* = 8.3 Hz, 1H), 7.8 ppm (d, *J* = 8.9 Hz, 2H)

¹³C NMR (125 MHz, CDCl₃) 14.82 (CH₃), 45.06, 60.01, 67.67 (CH₂), 111.57 (C), 111.81 (CH), 114.51 (2 CH), 116.99 (C), 128.08, 128.28, 129.41 (CH), 131.36 (3 CH), 132.20 (CH), 133.69, 133.80, 134.02, 134.35, 152.85, 153.44, 166.26 ppm (C)

UPLC-MS: Method B

RT: 2.39 min, Purity 100%

MS: 158.9 [C₇H₅Cl₂]⁺

Ethyl 4-((3-bromo-4-((2,4-dichlorobenzyl)oxy)-5-methoxybenzyl)amino)benzoate (41)

Chemical Formula: C₂₄H₂₂BrCl₂NO₄; MW: 539.25; CAS-Nr: 1243831-38-0

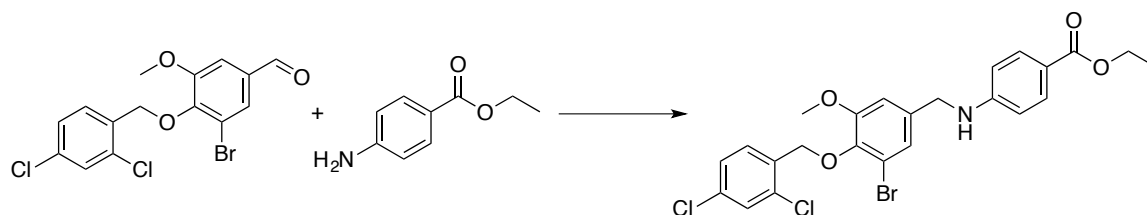


Table 38: Chemicals for the preparation of 41

Name	Eq.	mmol	MW	g	(d) g/cm ³	V (mL)
3-bromo-4-((2,4-dichlorobenzyl)oxy)-5-methoxybenzaldehyde	1	0.563	390.05	0.200		
Ethyl 4-aminobenzoate	1	0.513	165.19	0.085		
CH ₃ COOH	2	1.024	60.05	0.0615	1.049	0.059
DCE (solvent)			98.95			5
NaBH(OAc) ₃	2	1.024	211.94	0.217		

Procedure: General procedure 2

Purification: Biotage Isolera One

Cartridge: SNAP KP-SIL Ultra 25 g

Elution system: *n*-hexane – EtOAc (25 CV linear gradient reaching 20:80 (v/v)) Eluted at 9-12 CV

TLC: *n*-hexane-Ethylacetate 7:3 (v/v) R_f = 0.37. (Impurity R_f = 0.31; both broad spots)

Outcome: Yield: 141.2 mg, 46 %, off-white solid

Analysis:

¹H NMR (500 MHz, CDCl₃) 1.29 (t, *J* = 7.1 Hz, 3H), 3.74 (s, 2H), 4.23-4.27 (m, 4H), 4.42 (t, *J* = 5.5 Hz, 1H), 5.02 (s, 2H), 6.50-6.53 (m, 2H), 6.78 (d, *J* = 1.9 Hz, 1H), 7.07 (d, *J* = 1.9 Hz, 1H), 7.23 (dd, *J* = 8.3 Hz, *J* = 2.1 Hz, 1H), 7.32 (d, *J* = 2.1 Hz, 1H), 7.66 (d, *J* = 8.3 Hz, 1H), 7.80-7.83 ppm (m, 2H)

¹³C NMR (125 MHz, CDCl₃) 14.45 (CH₃), 47.16 (CH₂), 56.15 (CH₃), 60.30, 70.82 (CH₂), 110.55 (CH), 111.78 (2 CH), 118.15, 119.58 (C), 123.30, 127.16, 128.88, 130.38 (CH), 131.54 (2 CH), 133.12, 133.89, 133.97, 136.19, 144.25, 151.31, 153.98, 166.75 ppm (C)

UPLC-MS: Method B

RT: 2.416 min, Purity: 100%

MS: 158.9 [C₇H₅Cl₂]⁺

Ethyl 4-((4-((2,4-dichlorobenzyl)oxy)benzyl)amino)benzoate (42)

Chemical Formula: C₂₃H₂₁Cl₂NO₃; MW: 430.33; CAS-Nr: 1243692-81-0

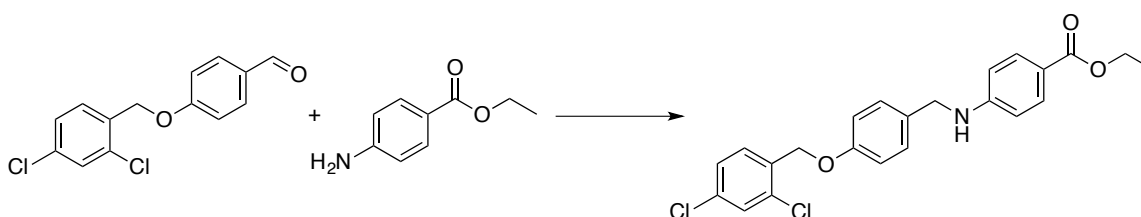


Table 39: Chemicals for the preparation of 42

Name	Eq.	mmol	MW	g	(d) g/cm ³	V (mL)
4-((2,4-Dichlorobenzyl)oxy)benzaldehyde	1	0.284	281.13	0.08		
Ethyl 4-aminobenzoate	1	0.284	165.19	0.047		
CH ₃ COOH	1.5	0.426	60.05	0.025	1.049	0.024
DCE (solvent)			98.95			5
NaBH(OAc) ₃	2	0.568	211.94	0.12		

Procedure: General procedure 2

The amount of acetic acid has been adapted to 1.5 eq.

Purification: Biotage Isolera One

Cartridge: SNAP KP-SIL 10 g

Elution system: *n*-hexane – EtOAc (7 CV 100:0 (v/v); 10 CV linear gradient reaching 40:60 (v/v); 3 CV 40:60 (v/v)); Eluted at 11-13.5 CV.

TLC: *n*-hexane-Ethylacetate 7:3 (v/v) R_f = 0.62.

Outcome: Yield: 81.6 mg, 66%, light yellow solid

Analysis:

¹H NMR (500 MHz, CDCl₃) 1.19 (t, *J* = 7.1 Hz, 3H), 4.13-4.17 (m, 4H), 4.26 (s, 1H), 4.95 (s, 2H), 6.42 (d, *J* = 8.5 Hz, 2H), 6.79 (d, *J* = 8.3 Hz, 2H), 7.1-7.12 (m, 3H), 7.33 (d, *J* = 8.3 Hz, 1H), 7.71 ppm (d, *J* = 8.4 Hz, 2H)

¹³C NMR (125 MHz, CDCl₃) 14.47 (CH₃), 47.15, 60.22, 66.69 (CH₂), 111.63, 115.12 (CH), 119.07 (c), 127.34, 128.86, 129.25, 129.61 (CH), 131.21 (C), 131.52 (CH), 133.18, 133.34, 134.18, 151.65, 157.73, 166.82 ppm (C)

UPLC-MS: Method B

RT: 2.34 min, Purity 97.55%

MS: 265.0 [C₁₄H₁₁Cl₂O]⁺ 158.9 [C₇H₅Cl₂]⁺

Methyl 4-((3-bromo-4-((2,4-dichlorobenzyl)oxy)-5-ethoxybenzyl)amino)benzoate (43)

Chemical Formula: C₂₄H₂₂BrCl₂NO₄; MW: 539.25

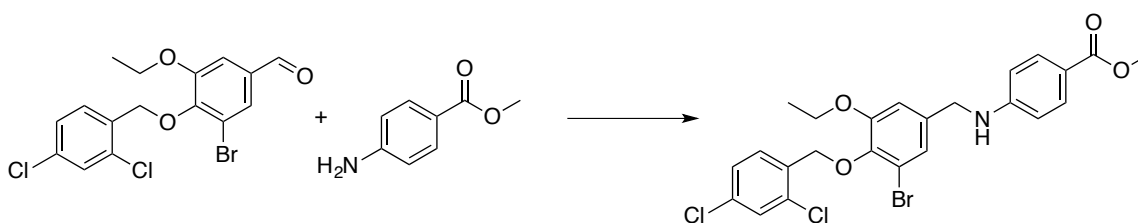


Table 40: Chemicals for the preparation of 43

Name	Eq.	mmol	MW	g	(d) g/cm ³	V (mL)
3-Bromo-4-((2,4-dichlorobenzyl)oxy)-5-ethoxybenzaldehyde	1	0.247	404.08	0.100		
Methyl 4-aminobenzoate	1	0.247	151.17	0.0374		
CH ₃ COOH	1	0.247	60.05	0.0148	1.049	0.014
DCE (solvent)			98.95			2
NaBH(OAc) ₃	2	0.494	211.94	0.1047		

Procedure: General procedure 2

Purification: Biotage Isolera One

Cartridge: SNAP KP-SIL 10 g

Elution system: *n*-hexane – EtOAc (3 CV 100:0 (v/v); 10 CV linear gradient reaching 40:60 (v/v); 3 CV 40:60 (v/v)); Eluted at 9 CV

TLC: *n*-hexane-Ethyl acetate 7:3 (v/v) R_f = 0.38

Outcome: Yield: 60.7 mg, 45 %, beige solid

Analysis:

¹H NMR (500 MHz, CDCl₃) 1.37 (t, *J* = 6.9 Hz, 3H), 3.85 (s, 3H), 4.02 (q, *J* = 6.9 Hz, 2H), 4.32 (d, *J* = 5.4 Hz, 2H), 4.49 (s, 1H), 5.11 (s, 2H), 6.58 (d, *J* = 8.2 Hz, 2H), 6.84 (s, 1H), 7.13 (s, 1H), 7.29 (d, *J* = 8.2 Hz, 1H), 7.39 (s, 1H), 7.75 (d, *J* = 8.3 Hz, 1H), 7.87 ppm (d, *J* = 8.4 Hz, 2H)

¹³C NMR (125 MHz, CDCl₃) 14.72 (CH₃), 47.12 (CH₂), 51.60 (CH₃), 64.71, 70.86 (CH₂), 111.57 (CH), 111.80 (2 CH), 118.07, 119.19 (C), 123.21, 127.13, 128.89, 130.59 (CH), 131.59 (2 CH), 133.21 (C), 134.01 (2 C), 135.97, 144.51, 151.40, 153.23 ppm (C)

UPLC-MS: Method B

RT: 2.47 min, Purity 100%

MS: 158.9 [C₇H₅Cl₂]⁺

***t*-butyl 4-((3-Bromo-4-((2,4-dichlorobenzyl)oxy)-5-ethoxybenzyl)amino)benzoate (44)**

Chemical Formula: C₂₇H₂₈BrCl₂NO₄; MW: 581.33

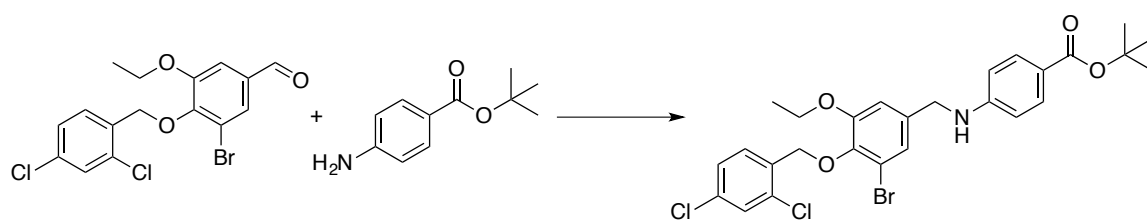


Table 41: Chemicals for the preparation of 44

Name	Eq.	mmol	MW	g	(d) g/cm ³	V (mL)
3-bromo-4-((2,4-dichlorobenzyl)oxy)-5-ethoxybenzaldehyde	1	0.247	404.08	0.100		
<i>t</i> -butyl 4-aminobenzoate	1	0.247	193.25	0.0477		
CH ₃ COOH	1	0.247	60.05	0.0148	1.049	0.014
DCE (solvent)			98.95			2
NaBH(OAc) ₃	2	0.494	211.94	0.1047		

Procedure: General procedure 2

Purification: Biotage Isolera One

Cartridge: SNAP KP-SIL 10 g

Elution system: *n*-hexane – EtOAc (3 CV 100:0 (v/v); 10 CV linear gradient reaching 40:60 (v/v); 3 CV 40:60 (v/v)); Eluted at 8 CV

TLC: *n*-hexane-Ethyl acetate 7:3 (v/v) R_f = 0.58

Outcome: Yield: 95 mg, 65%, yellow wax

Analysis:

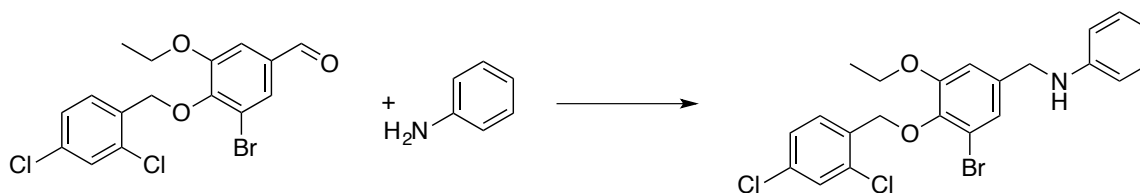
¹H NMR (500 MHz, CDCl₃) 1.26 (t, *J* = 7.0 Hz, 3H), 1.46 (s, 9H), 3.90 (q, *J* = 6.96 Hz, 2H), 4.19 (s, 2H), 4.62 (s, 1H), 5.00 (s, 2H), 6.45 (d, *J* = 8.7 Hz, 2H), 6.73 (s, 1H), 7.00 (s, 1H), 7.18 (d, *J* = 8.3 Hz, 1H), 7.28 (s, 1H), 7.65 (d, *J* = 8.3 Hz, 1H), 7.71 ppm (d, *J* = 8.7 Hz, 2H)

¹³C NMR (125 MHz, CDCl₃) 14.73 (CH₃), 28.33 (3 CH₃), 47.16, 64.69, 70.84, 80.00 (CH₂), 111.50 (CH), 111.70 (2 CH), 118.04, 121.15 (C), 123.14, 127.12, 128.88, 130.58 (CH), 131.36 (2 CH), 133.20, 133.97, 134.03, 136.15, 144.43, 151.03, 153.22, 166.02 ppm (C)

UPLC-MS: Method B

RT: min, Purity 100%

MS: 158.9 [C₇H₅Cl₂]⁺

***N*-(3-Bromo-4-((2,4-dichlorobenzyl)oxy)-5-ethoxybenzyl)aniline (45)**Chemical Formula: C₂₂H₂₀BrCl₂NO₂; MW: 481.21; CAS-Nr: 1311034-78-2**Table 42: Chemicals for the preparation of 45**

Name	Eq.	mmol	MW	g	(d) g/cm ³	V (mL)
3-bromo-4-((2,4-dichlorobenzyl)oxy)-5-ethoxybenzaldehyde	1	0.247	404.08	0.100		
Aniline	1	0.247	93.13	0.023	1.021	0.225*
CH ₃ COOH	1	0.247	60.05	0.0148	1.049	0.014
DCE (solvent)			98.95			2
NaBH(OAc) ₃	2	0.494	211.94	0.1047		

*Aniline dilution in DCE: 0.1 mL aniline + 0.9 mL DCE: 0.225mL = 0.023 g

Procedure: General procedure 2

Purification: Biotage Isolera One

Cartridge: SNAP KP-SIL 10 g

Elution system: *n*-hexane – EtOAc (3 CV 100:0 (v/v); 10 CV linear gradient reaching 40:60 (v/v); 3 CV 40:60 (v/v)). Eluted at 7 CV.

Outcome: Yield: 74.3 mg, 63 %, yellow-brown solid

Analysis:

¹H NMR (500 MHz, CDCl₃) 1.37 (t, *J* = 6.98 Hz, 3H), 4.03 (m, 3H), 4.26 (s, 2H), 5.11 (s, 2H), 6.63 (d, *J* = 7.9 Hz, 2H), 6.74 (t, *J* = 7.3 Hz, 1H), 6.89 (s, 1H), 7.16 (s, 1H), 7.19 (t, *J* = 7.4 Hz, 2H), 7.3 (d, *J* = 8.6 Hz, 1H), 7.39 (s, 1H), 7.77 ppm (d, *J* = 8.4 Hz, 1H)¹³C NMR (125 MHz, CDCl₃) 14.70 (CH₃), 47.86, 64.72, 70.84 (CH₂), 111.87 (CH), 113.01 (2 CH), 117.88 (C), 117.99, 123.34, 127.09, 128.86 (CH), 129.30 (2 CH), 130.59 (CH), 133.20, 133.94, 134.14, 137.12, 144.38, 147.84, 153.13 ppm (C)

UPLC-MS: Method B

RT: 2.52 min, Purity 100%

MS: 158.9 [C₇H₅Cl₂]⁺**4-((3-Bromo-4-((2,4-dichlorobenzyl)oxy)-5-ethoxybenzyl)amino)benzamide (46)**

Chemical Formula: C₂₃H₂₁BrCl₂N₂O₃; MW: 524.24

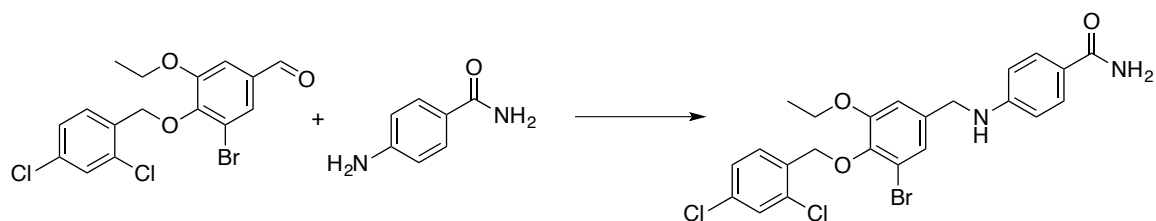


Table 43: Chemicals for the preparation of 46

Name	Eq.	mmol	MW	g	(d) g/cm ³	V (mL)
3-bromo-4-((2,4-dichlorobenzyl)oxy)-5-ethoxybenzaldehyde	1	0.247	404.08	0.100		
4-aminobenzamide	1	0.247	136.15	0.0336		
CH ₃ COOH	1	0.247	60.05	0.0148	1.049	0.014
DCE (solvent)			98.95			2
NaBH(OAc) ₃	2	0.494	211.94	0.1047		

Procedure: General procedure 2

Purification: The compound was crystallized in DCM and obtained by filtration.

Outcome: Yield: 9.8 mg, 7.6%, off-white solid

Analysis:

¹H NMR (500 MHz, CDCl₃) 1.37 (t, *J* = 6.98 Hz, 3H), 4.06 (q, *J* = 7.0 Hz, 2H), 4.35 (s, 2H), 4.61 (br s, 1H), 5.11 (s, 2H), 6.61-6.64 (m, 2H), 7.02 (d, *J* = 1.9 Hz, 1H), 7.15 (d, *J* = 1.9 Hz, 1H), 7.37 (dd, *J* = 8.3 Hz, *J* = 2.1 Hz, 1H), 7.49 (d, *J* = 2.1 Hz, 1H), 7.66-7.69 (m, 2H), 7.72 ppm (d, *J* = 8.4 Hz, 1H)

¹³C NMR (125 MHz, CDCl₃) 13.66 (CH₃), 45.73, 64.29, 70.39 (CH₂), 111.36 (2 CH), 117.21, 120.49 (C), 122.29, 122.50, 126.82, 128.45 (CH), 129.04 (2 CH), 131.07, 133.46, 133.95, 134.07, 137.62, 143.62, 151.80, 153.02, 171.39 ppm (C)

UPLC-MS: Method B

RT: 2.155 min; Purity: 98 %

MS: 524.9 [M+H]⁺

4-((3-Bromo-4-((2,4-dichlorobenzyl)oxy)-5-ethoxybenzyl)amino)benzamide (47)

Chemical Formula: C₂₃H₁₉BrCl₂N₂O₂; MW: 506.22

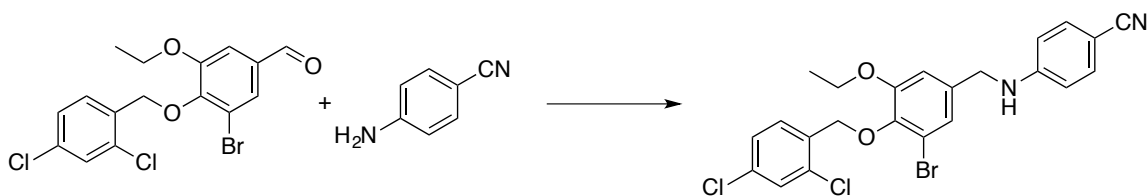


Table 44: Chemicals for the preparation of 47

Name	Eq.	mmol	MW	g	(d) g/cm ³	V (mL)
3-bromo-4-((2,4-dichlorobenzyl)oxy)-5-ethoxybenzaldehyde	1	0.247	404.08	0.100		
4-aminobenzonitrile	1	0.247	118.14	0.0292		
CH ₃ COOH	1	0.247	60.05	0.0148	1.049	0.014
DCE (solvent)			98.95			2
NaBH(OAc) ₃	2	0.494	211.94	0.1047		

Procedure: General procedure 2

Purification: Biotage Isolera One

Cartridge: SNAP KP-SIL 50 g

Elution system: *n*-hexane – EtOAc (3 CV 100:0 (v/v); 10 CV linear gradient reaching 40:60 (v/v); 3 CV 40:60 (v/v)). The product was eluted at 9.4-10.8 CV.

Outcome: Yield: 71.3 mg, 57%, yellowish solid

Analysis:

¹H NMR (500 MHz, CDCl₃) 1.30 (t, *J* = 7.0 Hz, 3H), 3.94 (q, *J* = 7.0 Hz, 2H), 4.21 (d, *J* = 5.5 Hz, 2H), 4.67 (t, *J* = 4.9 Hz, 1H), 5.03 (s, 2H), 6.49 (d, *J* = 8.6 Hz, 2H), 6.73 (s, 1H), 7.01 (s, 1H), 7.20 (d, *J* = 8.3 Hz, 1H), 7.32 (m, 3H), 7.66 ppm (d, *J* = 8.3 Hz, 1H)

¹³C NMR (125 MHz, CDCl₃) 14.72 (CH₃), 46.97, 64.97, 70.87 (CH₂), 99.76 (C), 111.49 (CH), 112.56 (2 CH), 118.19, 120.17 (C), 123.16, 127.14, 128.91, 130.59 (CH), 133.22 (C), 133.8 (2 CH), 133.94, 134.05, 135.35, 150.78, 153.31 ppm (C).

UPLC-MS: Method B

RT: 2.37 min; Purity: 97.7%

MS: 158.9 [C₇H₅Cl₂]⁺

4-((3-Bromo-4-((2,4-dichlorobenzyl)oxy)-5-ethoxybenzyl)amino)benzoic acid (48)

Chemical Formula: C₂₃H₂₀BrCl₂NO₄; MW: 525.22; CAS-Nr: 1243526-11-5

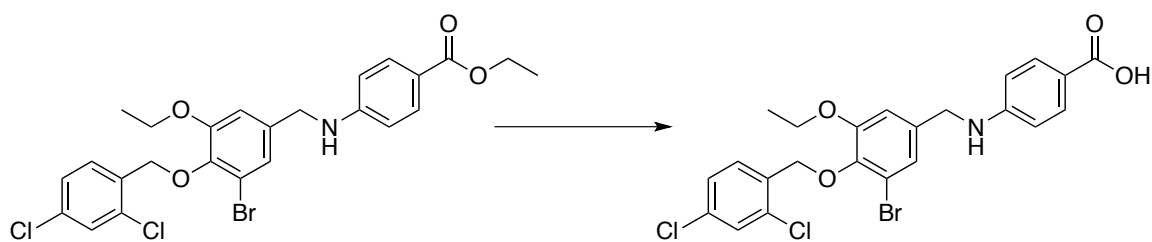


Table 45: Chemicals for the preparation of 48

Name	Eq.	mmol	MW	g	(d) g/cm ³	V (mL)
10	1	0.0904	553.27	0.050		
LiOH	3	0.27	23.95	0.006		

Procedure:

3 equivalents of lithium hydroxide were added to a stirring mixture of Ethyl 4-((3-bromo-4-((2,4-dichlorobenzyl)oxy)-5-ethoxybenzyl)amino)benzoate (10) in 5 mL of THF/MeOH/Water (4:2:2) and heated to 40°C for 24h. The reaction was monitored with TLC. Then the reaction was heated to 50°C and other 3 equivalents of LiOH were added and stirred for 6h. Then the reaction was heated to 70°C and reached completion after 12h.

Work up:

The reaction was stopped and cooled down to 25°C. A yellow-white solid was formed. Water was added to dilute the reaction mixture. The reaction was extracted once with 10 mL of DCM. Then the water phase was acidified with 3M HCl. The solid dissolved and the aqueous phase was extracted three times with DCM.

The organic layer was dried over Na₂SO₄ and evaporated under reduced pressure. A white solid was obtained which poorly dissolved in Chloroform. The NMR analysis was therefore carried out in DMSO.

TLC: *n*-hexane-Ethylacetate 7:3 (v/v). Spot remains on the baseline indicating the free acid. No more starting material can be seen on the TLC.

Outcome: Yield: 40.5 mg, 85.5%, off-white solid

Analysis:

¹H NMR (500 MHz, DMSO) 1.31 (t, *J* = 6.9 Hz, 3H), 4.05 (q, *J* = 6.8 Hz, 2H), 4.29 (d, *J* = 6.2 Hz, 2H), 5.05 (s, 2H), 6.60 (d, *J* = 8.7 Hz, 2H), 7.0 (t, *J* = 5.6 Hz, 1H), 7.09 (s, 1H), 7.13 (s, 1H), 7.49 (d, *J* = 8.3 Hz, 1H), 7.65-7.69 (m, 4H), 12.01 ppm (s, 1H).

¹³C NMR (125 MHz, CDCl₃) 14.97 (CH₃), 45.61, 64.75, 70.77 (CH₂), 111.77, 112.84 (CH), 117.21, 118.01 (C), 122.82 (CH), 127.89 (2 CH), 129.11 (CH), 131.57 (2 CH), 132.29 (CH), 133.84, 133.95, 134.25, 138.11, 143.42, 152.60, 152.96, 167.92 ppm (C).

UPLC-MS: Method B

RT: 2.257 min; Purity: 100%

MS: 525.09, [M+H]⁺

4-((2-Bromo-6-ethoxy-4-formylphenoxy)methyl)benzonitrile (49)

Chemical Formula: C₁₇H₁₄BrNO₃; MW: 360.21; CAS-Nr: 928708-50-3

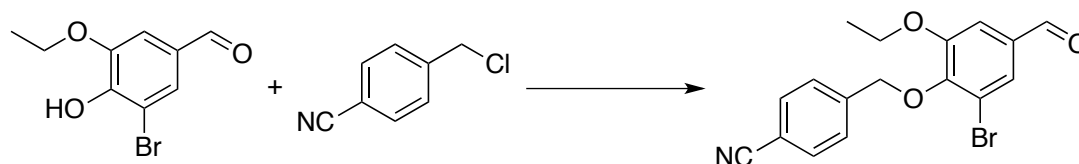


Table 46: Chemicals for the preparation of 49

Name	Eq.	mmol	MW	g	(d)	V
3-bromo-4-hydroxy-5-ethoxybenzaldehyde	1	0.816	245.07	0.200		
4-Chloromethyl benzonitrile	1	0.816	151.59	0.124		
K ₂ CO ₃	1.5	1.224	138.2	0.169		
DMF anhydrous					944 kg/m ³	3 mL

Procedure: General procedure 1

Purification: The compound was forming a solid when the reaction was poured into water and was therefore filtrated under vacuum to yield the clean product.

TLC: *n*-hexane-Ethyl acetate 7:3 (v/v) R_f = 0.42

Outcome: Yield: 141.5 mg, 48%, white solid

Analysis:

¹H NMR (500 MHz, CDCl₃) 1.46 (t, *J* = 7.0 Hz, 3H), 4.16 (q, *J* = 7.0 Hz, 2H), 5.22 (s, 2H), 7.39 (d, *J* = 1.78 Hz, 1H), 7.65 – 7.7 (m, 5H), 9.85 ppm (s, 1H)

¹³C NMR (125 MHz, CDCl₃) 14.67 (CH₃), 64.98, 73.64 (CH₂), 111.02 (CH), 112.01, 118.17, 118.71 (C), 128.37 (2 CH), 128.56 (CH), 132.25 (2 CH), 133.41, 142.09, 150.17, 153.31 (C), 189.76 ppm (CH)

4-((2-Ethoxy-4-formylphenoxy)methyl)benzonitrile (50)

Chemical Formula: C₁₇H₁₅NO₃; MW: 281.31; CAS-Nr: 928708-42-3

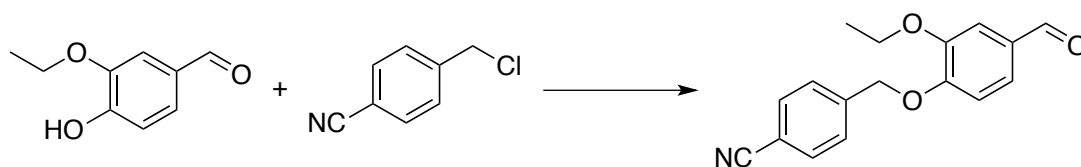


Table 47: Chemicals for the preparation of 50

Name	Eq.	mmol	MW	g	(d)	V
3-ethoxy -4-hydroxy benzaldehyde	1	2.407	166.18	0.400		
4-Chloromethyl benzonitrile	1	2.407	151.59	0.365		
K ₂ CO ₃	1.5	3.61	138.2	0.499		
DMF anhydrous					944 kg/m ³	8 mL

Procedure: General procedure 1

Purification: The reaction was stopped by pouring the mixture into water and the product formed a precipitate, that was vacuum filtrated and washed with water to remove the remains of DMF. The compound was dried with the rotary evaporator and under high vacuum.

TLC: *n*-hexane-Ethyl acetate 7:3 (v/v) R_f = 0.29

Outcome: Yield: 554.7 mg, 82%, white solid

Analysis:

¹H NMR (500 MHz, CDCl₃) 1.50 (t, *J* = 7.0 Hz, 3H), 4.19 (q, *J* = 7.0 Hz, 2H), 5.28 (s, 2H), 6.95 (d, *J* = 8.2 Hz, 1H), 7.39 (dd, *J* = 8.2 Hz, *J* = 1.9 Hz, 1H), 7.44 (d, *J* = 1.85 Hz, 1H), 7.57 (dd, *J* = 8.0 Hz, *J* = 0.6 Hz, 2H), 7.69 (dd, *J* = 8.4 Hz, *J* = 1.8 Hz, 2H), 9.85 ppm (s, 1H)

¹³C NMR (125 MHz, CDCl₃) 14.71 (CH₃), 64.60, 69.84 (CH₂), 110.99 (CH), 111.97 (C), 113.02 (CH), 118.58 (C), 126.09 (CH), 127.30 (2 CH), 130.95 (C), 132.51 (2 CH), 141.77, 149.57, 153.11 (C), 190.85 ppm (CH)

4-((4-Formylphenoxy)methyl)benzonitrile (51)

Chemical Formula: C₁₅H₁₁NO₂; MW: 237.26; CAS-Nr: 109702-08-1; (Amombo et al. 2012)

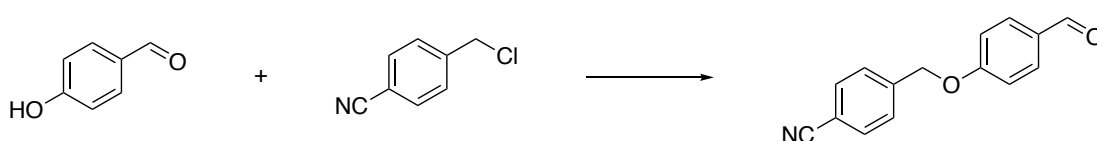


Table 48: Chemicals for the preparation of 51

Name	Eq.	mmol	MW	g	(d)	V
4-hydroxy benzaldehyde	1	2.456	122.12	0.300		
4-Chloromethyl benzonitrile	1	2.456	151.59	0.372		
K ₂ CO ₃	1.5	3.684	138.2	0.509		
DMF						5 mL

Procedure: General procedure 1

Purification: No purification needed after work up.

Outcome: Yield: 453.6 mg, 77.8%, white solid

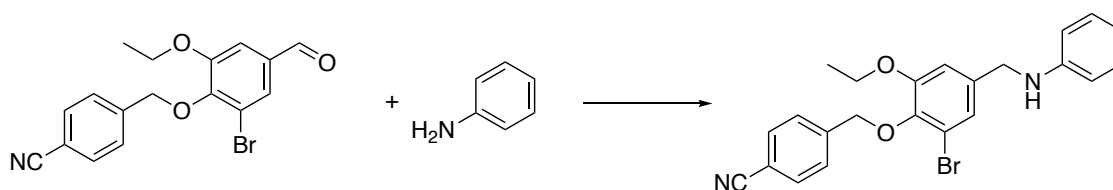
Analysis:

¹H NMR (500 MHz, CDCl₃) 5.22 (s, 2H), 7.05-7.08 (m, 2H), 7.55-7.57 (m, 2H), 7.69-7.72 (m, 2H), 7.84-7.87 (m, 2H), 9.90 ppm (s, 1H)

¹³C NMR (125 MHz, CDCl₃) 69.10 (CH₂), 112.17 (C), 115.09 (2 CH), 118.51 (C), 127.61 (2 CH), 130.60 (C), 132.08, 132.56 (2 CH), 141.35, 163.01 (C), 190.67 ppm (CH)

4-((2-Bromo-6-ethoxy-4-((phenylamino)methyl)phenoxy)methyl)benzonitrile (52)

Chemical Formula: C₂₃H₂₁BrN₂O₂; MW: 437.34

**Table 49: Chemicals for the preparation of 52**

Name	Eq.	mmol	MW	g	(d)	V
4-((2-bromo-6-ethoxy-4-formylphenoxy)methyl)benzonitrile	1	0.278	360.21	0.100		
Aniline	1	0.278	93.13	0.026	1.021	0.254*
CH ₃ COOH	1.5	0.417	60.05	0.025	1.049	0.023

NaBH(OAc) ₃	2	0.556	211.94	0.118		
DCE anhydrous					944 kg/m ³	3 mL

*Aniline dilution 0.1 mL aniline + 0.9 mL DCE

Procedure: General procedure 2

Purification: Biotage Isolera One

Cartridge: ZIP KP-SIL 30g

Elution system: *n*-hexane – EtOAc: 5CV 0% EtOAc, 10 CV linear gradient reaching 60% EtOAc, 2 CV 60% EtOAc. Eluted at 10-12 CV

TLC: *n*-hexane-Ethyl acetate 7:3 R_f: 0.52

Outcome: Yield: 45.3mg, 37.26%, beige solid

Analysis:

¹H NMR (500 MHz, CDCl₃) 1.42 (t, *J* = 7.0 Hz, 3H), 4.07 (q, *J* = 7.0 Hz, 3H), 4.29 (s, 2H), 5.1 (s, 2H), 6.65 (d, *J* = 7.7 Hz, 2H), 6.77 (t, *J* = 7.3 Hz, 1H), 6.92 (d, *J* = 1.7 Hz, 1H), 7.19 (d, *J* = 1.5 Hz, 1H), 7.21 (t, *J* = 7.9 Hz, 2H), 7.70 ppm (s, 3H)

¹³C NMR (125 MHz, CDCl₃) 14.78 (CH₃), 47.80, 64.64, 73.49 (CH₂), 111.67 (C), 111.70, 112.98 (CH), 117.85 (C), 118.00 (CH), 118.88 (C), 123.27, 128.43, 129.32, 132.15 (CH), 137.32, 142.81, 144.07, 147.80, 153.03 ppm (C)

UPLC-MS: Method B

RT: 2.2 min; Purity: 98 %

MS: 437.2, [M+H]⁺ (one Br in structure, characteristic split observed)

Ethyl 4-((4-((4-cyanobenzyl)oxy)-3-ethoxybenzyl)amino)benzoate (53)

Chemical Formula: C₂₆H₂₆N₂O₄; MW: 430.50

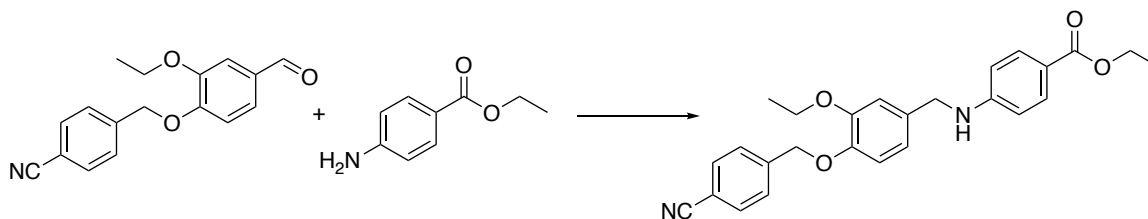


Table 50: Chemicals for the preparation of 53

Name	Eq.	mmol	MW	g	(d)	V
4-((2-ethoxy-4-formylphenoxy)methyl)benzonitrile	1	0.711	281.31	0.200		
Ethyl 4-aminobenzoate	1	0.711	165.19	0.117		

CH ₃ COOH	1.5	1.066	60.05	0.064	1.049	0.061
NaBH(OAc) ₃	2	1.421	211.94	0.301		
DCE anhydrous					944 kg/m ³	3 mL

Procedure: General procedure 2

Purification: Biotage Isolera One

Cartridge: SNAP KP-SIL 50g

Elution system: *n*-hexane – EtOAc (5 CV 93% (v/v) *n*-hexane, 10 CV linear gradient reaching 40:60 (v/v) *n*-hexane : EtOAc) Eluted at 12.5 CV

TLC: *n*-hexane-Ethyl acetate 7:3, R_f = 0.27

Outcome: Yield: 78 mg, 25.48%, off-white solid

Analysis:

¹H NMR (500 MHz, CDCl₃) 1.38 (t, *J* = 7.1 Hz, 3H), 1.47 (t, *J* = 7.0 Hz, 3H), 4.11 (q, *J* = 7.0 Hz, 2H), 4.32-4.36 (m, 4H), 4.43 (t, *J* = 5.3 Hz, 1H), 5.2 (s, 2H), 6.60-6.62 (m, 2H), 6.85 (d, *J* = 1.0 Hz, 2H), 6.94 (s, 1H), 7.59 (d, *J* = 8.6, 2H), 7.68-7.70 (m, 2H), 7.88-7.91 ppm (m, 2H).

¹³C NMR (125 MHz, CDCl₃) 14.46, 14.88 (CH₃), 47.52, 60.24, 64.55, 70.56 (CH₂), 111.59 (CH), 111.66 (C), 112.85, 115.35 (CH), 118.76, 119.21 (C), 119.66, 127.44, 131.50, 132.35 (CH), 132.54, 142.93, 147.30, 149.59, 151.61, 166.79 ppm (C).

UPLC-MS: Method B

RT: 2.065 min; Purity: 100 %

MS: 453.3 [M+Na]⁺

4-((2-Ethoxy-4-((phenylamino)methyl)phenoxy)methyl)benzonitrile (54)

Chemical formula: C₂₃H₂₂N₂O₂; MW: 358.44

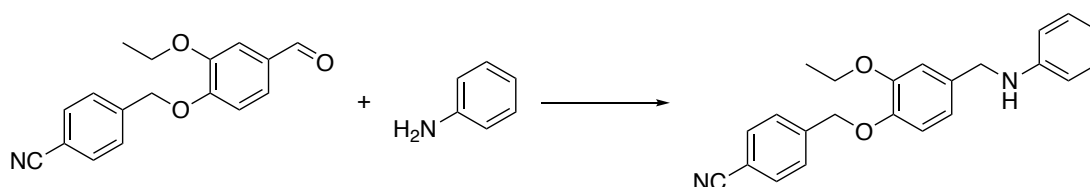


Table 51: Chemicals for the preparation of 54

Name	Eq.	mmol	MW	g	(d)	V
4-((2-ethoxy-4-formylphenoxy)methyl)benzonitrile	1	0.711	281.31	0.200		

Aniline	1	0.711	93.13	0.066	1.021	0.648*
CH ₃ COOH	1.5	1.066	60.05	0.025	1.049	0.023
NaBH(OAc) ₃	2	1.412	211.94	0.118		
DCE anhydrous					944 kg/m ³	3 mL

*Aniline dilution 0.1 mL aniline + 0.9 mL DCE

Procedure: General procedure 2

Purification: Biotage Isolera One

Cartridge: SNAP KP-SIL 50g

Elution system: *n*-hexane – EtOAc (3 CV 93% (v/v) *n*-hexane, 10 CV linear gradient reaching 40:60 (v/v) *n*-hexane : EtOAc, 2 CV 40:60 (v/v) *n*-hexane : EtOAc)) Eluted at 8.5 CV

TLC: *n*-hexane : Ethyl acetate 7:3; R_f = 0.53

Outcome: Yield: 153.8 mg, 60.3%, light yellow solid

Analysis:

¹H NMR (500 MHz, CDCl₃) 1.47 (t, *J* = 7.9 Hz, 3H), 3.99 (s, 1H), 4.12 (q, *J* = 7.0 Hz, 2H), 4.28 (s, 2H), 5.2 (s, 2H), 6.66 (dd, *J* = 7.7 Hz, *J* = 1.0 Hz, 2H), 6.75 (tt, *J* = 7.3 Hz, *J* = 1.0 Hz, 1H), 6.84-6.89 (m, 2H), 6.98 (d, *J* = 1.8 Hz, 1H), 7.2 (dd, *J* = 8.6 Hz, *J* = 7.4 Hz, 2H), 7.58-7.60 (m, 2H), 7.68-7.70 ppm (m, 2H).

¹³C NMR (125 MHz, CDCl₃) 14.90 (CH₃) 48.17, 64.49, 70.60 (CH₂), 111.55 (C), 112.90 (2 CH), 112.99, 115.34, 117.69 (CH), 118.79 (C), 119.69 (CH), 127.45, 129.27, 132.33 (3x2 CH), 133.66, 143.04, 147.07, 148.07, 148.14, 149.50 ppm (C).

UPLC-MS: Method B

RT: 2.032 min; Purity: 100 %

MS: 226.1 [C₁₇H₁₇NO₂]⁻ negative mode

Ethyl 4-((4-((4-cyanobenzyl)oxy)benzyl)amino)benzoate (55)

Chemical Formula: C₂₄H₂₂N₂O₃; MW: 386.45

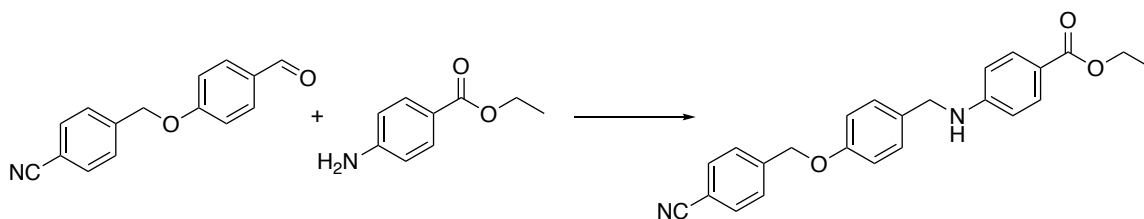


Table 52: Chemicals for the preparation of 55

Name	Eq.	mmol	MW	g	(d) g/cm ³	V (mL)
4-((4-formylphenoxy)methyl)benzonitrile	1	0.6322	237.26	0.150		
Ethyl-4-amino benzoate	1	0.6322	165.19	0.104		
CH ₃ COOH	1.5	0.9483	60.05	0.057	1.049	0.054
NaBH(OAc) ₃	2	1.2644	211.94	0.268		
DCE (dry) solvent						4

Procedure: General procedure 2

Purification: Crystallisation in EtOAc and filtration

Outcome: Yield: 134.4 mg, 55%, white needles

Analysis:

¹H NMR (500 MHz, CDCl₃) 1.38 (t, *J* = 7.1 Hz, 3H), 4.31-4.38 (m, 4H), 4.44 (s, 1H), 5.15 (s, 2H), 6.60-6.62 (m, 2H), 6.94-6.97 (m, 2H), 7.29-7.32 (m, 2H), 7.56-7.59 (m, 2H), 7.69-7.73 (m, 2H), 7.88-7.91 ppm (m, 2H)

¹³C NMR (125 MHz, (CD₃)₂SO) 14.82 (CH₃), 45.80, 59.95, 68.78 (CH₂), 110.93 (C), 111.74, 115.25 (2 CH), 116.83, 119.18 (C), 128.45, 129.01, 131.30 (3x2 CH), 132.24 (C), 132.85 (2 CH), 143.52, 153.10, 157.45, 166.30 ppm (C)

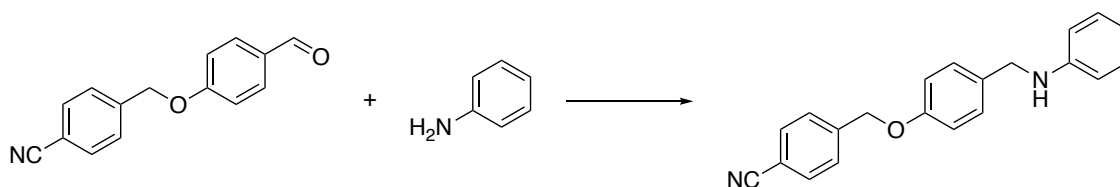
UPLC-MS: Method B

RT: 2.070 min; Purity: 100 %

MS: 409.3 [M+Na]⁺

4-((4-((Phenylamino)methyl)phenoxy)methyl)benzonitrile (56)

Chemical Formula: C₂₁H₁₈N₂O; MW: 314.39

**Table 53: Chemicals for the preparation of 56**

Name	Eq.	mmol	MW	g	(d)	V
4-((4-formylphenoxy)methyl)benzonitrile	1	0.6322	237.26	0.150		

methyl)benzotrile						
Aniline*	1	0.6322	93.13	0.059	1.021	0.577*
CH ₃ COOH	1.5	0.9483	60.05	0.057	1.049	0.054
NaBH(OAc) ₃	2	1.2644	211.94	0.268		
DCE						5 mL

*Aniline dilution in DCE 0.1ml Aniline + 0.9ml DCE

Procedure: General procedure 2

Purification:

The organic phase contained the clean product. No further purification was carried out.

Outcome: Yield: 145.8 mg; 73.4%; mustard shiny solid

Analysis:

¹H NMR (500 MHz, CDCl₃) 4.02 (bs, 1H), 4.30 (s, 2H), 5.15 (s, 2H), 6.65-6.68 (m, 2H), 6.75 (tt, *J* = 7.3 Hz, *J* = 1.0 Hz, 1H), 6.94-6.97 (m, 2H), 7.18-7.22 (m, 2H), 7.32-7.35 (m, 2H), 7.57 (dd, *J* = 8.0 Hz, *J* = 0.6 Hz, 2H), 7.70-7.71 ppm (m, 2H)

¹³C NMR (125 MHz, CDCl₃) 47.74, 69.00 (CH₂) 111.75 (C), 112.90, 114.96 (2 CH), 117.65 (CH), 118.69 (C), 127.55, 128.91, 129.28 (2 CH), 132.39 (C), 132.43 (2 CH), 142.55, 148.08, 157.46 ppm (C)

UPLC-MS: Method B

RT: 2.012 min; Purity: 100 %

MS: 315.20 [M+H]⁺

4-((2,4-Dichlorobenzyl)oxy)-3-fluorobenzaldehyde (57)

Chemical Formula: C₁₄H₉Cl₂FO₂; MW: 299.12; CAS-Nr: 1962269-13-1

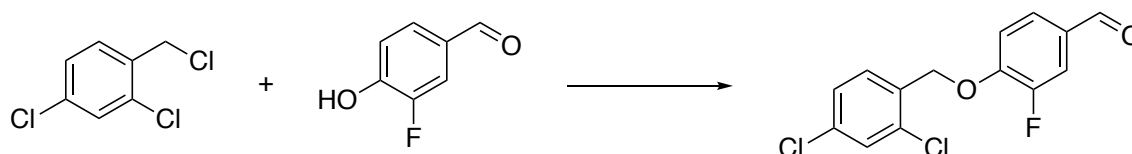


Table 54: Chemicals for the preparation of 57

Name	Eq.	mmol	MW	g	(d) g/cm ³	V (mL)
2,4-Dichloro benzylchloride	1	1.427	195.47	0.279	1.407	0.198
3-Fluoro-4-hydroxy benzaldehyde	1	1.427	140.11	0.200		
K ₂ CO ₃	1.5	2.140	138.2	0.295		
DMF (solvent)						5

Procedure: General procedure 1

Outcome: Yield: 310.3 mg, 72,7%, white powder

Analysis:

^1H NMR (500 MHz, CDCl_3) 5.28 (s, 2H), 7.12 (t, $J = 8.2$ Hz, 1H), 7.31 (dd, $J = 8.3$ Hz, $J = 2.1$ Hz, 1H), 7.45 (d, $J = 2.1$ Hz, 1H), 7.52 (d, $J = 8.4$ Hz, 1H), 7.62-7.67 (m, 2H), 9.88 ppm (d, $J = 2.13$ Hz, 1H)

^{13}C NMR (125 MHz, CDCl_3) 67.79 (CH_2), 114.553 ($J_{\text{CF}} = 1.68$ Hz, CH), 116.109 ($J_{\text{CF}} = 18.94$ Hz, CH), 127.57 (CH), 127.8655 ($J_{\text{CF}} = 2.87$ Hz, CH), 129.43, 129.60 (CH), 130.885 ($J_{\text{CF}} = 5.01$ Hz, C), 131.93, 133.25, 134.86, 151.5855 ($J_{\text{CF}} = 10.95$ Hz), 153.87 (C), 189.629 ppm ($J_{\text{CF}} = 1.86$ Hz, C=O)

^{19}F NMR (470 MHz, CDCl_3) -131.97 (F_{arom}) ppm

4-((2,4-Dichlorobenzyl)oxy)-3-fluoro-5-methoxybenzaldehyde (58)

Chemical Formula: $\text{C}_{15}\text{H}_{11}\text{Cl}_2\text{FO}_3$; MW: 329.148

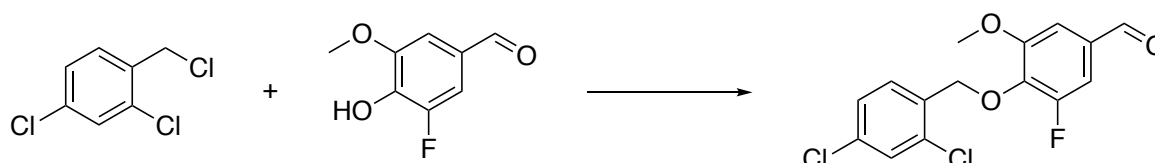


Table 55: Chemicals for the reaction of 58

Name	Eq.	mmol	MW	g	(d) g/cm ³	V (mL)
2,4-Dichloro benzylchloride	1	1.175	195.47	0.2297	1.407	0.163
3-Fluoro-4-hydroxy benzaldehyde	1	1.175	170.14	0.200		
K_2CO_3	1.5	1.763	138.2	0.2436		
DMF (solvent)						5

Procedure: General procedure 1

Outcome: Yield: 257.1 mg, 66.5 %, pale yellow solid

Analysis:

^1H NMR (500 MHz, CDCl_3) 3.94 (s, 3H), 5.29 (s, 2H), 7.24-7.30 (m, 4H incl. $\text{CDCl}_3 = 3\text{H}$), 7.40 (d, $J = 2.1$ Hz, 1H), 7.60 (d, $J = 8.4$ Hz, 1H), 9.85 ppm (d, $J = 1.25$ Hz, 1H)

^{13}C NMR (125 MHz, CDCl_3) 56.50 (CH_3), 71.805 ($J_{\text{CF}} = 3.93$ Hz, CH_2), 107.35 ($J_{\text{CF}} = 2.51$ Hz, CH), 111.865 ($J_{\text{CF}} = 20.76$ Hz, CH), 127.22, 129.17, 130.37 (CH), 131.8155 ($J_{\text{CF}} = 6.57$ Hz, C), 133.30, 133.63, 134.54, 141.25, 154.94, 156.92 (C), 189.76 ppm ($J_{\text{CF}} = 2.18$ Hz, CH)

^{19}F NMR (470 MHz, CDCl_3) -127.595 (F_{arom}) ppm

3-Chloro-4-((2,4-dichlorobenzyl)oxy)benzaldehyde (59)

Chemical Formula: $\text{C}_{14}\text{H}_9\text{Cl}_3\text{O}_2$; MW: 315.574; CAS-Nr: 433242-78-5

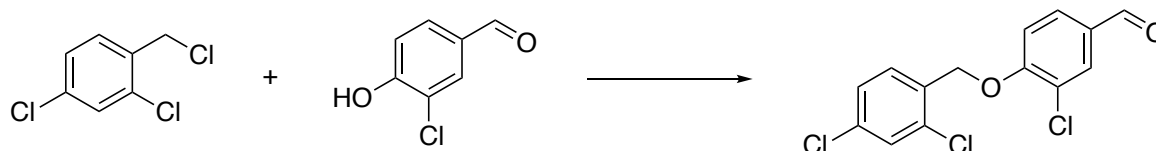


Table 56: Chemicals for the synthesis of 59

Name	Eq.	mmol	MW	g	(d) g/cm^3	V (mL)
2,4-Dichloro benzylchloride	1	1.277	195.47	0.249	1.407	0.177
3-chloro-4-hydroxybenzaldehyde	1	1.277	156.57	0.200		
K_2CO_3	1.5	1.916	138.2	0.264		
DMF (solvent)						4

Procedure: General procedure 1

Outcome: Yield: 214 mg, 53.1 %, white solid

Analysis:

^1H NMR (500 MHz, CDCl_3) 5.28 (s, 2H), 7.09 (d, $J = 8.5$ Hz, 1H), 7.32 (dd, $J = 8.3$ Hz, $J = 2.1$ Hz, 1H), 7.44 (d, $J = 2.1$ Hz, 1H), 7.60 (d, $J = 8.4$ Hz, 1H), 7.77 (dd, $J = 8.5$ Hz, $J = 2.0$ Hz, 1H), 7.95 (d, $J = 2.0$ Hz, 1H), 9.87 ppm (s, 1H)

^{13}C NMR (125 MHz, CDCl_3) 68.11 (CH_2), 114.67 (CH), 122.90 (C), 128.18, 129.55 (CH), 131.02 (C), 131.12, 131.24, 131.69 (CH), 132.97, 134.11, 134.40, 158.34 (C), 191.16 ppm (CH)

3-Chloro-4-((2,4-dichlorobenzyl)oxy)-5-ethoxybenzaldehyde (60)

Chemical Formula: $\text{C}_{16}\text{H}_{13}\text{Cl}_3\text{O}_3$; MW: 359.627; CAS-Nr: 428499-14-3

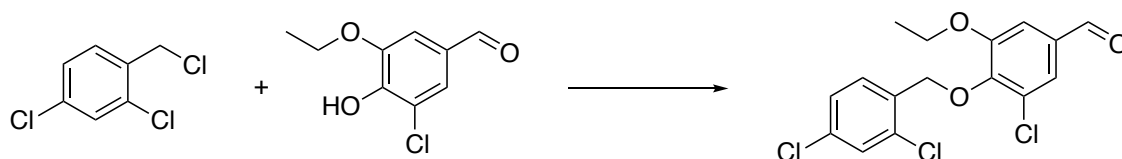


Table 57: Chemicals for the synthesis of 60

Name	Eq.	mmol	MW	g	(d) g/cm ³	V (mL)
2,4-Dichloro benzylchloride	1	0.997	195.47	0.195	1.407	0.138
3-chloro-5-ethoxy-4-hydroxybenzaldehyde	1	0.997	200.62	0.200		
K ₂ CO ₃	1.5	1.495	138.2	0.206		
DMF (solvent)						6

Procedure: General procedure 1

Outcome: Yield: 68 mg, 19.2 %, off-white solid

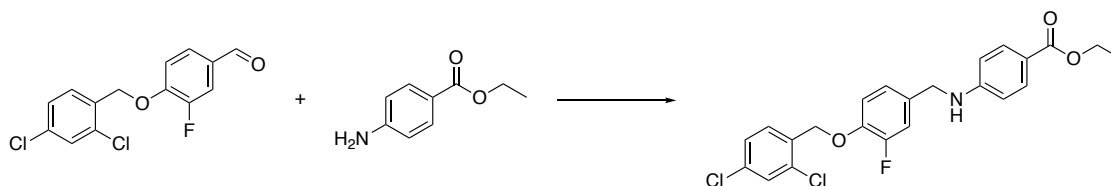
Analysis:

¹H NMR (500 MHz, CDCl₃) 1.47 (t, *J* = 7.0 Hz, 3H), 4.18 (q, *J* = 7.0 Hz, 2H), 5.29 (s, 2H), 7.33 (dd, *J* = 8.3 Hz, *J* = 2.1 Hz, 1H), 7.38 (d, *J* = 1.8 Hz, 1H), 7.44 (d, *J* = 2.1 Hz, 1H), 7.53 (d, *J* = 1.8 Hz, 1H), 7.72 (d, *J* = 8.3 Hz, 1H), 9.88 ppm (s, 1H)

¹³C NMR (125 MHz, CDCl₃) 14.88 (CH₃), 65.22, 71.27 (CH₂), 112.31, 124.34, 127.95 (CH), 128.32 (C), 129.24, 132.48 (CH), 133.21, 133.86, 134.20, 134.27, 148.56, 153.55 (C), 191.55 ppm (CH)

Ethyl 4-((4-((2,4-dichlorobenzyl)oxy)-3-fluorobenzyl)amino)benzoate (61)

Chemical Formula: C₂₃H₂₀Cl₂FNO₃; MW: 448.315

**Table 58: Chemicals for the synthesis of 61**

Name	Eq.	mmol	MW	g	(d) g/cm ³	V (mL)
4-((2,4-dichlorobenzyl)oxy)-3-fluorobenzaldehyde	1	0.668	299.12	0.200		
Ethyl-4-amino benzoate	1	0.668	165.19	0.110		
CH ₃ COOH	1	0.668	60.05	0.040	1.049	0.038
NaBH(OAc) ₃	2	1.337	211.94	0.283		
DCE (dry) solvent						5

Procedure: General procedure 2

Purification: Biotage Isolera One

Cartridge: SNAP KP-SIL 25g

Elution system: *n*-hexane – EtOAc (5 CV 93% (v/v) *n*-hexane, 10 CV linear gradient reaching 40:60 (v/v) *n*-hexane : EtOAc, 5 CV 40:60 (v/v) *n*-hexane : EtOAc) Eluted at 8-10 CV

Outcome: Yield: 174.1 mg, 58 %, off white solid

Analysis:

¹H NMR (500 MHz, CDCl₃) 1.38 (t, *J* = 7.1 Hz, 3H), 4.32-4.37 (m, 4H), 4.49 (t, *J* = 5.5 Hz, 1H), 5.20 (s, 2H), 6.58-6.61 (m, 2H), 7.00 (t, *J* = 8.3 Hz, 1H), 7.1 (dd, *J* = 8.3 Hz, *J* = 1.1 Hz, 1H), 7.14 (dd, *J* = 11.74 Hz, *J* = 2.1 Hz, 1H), 7.31 (dd, *J* = 8.3 Hz, *J* = 2.1 Hz, 1H), 7.45 (d, *J* = 2.1 Hz, 1H), 7.56 (d, *J* = 8.34 Hz, 1H), 7.88-7.91 ppm (m, 2H)

¹³C NMR (125 MHz, CDCl₃) 14.46 (CH₃), 46.79, 60.25, 68.03 (CH₂), 111.71 (2 CH), 115.4665 (*J*_{CF} = 19.0, CH), 115.909 (*J*_{CF} = 1.7 Hz, CH), 119.42 (C), 123.001 (*J*_{CF} = 3.5 Hz, CH), 127.43, 129.25, 129.64 (CH), 131.53 (2 CH), 132.69, 132.91, 133.13, 134.37, 151.34, 166.74 ppm (C)

¹⁹F NMR (470 MHz, CDCl₃) -132.76 ppm (F_{arom})

UPLC-MS: Method B

RT: 2.322 min; Purity: 100%

MS: 446.23 [M-H]⁻ negative mode

Ethyl 4-((4-((2,4-dichlorobenzyl)oxy)-3-fluoro-5-methoxybenzyl)amino)benzoate (62)

Chemical Formula: C₂₄H₂₂Cl₂FNO₄; MW: 478.341

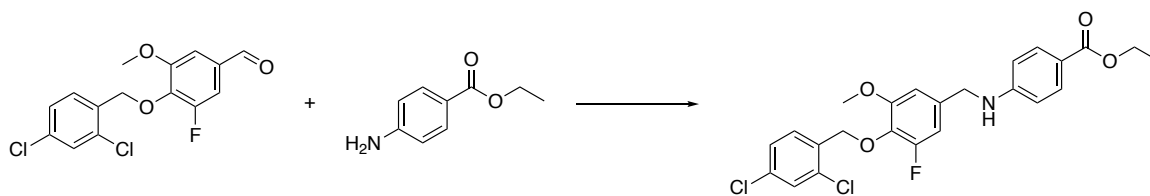


Table 59: Chemicals for the synthesis of 62

Name	Eq.	mmol	MW	g	(d) g/cm ³	V (mL)
4-((2,4-dichlorobenzyl)oxy)-3-fluoro-5-methoxybenzaldehyde	1	0.608	329.15	0.200		
Ethyl-4-amino benzoate (46)	1	0.608	165.19	0.100		
CH ₃ COOH	1	0.608	60.05	0.037	1.049	0.035

NaBH(OAc) ₃	2	1.215	211.94	0.258		
DCE (dry) solvent						5

Procedure: General procedure 2

Purification: Biotage Isolera One

Cartridge: SNAP KP-SIL 10g, Flowrate: 20 mL/min, 5 mL fraction volume

Elution system: DCM (solvent A) – MeOH 1% in DCM (solvent B) (1 CV 100% (v/v) solvent A, 5.1 CV linear gradient reaching 5% (v/v) solvent B, 5 CV reaching 10% (v/v) solvent B, 3 CV 10% solvent B. Eluted at 2.5-3.5 CV

Outcome: Yield: 35 mg; 12 %; white waxy solid

Analysis:

¹H NMR (500 MHz, CDCl₃) 1.28 (t, *J* = 7.1 Hz, 3H), 3.74 (s, 3H), 4.21-4.26 (m, 4H), 4.47 (t, *J* = 5.3 Hz, 1H), 5.06 (s, 2H), 6.48- 6.51 (m, 2H), 6.60 (s, 1H), 6.63 (dd, *J* = 10.6 Hz, *J* = 1.8 Hz, 1H), 7.19 (dd, *J* = 8.1 Hz, *J* = 1.8 Hz, 1H), 7.30 (d, *J* = 2.1 Hz, 1H), 7.55 (d, *J* = 8.3 Hz, 1H), 7.78-7.81 ppm (m, 2H)

¹³C NMR (125 MHz, CDCl₃) 14.46 (CH₃), 47.35 (CH₂), 56.29 (CH₃), 60.29, 71.77 (CH₂), 106.47 (CH) 107.58 (*J*_{CF} = 20.59 Hz, CH), 111.77 (2 CH), 119.48 (C), 127.13, 128.97, 130.38 (CH), 131.52 (2 CH), 133.43, 133.88, 134.12 (C), 134.95 (*J*_{CF} = 8.18 Hz, C), 151.38 (C), 154.18 (*J*_{CF} = 5.29 Hz, C), 155.27, 157.23, 166.76 ppm (C)

¹⁹F NMR (470 MHz, CDCl₃) –128.87 (F_{arom}) ppm

UPLC-MS: Method B

RT: 2.350, Purity 100 %

MS: 158.9 [C₇H₅Cl₂]⁺

Ethyl 4-((3-chloro-4-((2,4-dichlorobenzyl)oxy)benzyl)amino)benzoate (63)

Chemical Formula: C₂₃H₂₀Cl₃NO₃; MW: 464.767

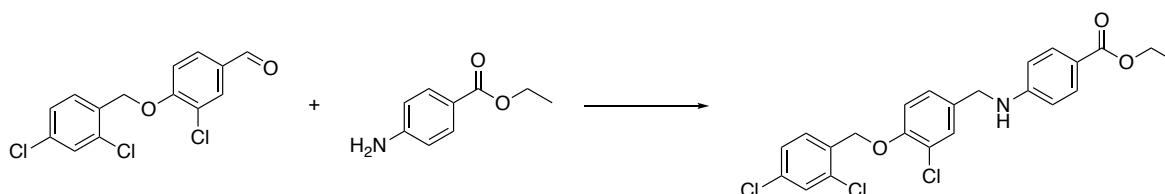


Table 60: Chemicals for the synthesis of 63

Name	Eq.	mmol	MW	g	(d) g/cm ³	V (mL)
3-chloro-4-((2,4-dichlorobenzyl)oxy) benzaldehyde	1	0.317	315.57	0.100		
Ethyl-4-amino benzoate	1	0.317	165.19	0.052		
CH ₃ COOH	1	0.317	60.05	0.019	1.049	0.018
NaBH(OAc) ₃	2	0.634	211.94	0.134		
DCE (dry) solvent						5

Procedure: General procedure 2

Purification: Crystallisation in DCM and filtration

Outcome: Yield: 23.9 mg, 16 %, white solid

Analysis:

¹H NMR (500 MHz, CDCl₃) 1.35 (t, *J* = 7.1 Hz, 3H), 4.29-4.33 (m, 4H), 4.46 (s, 1H), 5.18 (s, 2H), 6.56-6.58 (m, 2H), 6.93 (d, *J* = 8.4 Hz, 1H), 7.17 (d, *J* = 2.2 Hz, 1H), 7.3 (dd, *J* = 8.3 Hz, *J* = 2.1 Hz, 1H), 7.40 (d, *J* = 2.2 Hz, 1H), 7.42 (d, *J* = 2.0 Hz, 1H), 7.61 (d, *J* = 8.2 Hz, 1H), 7.86-7.88 ppm (m, 2H)

¹³C NMR (125 MHz, CDCl₃) 14.43 (CH₃), 46.74, 60.20, 67.54 (CH₂), 111.73 (2 CH), 114.20 (CH), 119.56, 123.72 (C), 126.64 (CH), 126.86 (C), 127.47, 129.15, 129.38, 129.42 (CH), 131.52 (2 CH), 132.60, 132.90, 151.33, 153.10, 165.25, 166.68 ppm (C)

UPLC-MS: Method B

RT: 2.382 min; Purity 100%

MS: 464.25 [M-H]⁻ negative mode

Ethyl 4-((3-chloro-4-((2,4-dichlorobenzyl)oxy)-5-ethoxybenzyl)amino)benzoate (64)

Chemical Formula: C₂₅H₂₄Cl₃NO₄; MW: 508.82; CAS-Nr: 1311001-22-5

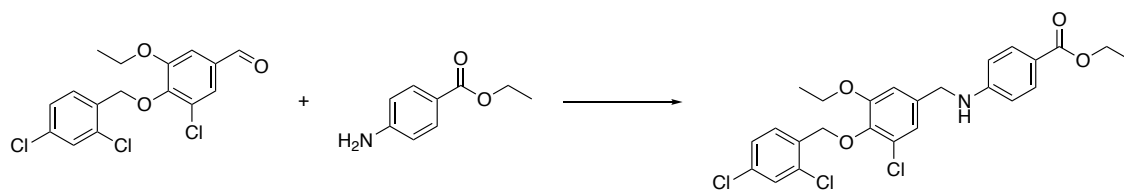


Table 61: Compounds for the synthesis of 64

Name	Eq.	mmol	MW	g	(d) g/cm ³	V (mL)
3-chloro-4-((2,4-dichlorobenzyl)oxy)-5-ethoxybenzaldehyde	1	0.139	359.63	0.050		
Ethyl-4-amino benzoate	1	0.139	165.19	0.023		
CH ₃ COOH	1	0.139	60.05	0.008	1.049	0.008
NaBH(OAc) ₃	2	0.278	211.94	0.059		
DCE (dry) solvent						3

Procedure: General procedure 2

Purification: Biotage Isolera One

Cartridge: SNAP KP-Sil 10g

Elution system: *n*-hexane – EtOAc (7 CV 100% (v/v) *n*-hexane, 20 CV linear gradient reaching 100% (v/v) EtOAc, 9 CV 100% (v/v) EtOAc) Eluted at 13.5-14.5 CV

Outcome: Yield: 14 mg, 20 %, white solid

Analysis:

¹H NMR (500 MHz, CDCl₃) 1.29 (2t overlaid *J* = 7.1 Hz, *J* = 7.0 Hz, 6H), 3.94 (q, *J* = 7.0 Hz, 2H), 4.22-4.26 (m, 4H), 4.45 (t, *J* = 5.6 Hz, 1H), 5.05 (s, 2H), 6.50 (d, *J* = 8.9 Hz, 2H), 6.72 (d, *J* = 2.0 Hz, 1H), 6.89 (d, *J* = 2.0 Hz, 1H), 7.22 (dd, *J* = 8.3 Hz, *J* = 2.1 Hz, 1H), 7.31 (d, *J* = 2.0, 1H), 7.65 (d, *J* = 8.3 Hz, 1H), 7.80 ppm (d, *J* = 8.9 Hz, 2H)

¹³C NMR (125 MHz, CDCl₃) 14.45, 14.72 (CH₃), 47.22, 60.28, 64.69, 70.99 (CH₂), 110.76 (CH), 111.76 (2 CH), 119.49 (C), 120.33, 127.11 (CH), 128.66 (C), 128.89, 130.54 (CH), 131.52 (2 CH), 133.27, 134.00, 134.04, 135.43, 143.40, 151.35, 153.37, 166.75 ppm (C)

UPLC-MS: Method B

RT: 2.500 min; Purity: 98 %

MS: 158.9 [C₇H₅Cl₂]⁺

Ethyl 6-((3-bromo-4-((2,4-dichlorobenzyl)oxy)-5-ethoxybenzyl)amino)nicotinate (65)

Chemical Formula: C₂₄H₂₃BrCl₂N₂O₄; MW: 554.262

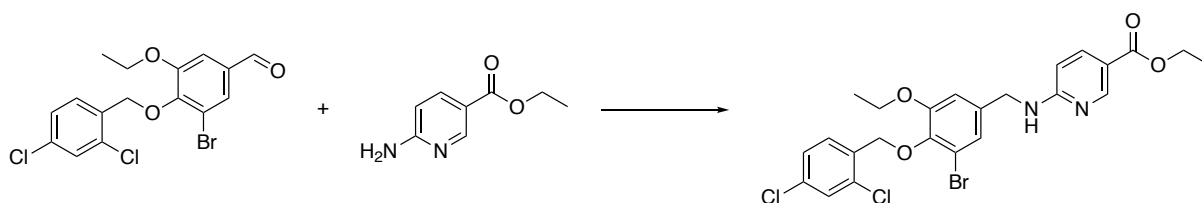


Table 62: Chemicals for the synthesis of 65

Name	Eq.	mmol	MW	g	(d) g/cm ³	V (mL)
3-bromo-4-((2,4-dichlorobenzyl)oxy)-5-ethoxybenzaldehyde	1	0.4949	404.08	0.200		
Ethyl 6-aminonicotinate	1.22	0.608	166.18	0.101		
TiCl(O ⁱ Pr) ₃	2.7	1.338	260.58	0.349	1.091	0.319
CH ₃ COOH			60.05			3 gtt
STAB	6.1	3.04	211.94	0.644		
DCM dry (solvent)						5

Procedure: General procedure 3

Purification: The compound was obtained by crystallization in *n*-hexane and filtration.

Outcome: Yield: 8 mg; 3 %; white shiny solid

Analysis:

¹H NMR (500 MHz, CDCl₃) 1.40 (2t overlaid, *J* = 7.1 Hz, *J* = 7.0 Hz, 6H), 4.05 (q, *J* = 7.0 Hz, 2H), 4.36 (q, *J* = 7.1 Hz, 2H), 4.55 (d, *J* = 5.9 Hz, 2H), 5.13 (s, 2H), 5.28 (s, 1H), 6.39 (dd, *J* = 8.8 Hz, *J* = 0.7 Hz, 1H), 6.87 (d, *J* = 1.9 Hz, 1H), 7.14 (d, *J* = 1.9 Hz, 1H), 7.32 (dd, *J* = 8.32 Hz, *J* = 2.1 Hz, 1H), 7.42 (d, *J* = 2.1 Hz, 1H), 7.78 (d, *J* = 8.3 Hz, 1H), 8.04 (dd, *J* = 8.7 Hz, *J* = 2.3 Hz, 1H), 8.81 ppm (d, *J* = 2.1 Hz, 1H)

¹³C NMR (125 MHz, CDCl₃) 14.40, 14.72 (CH₃), 45.36, 60.54, 64.70, 70.84 (CH₂), 106.04, 111.71 (CH), 116.30, 118.00 (C), 123.29, 127.12, 128.88, 130.59 (CH), 133.21, 133.99, 134.00, 135.94 (C), 138.64 (CH), 144.52 (C), 151.48 (CH), 153.20, 160.41, 165.89 ppm (C)

UPLC-MS: Method B

RT: 2.229 min, Purity 100 %

MS: 158.9 [C₇H₅Cl₂]⁺

Ethyl 2-((3-bromo-4-((2,4-dichlorobenzyl)oxy)-5-ethoxybenzyl)amino)pyrimidine-5-carboxylate (66)

Chemical Formula: C₂₃H₂₂BrCl₂N₃O₄; MW: 555.250

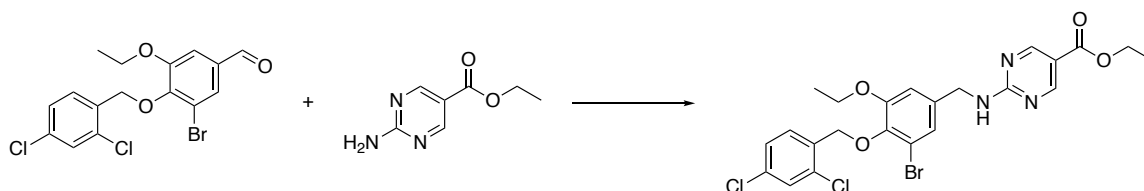


Table 63: Chemicals for the synthesis of 66

Name	Eq.	mmol	MW	g	(d) g/cm ³	V (mL)
3-bromo-4-((2,4-dichlorobenzyl)oxy)-5-ethoxybenzaldehyde	1	0.4949	404.08	0.200		
Ethyl 2-aminopyrimidine-5-carboxylate	1.1	0.544	167.17	0.091		
TiCl(O ^{<i>i</i>} Pr) ₃	2.2	1.089	260.58	0.284	1.091	0.26
CH ₃ COOH			60.05			3 gtt
STAB	5	2.4747	211.94	0.525		
DCM dry (solvent)						5

Procedure: General procedure 3

Purification: Biotage Isolera One

Cartridge: SNAP KP-SIL 10g, Flowrate: 25 mL/min, 5 mL fraction volume

Elution system: DCM (solvent A) – MeOH 1% in DCM (solvent B) (3 CV 100% (v/v) solvent A, 0.7 CV linear gradient reaching 1% (v/v) solvent B, 5.5 CV 1% (v/v) solvent B, 6 CV linear gradient reaching 13% solvent B, 3 CV 13% solvent B, 0.4 CV linear gradient 13-14%, solvent B, 2.9 CV 14% solvent B, 2.2 CV linear gradient 14-19% solvent B, 1.4 CV linear gradient 19-30% solvent B, 2 CV linear gradient 30-40% solvent B. Eluted at 6-8 CV

Outcome: Yield: 9.6 mg, 3.3%, white solid

Analysis:

¹H NMR (500 MHz, CDCl₃) 1.31 (2t overlaid, *J* = 7.1 Hz, *J* = 7.0 Hz, 6H), 3.96 (q, *J* = 7.0 Hz, 2H), 4.29 (q, *J* = 7.1 Hz, 2H), 4.57 (d, *J* = 6.2 Hz, 2H), 5.03 (s, 2H), 5.82 (t, *J* = 6.0 Hz, 1H), 6.77 (d, *J* = 1.9 Hz, 1H), 7.05 (d, *J* = 1.9 Hz, 1H), 7.23 (dd, *J* = 8.3 Hz, *J* = 2.1 Hz, 1H), 7.32 (q, *J* = 2.2 Hz, 1H), 7.68 (d, *J* = 8.3 Hz, 1H), 8.80 ppm (br d, *J* = 14.46 Hz, 2H)

¹³C NMR (125 MHz, CDCl₃) 14.35, 14.74 (CH₃), 44.88, 60.83, 64.71 70.85 (CH₂), 111.90 (CH), 114.62, 117.97 (C), 123.46, 127.12, 128.88, 130.59 (CH), 133.21, 133.97, 134.01, 135.74, 144.59, 153.13 (C), 160.29 (2 CH), 163.27, 164.58 ppm (C)

UPLC-MS: Method B

RT: 2.450 min; Purity: 96%

MS: 555.9 [M+H]⁺

Ethyl 5-((3-bromo-4-((2,4-dichlorobenzyl)oxy)-5-ethoxybenzyl)amino)pyrazine-2-carboxylate (67)

Chemical Formula: C₂₃H₂₂BrCl₂N₃O₄; MW: 555.25

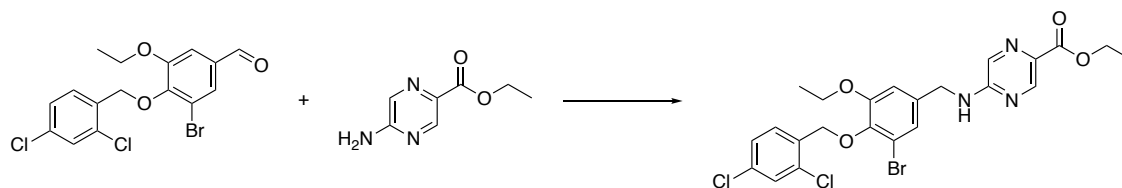


Table 64: Chemicals for the synthesis of 67

Name	Eq.	mmol	MW	g	(d) g/cm ³	V (mL)
3-bromo-4-((2,4-dichlorobenzyl)oxy)-5-ethoxybenzaldehyde	1	0.4949	404.08	0.200		
Ethyl-5-aminopyrazine-2-carboxylate	1.1	0.544	167.17	0.091		
TiCl(O ⁱ Pr) ₃	2.2	1.089	260.58	0.284	1.091	0.26
CH ₃ COOH			60.05			3 gtt
STAB	5	2.4747	211.94	0.525		
DCM dry (solvent)						5

Procedure: General procedure 3

Purification: Biotage Isolera One

Cartridge: SNAP KP-SIL 25g, Flowrate: 50 mL/min, 5 mL fraction volume

Elution system: Diethylether (solvent A) – DCM (solvent B) 3 CV 0% (v/v) solvent B, 2.2 CV linear gradient reaching 15% (v/v) solvent B, 2 CV 15% (v/v) solvent B, 12.7 CV linear gradient reaching 100% solvent B, 2 CV 100% solvent B. Eluted at 6-9 CV

TLC: Diethylether/DCM 5:5; R_f = 0.46

Outcome: Yield: 7 mg, 2.5% ,off-white solid

Analysis:

¹H NMR (500 MHz, CDCl₃) 1.34 (dt, *J* = 7.0 Hz, *J* = 7.1 Hz, 6H), 3.97 (q, *J* = 7.0 Hz, 2H), 4.37 (q, 7.1 Hz, 2H), 4.52 (d, *J* = 5.9 Hz, 2H), 5.04 (s, 2H), 5.31 (t, *J* = 5.9 Hz, 1H), 6.78 (d, *J* = 2.0 Hz, 1H), 7.05 (d, *J* = 2.0 Hz, 1H), 7.23 (dd, *J* = 8.3 Hz, *J* = 2.1 Hz, 1H), 7.32 (d, *J* = 2.0 Hz, 1H), 7.67 (d, *J* = 8.3 Hz, 1H), 7.90 (d, *J* = 1.4 Hz, 1H), 8.75 ppm (d, *J* = 1.2 Hz, 1H)

¹³C NMR (125 MHz, CDCl₃) 14.43, 14.73 (CH₃), 44.78, 61.33, 64.76, 70.88 (CH₂), 112.08 (CH), 118.08 (C), 123.61, 127.13, 128.91, 130.61, 131.75 (CH), 132.43, 133.26, 133.91, 134.04, 135.20, 144.76 (C), 145.68 (CH), 153.23, 155.13, 164.71 ppm (C)

UPLC-MS: Method B

RT: 2.357 min, Purity 95 %

MS: 555.9 [M+H]⁺

3-Bromo-4-((2,4-dichloropyrimidin-5-yl)methoxy)-5-ethoxybenzaldehyde (68)

Chemical Formula: C₁₄H₁₁BrCl₂N₂O₃; MW: 406.057

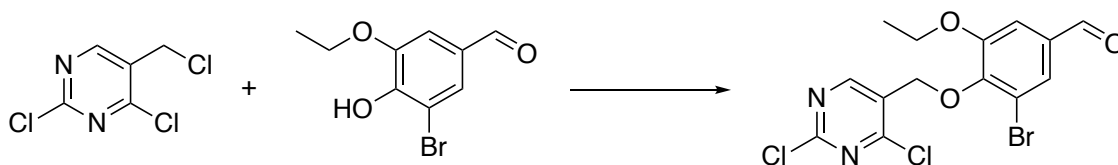


Table 65: Chemicals for the synthesis of 68

Name	Eq.	mmol	MW	g	(d) g/cm ³	V (mL)
2,4-dichloro-5-(chloromethyl)pyrimidine	1	0.816	245.07	0.200		
3-bromo-5-ethoxy-4-hydroxybenzaldehyde	1	0.816	197.44	0.161		
K ₂ CO ₃	1.5	1.224	138.20	0.169		
DMF						5

Procedure: General procedure 1

Outcome: Yield: 238.8 mg, 72 %, off-white solid

Analysis:

¹H NMR (500 MHz, CDCl₃) 1.44 (t, *J* = 7.0 Hz, 3H), 4.16 (q, *J* = 7.0 Hz, 2H), 5.26 (d, *J* = 0.7 Hz, 2H), 7.41 (d, *J* = 1.8 Hz, 1H), 7.67 (d, *J* = 1.8 Hz, 1H), 8.98 (s, 1H), 9.87 ppm (s, 1H)

¹³C NMR (125 MHz, CDCl₃) 14.77 (CH₃), 65.24, 68.79 (CH₂), 112.85 (CH), 117.87 (C), 127.34 (CH), 129.01, 134.13, 149.17, 153.25, 159.18, 161.91 (C), 161.96, 191.52 ppm (CH)

Ethyl 4-((3-bromo-4-((2,4-dichloropyrimidin-5-yl)methoxy)-5-ethoxybenzyl)amino)benzoate (69)

Chemical Formula: C₂₃H₂₂BrCl₂N₃O₄; MW: 555.25

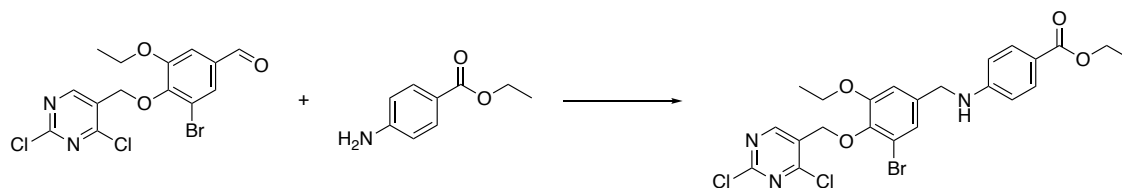


Table 66: Chemicals for the synthesis of 69

Name	Eq.	mmol	MW	g	(d) g/cm ³	V (mL)
3-bromo-4-((2,4-dichloropyrimidin-5-yl)methoxy)-5-ethoxybenzaldehyde	1.1	0.2462	406.06	0.100		
Ethyl-4-aminobenzoate	1	0.2239	165.19	0.037		
TFA*	2	0.4478	114.02	0.051	1.49	0.034
STAB	1.2	0.2687	211.94	0.057		
EtOAc						3

* TFA dilution in EtOAc: 0.34 mL in 1 mL of EtOAc

Procedure: adapted from (McLaughlin et al. 2006)

Benzocaine is added together with **68** into a round bottom flask with EtOAc. TFA is added and stirred. Then NaBH(OAc)₃ is added portionwise and the mixture is heated to 40° C for 10 min.

Work up: 10% wt aqueous NaOH is added to the mixture to a pH of 8-9. Phases were allowed to separate. The organic phase was washed with brine and dried over Na₂SO₄ and then evaporated.

Purification: Biotage Isolera One

Cartridge: SNAP KP-Sil 25g

Elution system: *n*-hexane – EtOAc: 5 CV 100% (v/v) *n*-hexane, 15 CV linear gradient reaching 100% EtOAc. Eluted at 12CV (fractions impure).

Product was obtained by precipitation in *n*-hexane/EtOAc.

Outcome: Yield: 8.6 mg, 6.3%, yellow solid

Analysis:

¹H NMR (500 MHz, CDCl₃) 1.35 (2t overlapped, *J* = 7.1 Hz, *J* = 7.0 Hz, 6H), 4.01 (q, *J* = 7.0 Hz, 2H), 4.31 (m, q & s overlapped, *J* = 7.1 Hz, 4H), 4.49 (t, *J* = 5.6 Hz, 1H), 5.12 (s, 2H), 6.55-6.57 (m, 2H), 6.84 (d, *J* = 1.8 Hz, 1H), 7.12 (d, *J* = 1.8 Hz, 1H), 7.85-7.88 (m, 2H), 8.96 ppm (s, 1H)

¹³C NMR (125 MHz, CDCl₃) 14.45, 14.70 (CH₃), 47.08, 60.32, 64.69, 68.19 (CH₂), 111.30 (CH), 111.80 (2 CH), 117.88, 119.69 (C), 123.12 (CH), 128.69 (C), 131.54 (2 CH), 136.80, 143.72, 151.20, 152.94, 159.69 (C), 160.02 (CH), 160.32, 166.71 ppm (C)

UPLC-MS: Method B

RT: 2.277 min; Purity 100%

MS: 555.9 [M+H]⁺

2.3.8.3 Biology

The biological testing was performed by our collaborators at Rega Institute for Medical Research at KU Leuven, their protocols are specified below.

2.3.8.3.1 Cells, viruses and compounds

African green monkey kidney cells, Vero A cells (ATCC CCL-81) and Buffalo green monkey kidney cells, BGM cells (ECACC 90092601) were maintained in minimal essential medium (MEM Rega-3, Gibco, Belgium) supplemented with 10% fetal bovine serum (FBS, Gibco), 1% L-glutamine (Gibco) and 1% sodium bicarbonate (Gibco). The virus propagation and antiviral assays were performed in the same medium except that it was supplemented with 2% FBS instead of 10%. All cell cultures were maintained at 37°C in an atmosphere of 5% CO₂ and 95%-99% relative humidity.

CHIKV Indian Ocean strain 899 (CHIKV-899; GenBank FJ959103.1) is a lab-adapted strain that was a kind gift of Professor C. Drosten (University of Bonn, Germany). The virus stocks were prepared in Vero A cells and stored at -80°C.

All compounds were dissolved in analytical grade DMSO to yield 10 mg/ml stocks. The compounds were protected from light and were stored at -20°C until used.

2.3.8.3.2 Cytopathic effect (CPE) reduction assay

Vero A cells were seeded at a density of 2.5×10⁴ in 96-well tissue culture plates (BD Falcon) and were allowed to adhere overnight. The next day, dilution series of the compounds was prepared in the medium, after which the cultures were infected with CHIKV-899 at MOI of 0.001. On day 5 post-infection, the antiviral effect was quantified using the MTS/PMS method as described by the manufacturer (Promega, The Netherlands). The cells were checked by microscope for minor signs of virus-induced cytopathic effects or compound-induced adverse

effects on cell and monolayer morphology. The 50% effective concentration (EC_{50}), which is defined as the concentration of compound that is required to inhibit virus-induced cell death by 50%, was determined using logarithmic interpolation. In parallel, the 50% cytotoxic/cytostatic concentration (CC_{50}), which is the concentration of compound that is required to reduce cell viability by 50%, was determined in non-infected cells using the MTS/PMS method.

2.3.8.3.3 Virus yield assay

Vero A cells were seeded in 96-well tissue culture plates at a density of 5×10^4 cells/well. Next day, cells were treated with a serial dilution of the selected compound and then infected with CHIKV-899 (MOI 0.001). After 2h of infection, the cells were washed to remove non-adsorbed virus, treated again with the same serial dilutions of compounds and incubated for 48h. At the end of the incubation period, supernatant was collected and virus RNA was quantified by real-time qRT-PCR, while the amount of infectious progeny virus was determined by titration assay ($CCID_{50}/ml$).

2.3.8.3.4 Quantitative reverse transcription PCR (qRT-PCR)

Extracellular viral RNA was isolated from 150 μ l supernatant using the NucleoSpin RNA virus kit (Macherey-Nagel, Düren, Germany), while the intracellular viral RNA was isolated using the Cells-to-cDNA™ lysis buffer (Life Technologies). The sequences of primers and probe used in qRT-PCR: forward primer 5'-CCGACTCAACCATCCTGGAT-3', reverse primer 5'-GGCAGACGCAGTGGTACTTCCT-3', probe 5'-FAM-TCCGACATCATCCTCCTTGCTGGC-TAMRA. The one-step, quantitative RT-PCR was performed in a total volume of 25 μ l, containing 13.94 μ l H₂O, 6.25 μ l master mix (Eurogentec, Seraing, Belgium), 0.375 μ l of forward primer, 0.375 μ l of reverse primer (final concentration of each primer 150 nM), 1 μ l of probe (final concentration 400 nM), 0.0625 μ l reverse transcriptase (Eurogentec) and 3 μ l RNA sample. The reaction was quantified using the Applied Biosystems 7500 Fast Real-Time PCR System using the following conditions: 30 min at 48°C and 10 min at 95°C, followed by 40 cycles of 15 s at 95°C and 1 min at 60°C. For quantification, standard curves were generated each run using 10-fold dilutions of a CHIKV standard cDNA.

2.3.8.3.5 Determination of $CCID_{50}$ per ml

Briefly, Vero A cells were seeded in 96-well microtiter plates at a density of 2.5×10^4 cells/well and were allowed to adhere overnight. The next day, 6 parallel 10-fold serial dilutions of the virus-containing samples were prepared. After 5 days of incubation, the cells were examined

microscopically for virus-induced cytopathic effects (CPE). A well was scored positive if any traces of virus-induced CPE were observed compared to the uninfected controls. The CCID₅₀/ml was calculated using the method of Reed and Muench (Reed & Muench 1938) and is defined as the virus dose that would infect 50% of the cell cultures.

2.3.8.3.6 Delay of treatment assay

Vero A cells were seeded in 96-well tissue culture plates at a density of 5×10^4 cells/well and incubated overnight. Two hours prior to CHIKV-899 infection, the selected compound was added at the condition -2 hours. Subsequently, at time point 0, the medium of all wells was removed and the cells were infected with virus with MOI of 1 for 1 hour at 37°C. The compounds were added at the previously mentioned concentrations at 0, 2, 4 and 6 hours after infection. Following 24h of incubation, cells were lysed the intracellular RNA was quantified by qRT-PCR as described before.

2.3.8.3.7 Entry assay using CHIKV pseudoparticles (CHIKVpp)

CHIKV pseudoparticles (CHIKVpp) were prepared as reported in Salvador *et al.*, 2009. BGM cells were seeded in white 96-well tissue culture plates (ViewPlate-96, PerkinElmer) at a density of 2.5×10^4 cells/well in assay medium and left to adhere overnight. The next day, serial dilutions of the compounds were added to the cells followed by infection with the appropriate dilution of CHIKVpp. Chloroquine was used as a positive control entry inhibitor. On day three post-infection, cells were lysed and the firefly luciferase activity in the cell lysate was detected using the Luciferase Assay System kit as described by the manufacturer (Promega).

2.3.8.3.8 Evaluation of 10 on different Alphaviruses and on West Nile Virus

This evaluation was performed at Aix-Marseille Université as a courtesy of Gilles Querat.

Virus strains

Chikungunya: Opy1 (Réunion LR2006_OPY1) EVAg 001v-EVA83; O’Nyong Nyong (Senegal IPD A234), Ross River Virus (NCPV 5281v); Venezuelan Equine Encephalitis Virus (P676 NCPV ref 0605153v); West Nile virus ((R94224) USA Wisconsin.

Virus yield reduction assay

The amount of each virus and the duration of the assay have been calibrated by trial so that the replication is still in log phase of growth at the time of readout and the Ct standard

deviations of qRT-PCR quantification (quadruplicate) is below 0.5. Approximate multiplicity of infection (MOI) range from 10^{-4} to 10^{-3} depending on the strain.

One day prior to infection 5×10^4 Vero E6 cells were seeded in 100 μ l of medium (with 2.5 % FCS) in each wells of a 96-well titer plates. The next day, 8 two-fold serial dilutions of the compounds (beginning at 20 μ M final concentration, down to 0.16 μ M), in duplicates or triplicates, were added to the cells (25 μ l/well, in 2.5 % FCS containing medium). Four Virus Control (VC) wells (per virus) were supplemented with 25 μ l medium. Fifteen minutes later, 25 μ l of a virus mix containing the appropriate amount of viral stock diluted in medium (2.5 % FCS) was added to the 96-well plates.

Cells were cultivated for 1.5 to 2 days after which 100 μ l of the supernatant were collected for viral RNA purification. The infected supernatants were transferred to 96 wells S-Bloc from QIAgen preloaded with VXL mix and extract by the Cador Pathogen 96 QIAcube HT kit run on QIAcube HT automat according to Qiagen protocol. Purified RNAs were eluted in 80 μ l of water.

Viral RNAs were then quantified by real time one step RT-PCR to determine viral RNA yield using 3.5 μ l of RNA and 6.5 μ l of RT-PCR mix using standard cycling parameters. The four control wells were replaced by four 2 log dilutions of an appropriate T7-generated RNA standards of known quantities for each viral genome (100 copies to 100 millions copies).

IC₅₀ (half maximal inhibitory concentration) and IC₉₀ (90% inhibitory concentration determination

Mean inhibition of virus yield is equal to $100 \times (\text{mean quantities of viral RNA in VC quadruplicates} - \text{mean quantities of viral RNA in drug treated triplicates}) / \text{mean quantities of viral RNA in VC}$. The inhibition values (expressed as percent inhibition, in linear scale) obtained for each drug concentration (expressed in μ M, in log scale) are plotted using Kaleidagraph plotting software (Synergy Software) and the best sigmoidal curve, fitting the mean values, is determined by a macro in the software: (Inhibition, Y is given by $Y = 100 / (1 + (m_0/m_1)^{m_2})$). This macro allows determining the best curve fit and the m1 and m2 parameters, where m1 corresponds to IC₅₀. The reverse equation $x = M1 \cdot ((100/y) - 1)^{1/m_2}$ allows to calculate x: IC₉₀ concentration for Y = 0.9.

Cytotoxicity assay

One day prior to the assay 5×10^4 Vero E6 cells were seeded in 100 μ l of medium (with 2.5 % FCS) in each wells of a 96-well titer plates. The next day, two-fold serial dilutions of the compounds (beginning at 200 μ M final concentration, down to 6.2 μ M), in triplicates, were added to the cells (25 μ l/well, in 2.5 % FCS containing medium). Six cell control (CC) wells were

supplemented with 25 μ l medium. Two wells were not seeded by cells and served as background control of fluorescence for the plates.

Cells were cultivated for 2 days after which the supernatant was removed and replaced with 70 μ l of medium containing CellTiter-Blue reagent (Promega) and further incubated for 90 min at 37 ° C. Fluorescence of the plates were then read on a TECAN Infinite M 200 Pro reader. The cell viabilities, in percent, were calculated as $100 \times (\text{mean value of X} - \text{Background without cells}) / (\text{CC} - \text{background})$.

2.4 Targeting the distal ribose pocket of CHIKV macro domain

A detailed description of the CHIKV macro domain was already provided in the previous section. In contrast to the previous chapter, where the entire ADP-ribose binding pocket was investigated, the aim of this part of the project is to have a closer look at the mode of action of mono-ADP-ribose (MAR) cleavage from other proteins. It was described that CHIKV macro domains are capable of binding and removing MAR moieties from host proteins especially from aspartate and glutamate residues, which are linked via the distal ribose of the ADP-ribose molecule at the 1''-OH group. (Li et al. 2016) The crystal structures currently available have only ADP-ribose, RNA or 2'-5'-oligoadenylate co-crystallised (Malet et al. 2009; Morin et al. 2014), and the mechanism of hydrolysis of the ester bond with the second protein containing the post-translational MARylation has not been conclusively elucidated, yet.

2.5 Aims

The aim of this part of the project was to use pharmacophore searches, molecular docking and consensus rescoring to identify new potential inhibitors for the CHIKV macro domain, at the distal ribose site, that were then purchased and tested in antiviral assays. Two different libraries were used: the SPECS library and the Prestwick chemical fragment library, the latter was kindly provided by Prestwick Chemical who are partners in the Antivirals ETN. Then the resulting compounds were docked into CHIKV macro domain and three different scoring functions were used to evaluate each compound. The compounds in the final selection were docked into several human macro domains in order to evaluate possible off-target effects. Visual inspection was crucial to decide on a final compound selection.

2.6 Methods

2.6.1 Hardware and Software

All stages of the computational work were carried out on Vigle genie processor Intel® core i7-4790 CPU@ 3.60 GHz x 8 running Linux Ubuntu 16.04 using Molecular Operating Environment (MOE) 2015.10 (Chemical Computing Group Inc. 2016), FlexX module in LeadIT 2.1.8 (BioSolveIT GmbH 2006), Protein Ligand ANT system (PLANTS) version 1.2 (Korb et al. 2009), Maestro (Schrödinger LLC 2018).

2.6.2 Protein preparation

The 3D crystal structure of the CHIKV macro domain in complex with ADP-ribose (PDB ID: 3GPO) was downloaded from Protein Drug Bank (PDB) (Berman et al. 2000) to Molecular Operating Environment (MOE). Structure preparation and Protonate3D were carried out with the tools comprised in MOE. All water molecules were removed and chain D was kept to create the pharmacophore queries. Structure preparation was also applied to the human proteins downloaded from the PDB: 2X47, 4IQY, 4ABK, 4J5R (Chen et al. 2011; Jankevicius et al. 2013; Forst et al. 2013; Sharifi et al. 2013).

The two libraries of compounds were used. First, the SPECS library (www.specs.net) of commercially available compounds which contains roughly 456,855 molecules, and the drug fragment library of Prestwick Chemical, which was provided for research purposes, as a courtesy of Prestwick chemical, and contains approximately 2,000 compounds. Conformational search on the SPECS library had previously been carried out in our group. The Prestwick drug fragment database was imported into an MOE database. A wash step was applied, using the default parameters, before a conformational search was performed on the database, resulting in 60,221 fragment conformations.

2.6.3 Pharmacophore queries and searches

The pharmacophore queries were prepared in MOE, using the query editor, to elucidate the pharmacophore features on the ligand. Pharmacophore searches were run against the prepared SPECS and Prestwick drug fragment databases separately, but the final files were merged.

2.6.4 Docking and Scoring

Results from the pharmacophore searches were docked into the prepared 3GPO structure using Glide SP and subsequently rescored with Glide XP, FlexX and PLANTS. The united database of unique compounds, resulting from the different pharmacophore searches, was prepared for docking using the ligprep tool of Maestro. The protein was prepared with the Protein Preparation Wizard, then a grid box for the docking was created with Glide grid generation. Coordinates of the ring oxygen of the distal ribose ($x=5.977$, $y=44.088$, $z=-16.55$) were used for the centre of the grid. The same coordinates were also used for the rescoring with FlexX and PLANTS. PLANTS can process the direct coordinates for the rescoring but FlexX needs a pseudo-ligand in the position of the grid centre; one nitrogen atom was used as a dummy.

The human macro domains were superimposed on to the 3GPO structure and equally prepared for the docking.

2.6.5 Biological evaluation

The compounds were tested in the same CPE assay as described before at the Rega Institute for Medical Research at the KU Leuven, Belgium.

2.6.6 Cross evaluation against human macro domains

After the docking and consensus rescoring of the database of roughly 250 molecules into the different human macro domains the compounds were visually inspected for their interactions with the respective macro domains. Compounds that seemed to fit better into human macro domains than into the CHIKV 3GPO protein were then removed from the penultimate selection of compounds that were chosen for the test against CHIKV in the CPE assay. The structures of Macro D1, Macro D2, TARG1 and PARP14 were downloaded from the protein databank using the accession codes 2X47, 4IQY, 4ABK, 4J5R (Chen et al. 2011; Jankevicius et al. 2013; Forst et al. 2013; Sharifi et al. 2013).

2.7 Results

2.7.1 Library preparation

Below, Figure 30 represents the workflow that was applied to the present project. At the beginning two libraries were selected for the screening of compounds with the pharmacophore models discussed in the next section. The SPECS library is a library of commercially available screening compounds that can be readily purchased and tested. In our group this is the most frequently used library and for computational purposes the virtual library is centrally prepared and curated. Therefore, it is readily available with a set of prepared conformations that can be screened with pharmacophore models or used for docking purposes. A total of 465,855 compounds were collected in this library in the version used for this study.

The fragment library from Prestwick chemical was kindly provided through our network collaborators, especially Thierry Langer and Marie-Louise Jung. It is a library of approximately 2000 fragments in .sdf format. The library needed to be prepared for searches requiring conformations of the compounds. The represented molecules are on average not larger than 300 Da in molecular weight and are derived from the Prestwick chemical library of FDA-approved drugs. Features of the fragments can be linked to their parent compound for which

the mode of action is often known, and certain conclusions on the activity can be drawn when they are used in screenings.

It is of importance that the two sets of libraries are not directly comparable. Smaller compounds might achieve lower scores in pharmacophore screening, and sometimes even docking, either because they are not capable of covering the physical space of the query compared to a larger molecule, or they might not be able to interact with as many hydrogen bond donors or acceptors when they do not have these features themselves. Therefore, a score of interactions per molecular weight, can help to level out this bias and is often implemented in the scoring function.

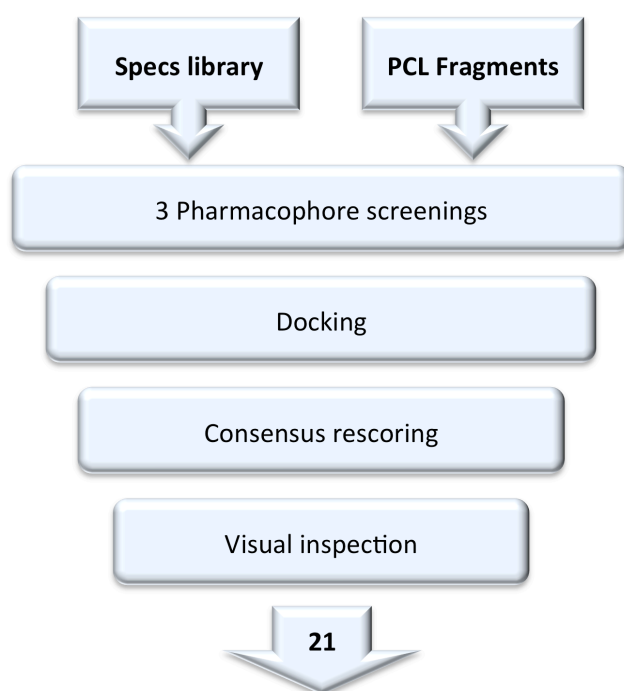


Figure 30: Screening workflow describing the course of the compound selection

2.7.2 Pharmacophore modelling

The binding pocket of ADP-ribose of CHIKV macro domain can be split into two parts: the adenosine binding part and the part that is occupied by the distal ribose. Both are linked by a narrow channel where two phosphate groups are residing that link the proximal and the distal ribose. The distal ribose is important in macro domains because it is the attachment point to another protein via the 1''-hydroxyl group of the sugar moiety that forms an ester bond with glutamic acid or aspartic acid. Other amino acids are also able to attach to the distal ribose, but CHIKV macro domain is only able to recognise and cleave from aspartic and glutamic acid-linked proteins as substrates. CHIKV macro domain can catalyse the cleavage of the ester-bond between the sugar and the amino acid. The linked MAR molecule serves as a signal molecule that is attached to proteins as a post-translational modification. In the context of a viral

infection this signal is supposed to increase the antiviral reaction of the cell or the organism to, for example, CHIKV (Li et al. 2016).

In order to exactly target the desired region in the spacious binding pocket of the CHIKV macro domain, several pharmacophore models were created that cover the space and features of the distal ribose. The MOE pharmacophore editor highlights and helps placing pharmacophore features on the ribose for the optimal interaction between the protein and its natural substrate.

Several queries were tested for the compounds that they retrieved and finally the three queries depicted in Figure 31, Figure 33 and Figure 35 were kept for the selection of compounds for the next steps in the workflow.

Query A contains four features. F1 and F2 are positions where either a donor or an acceptor can be placed. F3 requires any heavy atom in this precise position. F4 is set to require an acceptor and the receptor space was used to define an exclusion volume, where no atom of a potential compound can be placed. None of the features was set to essential in this query. It was run against both the SPECS and the Prestwick drug fragment library and retrieved 717 conformations of a total of 56 unique compounds for SPECS and 19 unique compounds for the Prestwick drug fragment library. A compound retrieved for each of the databases is depicted in Figure 32.

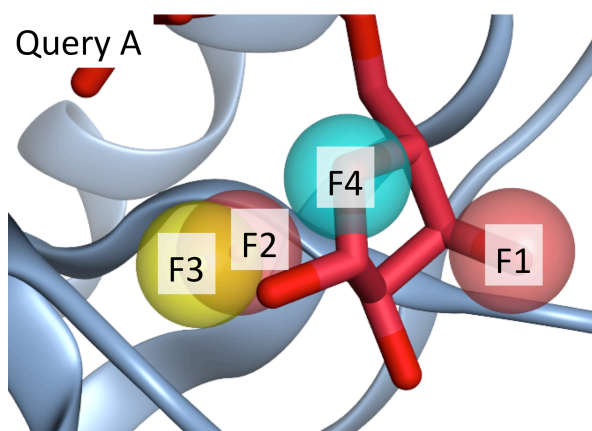


Figure 31: Query A on the distal ribose site
CHIKV macro domain 3GPO (blue), ADPR (red)
Pharmacophore features:
Acceptor (cyan), Acceptor or Donor (pink), AtomQ (yellow)

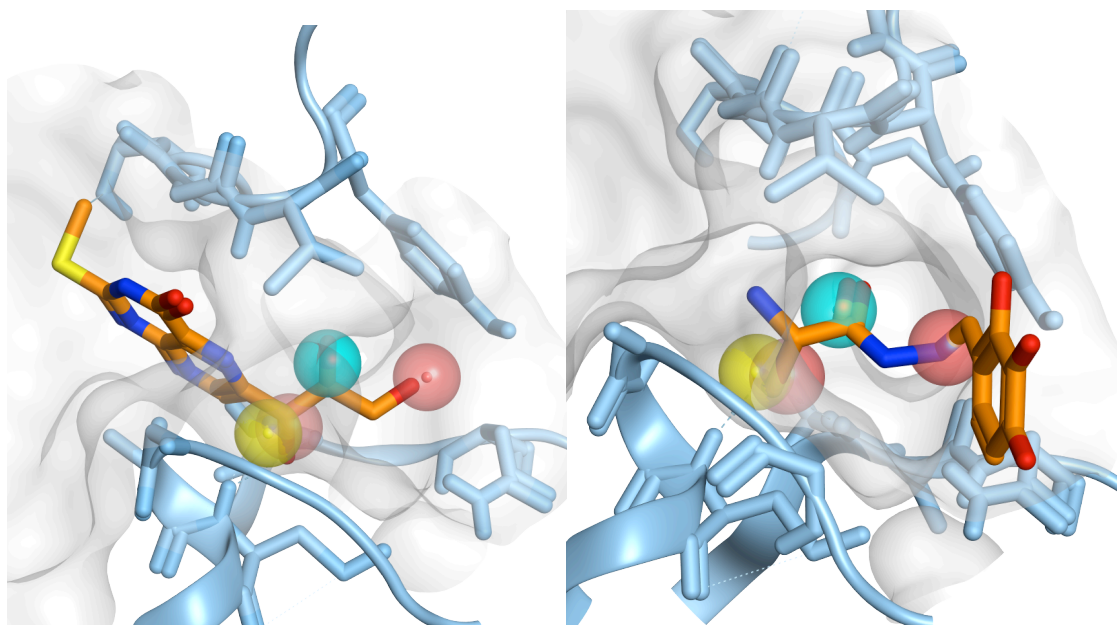


Figure 32: Example of query A with a compound (orange) from SPECS (left) and PCL (right)

Query B also had four features. F1 and F3 were hydrogen bond acceptors and F2 and F4 were both hydrogen bond donors. F3 and F4 were chosen to be essential, which means that both features had to be matched by a compound in order to be selected. As in query A an exclusion volume for the receptor space was added. Out of both libraries 49,736 conformations yielding a total of 1,785 unique compounds were satisfying the query. One result for each library is depicted in Figure 34

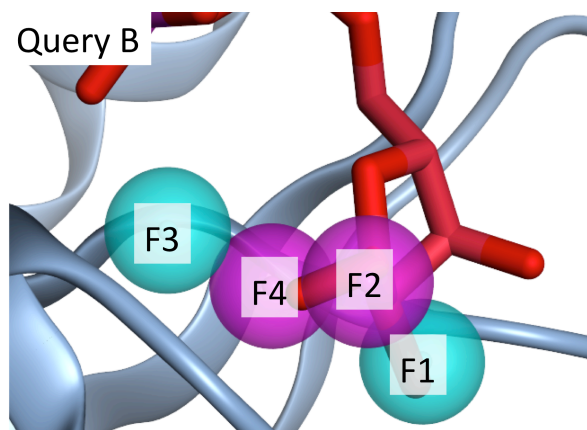


Figure 33: Query B on the distal ribose site
 CHIKV macro domain 3GPO (blue), ADPR (red)
 Pharmacophore features: Acceptor (cyan), Donor (purple)

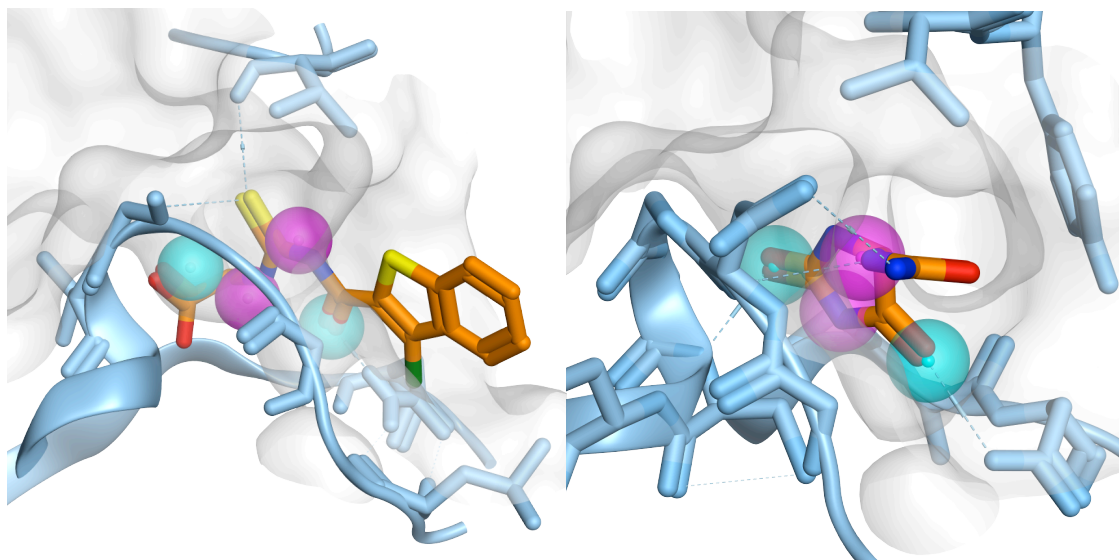


Figure 34: Example of query B with a compound (orange) from SPECS (left) and PCL (right)

Query C was composed of 4 acceptor features, which were quite widely distributed in the space around the distal ribose. Only F1 was set to essential and no exclusion volume was required, making this query rather permissive. 30,712 conformations of 1,500 compounds were resulting from this query. A result retrieved by the query for each library is depicted in Figure 36

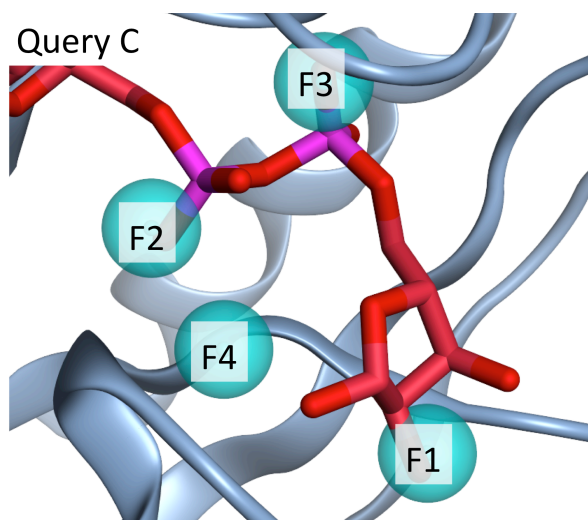


Figure 35: Query C on the distal ribose site
 CHIKV macro domain 3GPO (blue), ADPR (red); Pharmacophore features: Acceptor (cyan)

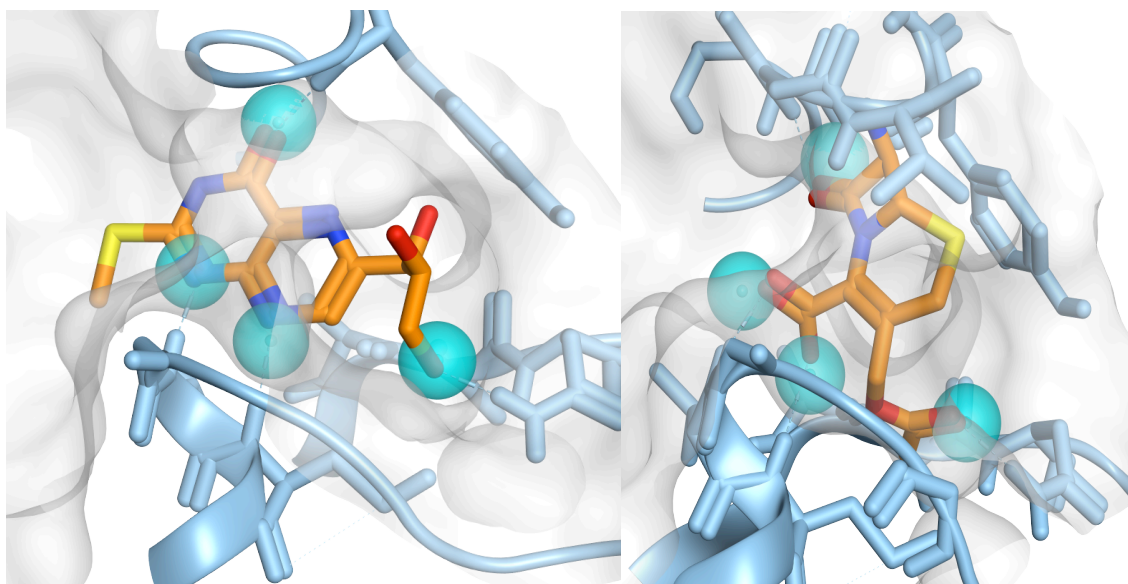


Figure 36: Example of query C with a compound (orange) from SPECS (left) and PCL (right)

The compounds were then incorporated into one common database where duplicates were again eliminated and only one conformation per compound was kept for the next steps. This database contained a total of 3,235 compounds.

2.7.3 Docking

The merged database was used in the docking procedure. The docking was carried out with Maestro Glide SP. To prepare the compound database for the use with Maestro it was subjected to the ligand preparation tool ligprep in Maestro. 32 conformations per compound were generated for the docking. Then the receptor needed to be prepared for the docking, as well. First the protein was treated in the Protein Preparation Wizard in Maestro using the default settings. Then the position on the protein for the compounds to be docked needed to be specified. A grid box was centred on C4'' of the distal ribose of ADPR complexed with chain A (coordinates $x=5.977$, $y=44.088$, $z=-16.55$). The size of the cubic box was set to 12 Å, which is the distance from the centre, in which most of the ligands can be accommodated and which still specifies the site precisely enough. During the docking procedure the receptor was kept as a rigid entity and the ligand was treated flexibly. 5 poses per ligand were generated in the docking. In the resulting database 38,043 poses were recorded which were then subjected to consensus rescoring. Starting with a relatively small compound library for the docking, the complete database was rescored.

2.7.4 Consensus rescoring

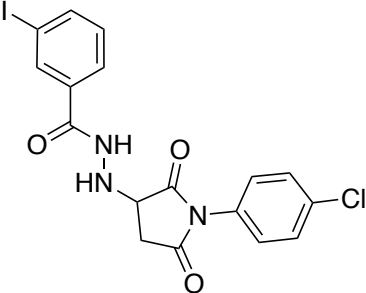
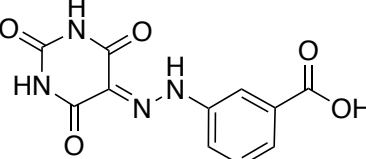
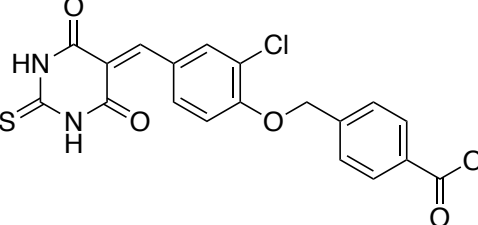
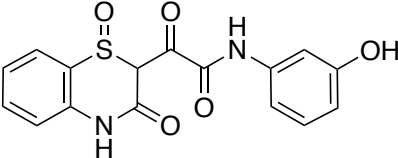
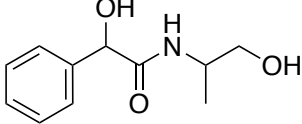
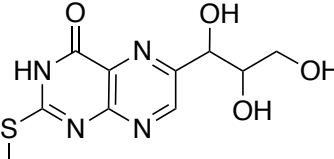
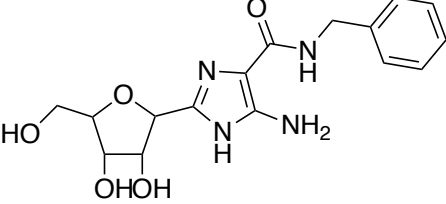
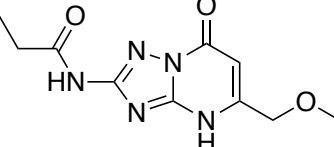
Consensus rescoring is a procedure that uses different scoring functions to re-evaluate the poses resulting from the docking. Already the Glide SP docking ranks the compounds according to the internal scoring function but Glide extra precision (XP) takes more factors into account and is therefore used as the first of the three rescoring functions. As the second scoring function FlexX score is used. FlexX is integrated in the Leadit software suite. Also there the receptor needed to be prepared specifically for the program and the centre needed to be specified. In this case we used a dummy atom on the position of the ring oxygen to specify the centre. The database resulting from the Glide XP scoring was used as an input file in .mol2 format. Maestro allows for the transformation of the data between the different file formats. The third scoring function is integrated in the software PLANTS. It's an algorithm based on ant colony theory and does not directly calculate a binding energy value. The score between the three scoring functions is in a different range and therefore not directly comparable. In order to take into account all three scoring functions equally, the rank by vote strategy (Wang & Wang 2002), was applied. Compounds with a vote of all three scoring functions were selected for visual inspection. In total 386 compounds were processed through visual inspection.

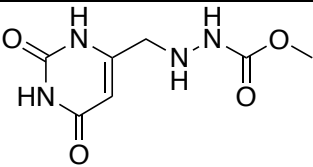
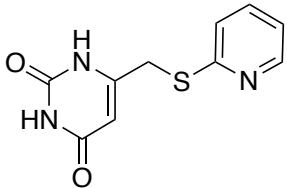
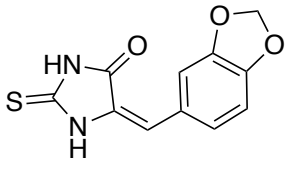
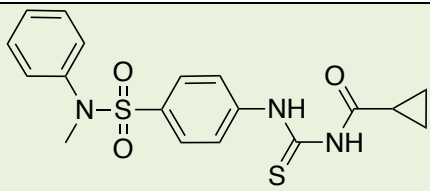
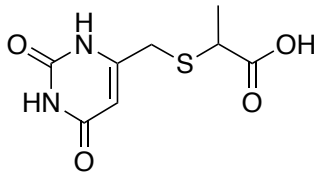
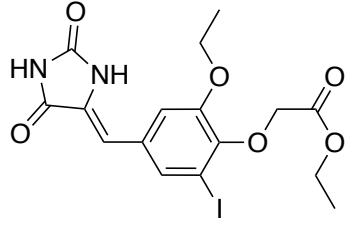
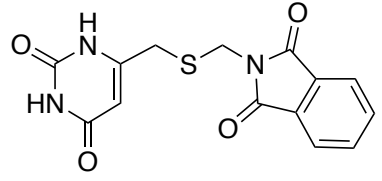
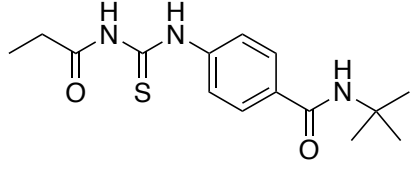
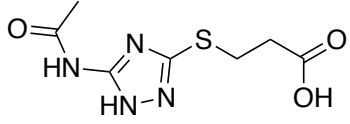
2.7.5 Visual inspection

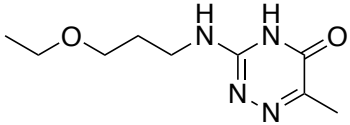
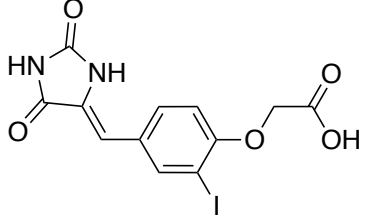
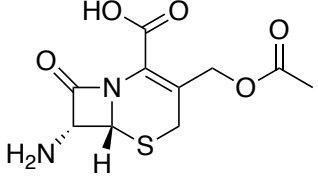
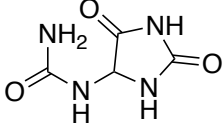
The visual inspection of the compounds selected in the scoring procedure is a lengthy process where the experience and judgement of the computational medicinal chemist plays an important role. The visual inspection and compound selection were primarily carried out by a project student, Adit Patel, under my direct supervision. In several rounds we reduced the number of compounds from 386 to 13. The main criteria for visual inspections were the space occupied by the compounds as we wanted to target the distal ribose site. The interactions with the key residues of the protein and the drug-likeness of the compound itself also played a role during the selection. Tools supporting the judgement were the ligand interaction tool in MOE, and several molecular descriptors like molecular weight, SlogP, number of hydrogen-bond donors and acceptors per molecule.

Finally, 13 compounds were selected (Table 67). 11 molecules were from the SPECS library dataset and 2 molecules from the Prestwick fragment library collection. The compounds from SPECS were purchased. Other compounds were selected from the penultimate round of visual inspection so that a total amount of 21 compounds was bought. Figure 37 shows the interactions with the protein of one representative compound within the binding pocket.

Table 67: List of selected compounds for biological testing

Nr	Compound structure	SPECS code/ Name	Activity
70		AG-690/13779103	Not active
71		AN-329/33311027	Not active
72		AH-487/41660565	Not active
73		AQ-911/41963634	Not active
74		AR-360/42760614	Not active
75		AB-323/13887441	Not active
76		AJ-030/14523202	Not active
77		AM-900/14782004	Not active

78		AI-204/31697013	Not active
79		AI-204/31697035	Not active
80		AG-690/33251021	Not active
81		AN-329/42612973	EC ₅₀ ≈ 5.1 μM
82		AI-204/31697018	Not active
83		AN-655/41063757	Not active
84		AI-204/31697039	Not active
85		AN-329/43448538	Not active
86		AP-853/43464285	Not active

87		AS-871/43477421	Not active
88		AA-487/40935550	Not available on SPECS
89		7-Amino cephalosporanic acid PCL fragments	Not tested
90		Allantoin PCL fragments	Not tested

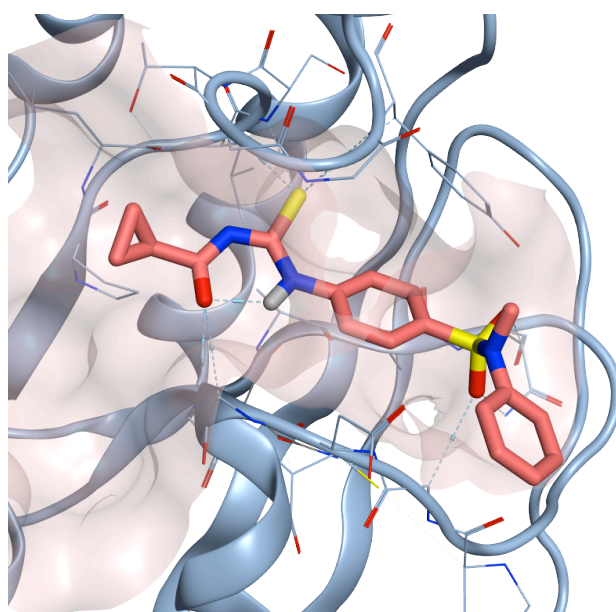


Figure 37: Compound 81 docked into 3GPO

2.7.6 Biological evaluation

The 19 compounds were then sent to our collaborators at the Rega Institute for Medical Research at the KU Leuven in Belgium. Compounds were evaluated for their antiviral activity against CHIKV in a cell-based CPE reduction assay. Simultaneously, the cytotoxicity was evaluated to differentiate the antiviral effect from possible toxicity against the host cell. Out of the 19 compounds sent for testing only one active compound was identified. The inhibition and cytotoxicity curves from the CPE assay are depicted in Figure 38 and Figure 39.

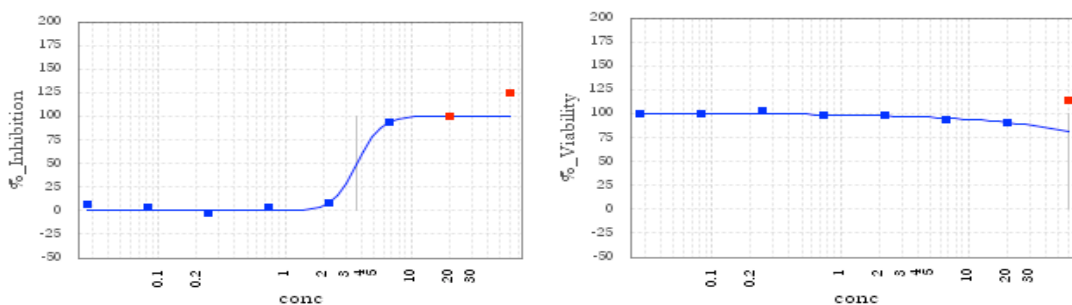


Figure 38: Assay 1 Compound **81** antiviral CPE-assay (left) and MTS assay (right);
 $EC_{50} = 6.5 \mu\text{M}$; $CC_{50} > 20 \mu\text{M}$

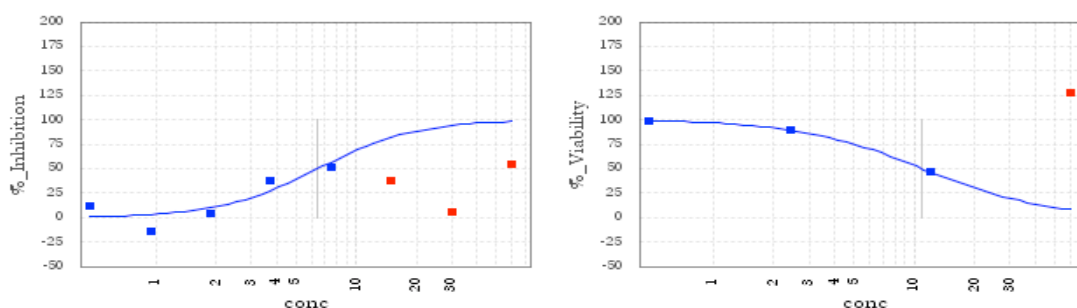


Figure 39: Assay 2 Compound **81** antiviral CPE-assay (left) and MTS assay (right);
 $EC_{50} = 3.7 \mu\text{M}$; $CC_{50} = 10.97 \mu\text{M}$; apparent precipitation of compound

Given the small amount of compounds tested in the cell-based assay the result of one slightly active compound gives hopes and possibilities to the chemist that this compound can be improved and potentially lead to more active molecules against CHIKV. Furthermore, cell-based assays are only the first indicator for the selection of a candidate that is worth investing time to elucidate the precise mode of action. Although the design of the study aimed at finding compounds targeting the CHIKV macro domain, it cannot be excluded that other mechanisms are responsible for the activity against the virus within the cell. A confirmation for the hypothesised mechanism of action would be a direct binding assay like TSA or ITC on the purified CHIKV macro domain, which was not possible for this set of compounds due to the lack of time during this project. A biochemical assay directly on CHIKV macro domain might also help to identify other compounds that are actually able to interact with the target but are

not penetrating the cell or have other unfavourable physicochemical properties such as solubility issues, for example.

In Figure 38 and Figure 39 also the cell-viability is depicted. The second to last measurement already shows very slightly decreased cell-viability to a degree of 90%. In assay 2 the cytotoxicity seems to be more pronounced, but further testing would be needed.

2.7.7 Computational evaluation against human macro domains

As mentioned in the introduction, there are several types of ADP-ribose binding proteins in all domains of life and in viruses. Here we aim at targeting the viral macro domain of CHIKV but there are several human macro domains that are essential for the normal function of the organism and should therefore not be inhibited by our compounds. In order to test this first with an *in silico* approach, the crystal structures of the human macro domains that are most similar to CHIKV macro domain, were selected from the protein databank and prepared for the use in MOE and Maestro as was 3GPO, previously. PDB structures of human D1 macro domain (2X47), human D2 macro domain (4IQY), TARG1 (4J5R) and PARP14 (4ABK) were retrieved (Chen et al. 2011; Jankevicius et al. 2013; Forst et al. 2013; Sharifi et al. 2013).

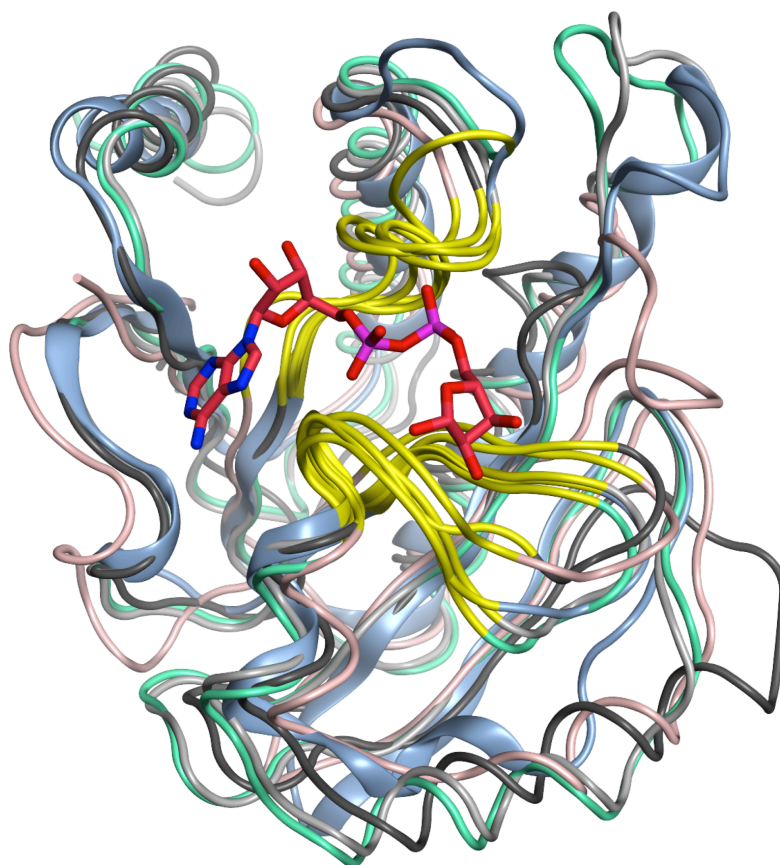


Figure 40: Superimposition of 4IQY (grey), 2X47 (cyan), 4J5R (pink), 4ABK (black) onto 3GPO (blue); Pocket region in yellow; ADP-ribose in red

Table 68: Identity and similarity of the human macrodomain containing proteins

3GPO vs.	4IQY	2X47	4J5R	4ABK
Identity	23.4%	22.8%	7.6%	14.6%
Identity Pocket	50.0%	50.0%	16.7%	30.6%
Similarity	36.7%	36.7%	25.9%	34.2%
Similarity Pocket	55.6%	55.6%	38.9%	41.7%
RMSD	1.76 Å	2.04 Å	2.07 Å	2.36 Å
RMSD Pocket	0.95 Å	1.8 Å	2.17 Å	1.4 Å

The four proteins were superimposed to 3GPO and compared for their amino acid identity and similarity in the distal ribose region and on the overall ADPR coordinating residues. For some human macro domains mechanisms of cleavage were postulated as discussed in the introduction to macro domains. CHIKV macro domain is most similar to human macro D2 domain, here represented by the structure 4IQY.

To evaluate if the selected compounds might show some off-target effects on human macro domains, a file with 263 unique compounds all ranked favourable by all three scoring functions (saved after the first round of visual inspection) was docked and rescored with the same procedure as for CHIKV macro domain. For 4IQY none of the top ranked compounds was coinciding with the selected compounds for 3GPO. In 2X47 compound **75** of the sum3 compounds from the consensus rescoring was also found favourable in 3GPO. While docking the compounds into 4J5R a problem with FlexX occurred and not all compounds could be ranked by its scoring function, leaving empty cells in the database, which were treated as if the value was zero. The compounds ranked with a sum of 3 contained 7-Aminocephalosporanic acid (**89**), which was also among the best molecules for the CHIKV dataset. The last docking and rescoring of the database into the human macro domain 4ABK for PARP14 picked up two compounds that were also present in the selected CHIKV compounds namely compound **74** and compound **90** (Allantoin).

None of the compounds tested showed cytotoxicity on VeroA cells in the MTS assay conducted in parallel with the antiviral evaluation apart from compound **80** and the active compound **81**, which decrease cell viability at concentrations higher than 43 μM and 20 μM respectively. For compound **81** the cell viability curve is falling minimally to a level of around 90% of the initial level and coincides with precipitation of compound in the medium. This does not directly apply to toxicity in humans and for human cell lines, but it would be interesting to test the cell viability also in more relevant cell lines. These efforts go beyond the scope of the project at the present stage. The two fragments from the Prestwick drug fragment library were not tested in the antiviral assay, so toxicity of these compounds in the cells could not be tested either.

Due to time constraints only a basic evaluation of the set of compounds on human macro domains was carried out. There are several ADPR binding modules that might accommodate the selected compounds and play important roles in the human physiology, therefore a more in depth analysis of compounds targeting viral macro domains, both from the computational side and *in vitro*, would be best practice.

2.7.8 Future perspectives

In this part of the project one compound against CHIKV was discovered. Unfortunately it was not possible to synthesise an improved set of compounds or to investigate the active hit in greater depth. Several points could be addressed in the future. First, probing if the compound does actually bind to the CHIKV macro domain. If the compound acts on CHIKV macro domain, it would be interesting to apply a target-based approach to improve the interactions resulting in a more potent compound. As CHIKV macro domain can be readily crystallised it would also be interesting to solve a crystal structure with the new inhibitor.

On the other hand, the physicochemical properties of the hit molecule are suboptimal for further biological characterisation. The CPE and cytotoxicity assays already showed a limited solubility of the compound in aqueous medium. Together with a structure-activity relationship studies it would also be useful to improve the compound in terms of bioavailability properties, especially solubility. This would be a key factor for target confirmation and mode-of-action studies based on resistance screening.

Furthermore, as there are several human macro domains that are quite related to the one of CHIKV, it would be important to test the off-target activity of the compounds and ensure the safety in vitro before the compound(s) could move forward to be tested in animals. This could be first achieved by testing them in several cell-lines and to evaluate them, as well, on the different isolated macro domain containing proteins.

Finally, not only CHIKV possesses a macro domain, several other viruses like HEV, Coronaviruses, but also many of the other alphaviruses, contain one or more macro domains. They were already evaluated and some of them are more similar than others. It could be interesting to test the compound also against the other macro domain containing viruses to evaluate if the compound(s) are active against them, as well. This could also be a step towards the better understanding of macro domains in the viral replication cycle and the interactions with their host, in which viral macro domains seem to be critical.

2.8 Modelling the P23 precursor protein

The non-structural proteins of CHIKV are synthesized as one polyprotein, which is then cleaved by the protease domain of nsP2. The last cleavage occurs between nsP2 and nsP3. The efficiency of the cleavages at the different cleavage sites was evaluated using small peptides, but during viral infection the spatial arrangement of the overall protein complex might account for cleavage efficiency and order, as well (Saisawang, Sillapee, et al. 2015). The P2/3 cleavage in SINV marks also the irreversible switch of the template RNA uniquely to the minus strand template in order to produce the positive sense genomic and subgenomic RNAs required in the late stage of the infection for the formation of new viral particles (Lemm et al. 1994; Shirako & Strauss 1994).

The parts of the nsPs that can be readily investigated with structure-based methods are the macro domain of nsP3 (PDB: 3GPG, 3GPO, 3GPQ), and the protease and methyltransferase-like domain of nsP2 (PDB: 3TRK). Previous work in our group has focussed on the protease domain (Bassetto et al. 2013) and the two previous sections were investigating the CHIKV macro domain in greater details. The generation of a homology model of CHIKV AUD is a logical and straightforward step due to the availability of a crystal structure for the closely related Sindbis virus. After obtaining this part of nsP3 all the present individual components from nsP2^{PRO} to nsP3^{AUD} could be linked together.

2.9 Aims

This part of the project was aimed to expand the available structural information for the precursor polyprotein that contains the individual non-structural proteins. The individual domains of CHIKV nsP2 and nsP3 that are crystallised are the protease and the methyltransferase-like domain for nsP2 and the macro domain for nsP3.

Using the structural information of the alphavirus unique domain of SINV (PDB: 4GUA) homology modelling was used to build the structural model for the closely related corresponding domain of CHIKV.

Although nsP2 protease and MT-like domain and nsP3 macro domain are crystallised they are not linked together and some of these domains lack a few residues on the N-terminal or C-terminal domain. Filling these gaps is the second aim of this chapter in order to have one structure of the P23 precursor that can then be used to find new druggable pockets.

2.10 Results and Discussion

2.10.1 Alphavirus unique domain (AUD)

In 2012 Shin and co-workers published the structure of a part of the polyprotein precursor spanning the protease domain of nsP2 to the AUD of nsP3. Part of the project is the elucidation of the mode of action of the non-structural proteins of CHIKV. Additional structural information of nsP3 is therefore of great interest. The crystal structure of Sindbis virus can readily be used as a template to generate a homology model for the AUD of CHIKV nsP3, for which no crystal structure is available yet (Shin et al. 2012).

2.10.1.1 Homology model of AUD

For the generation of the homology model of CHIKV AUD the sequence of the full-length precursor P1234 of CHIKV was retrieved from National Center for Biotechnology Information (NCBI sequence identifier ADJ19187) in FASTA (.fst) format and truncated to the region of interest. Because the nsP3 macro domain of CHIKV is already crystallised, only the amino acids 1494 to 1660 of the P1234 sequence were used as query. This corresponds to the amino acids 161 to 327 of nsP3 of CHIKV and represents the alphavirus unique domain (AUD).

The template crystal structure of the P23 precursor of SINV was retrieved with MOE (Chemical Computing Group Inc. 2016) using the PDB search (4GUA). The structure was then prepared in MOE and chain A was selected as the structure template.

The CHIKV sequence was aligned in MOE to the corresponding protein sequence of SINV (NCBI sequence identifier NP_740672) present in the crystal structure of the P23 precursor spanning nsP2 protease domain to nsP3 AUD (also referred to as zinc binding domain – ZBD). The alignment was manually adapted according to the alignment used in Shin et al. 2012 (Supplemental Fig. S2). The sequence identity between the sequences is very high with 60% pairwise identity. Similarity for the pairwise comparison between the CHIKV and the SINV sequence for the selected region was $\approx 77\%$.

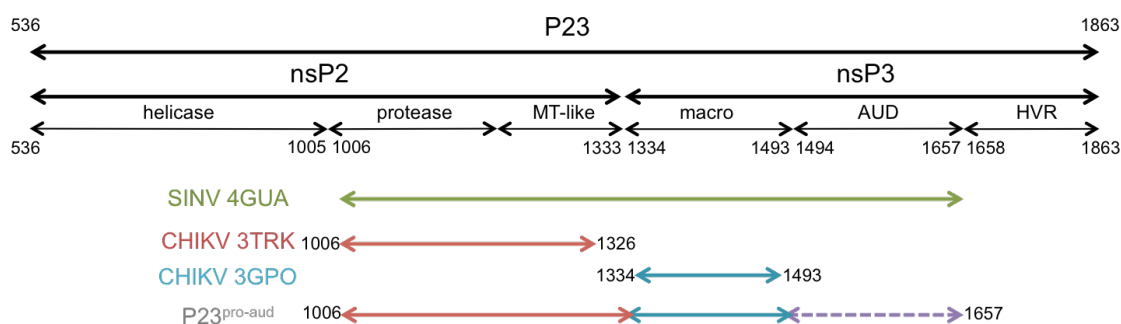


Figure 41: Schematic description of P23, nsP2 and nsP3 with available structural information

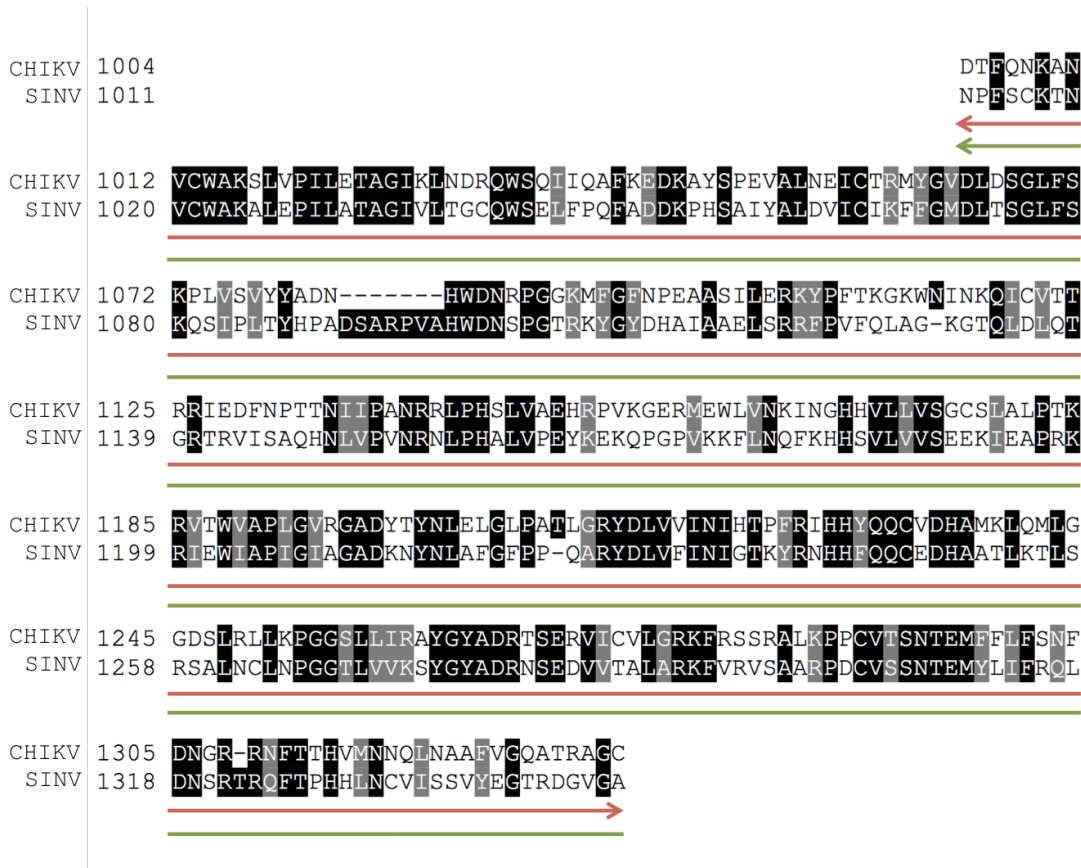


Figure 42: nsP2 protease and methyl-transferase like domain (3TRK in red, 4GUA in green)

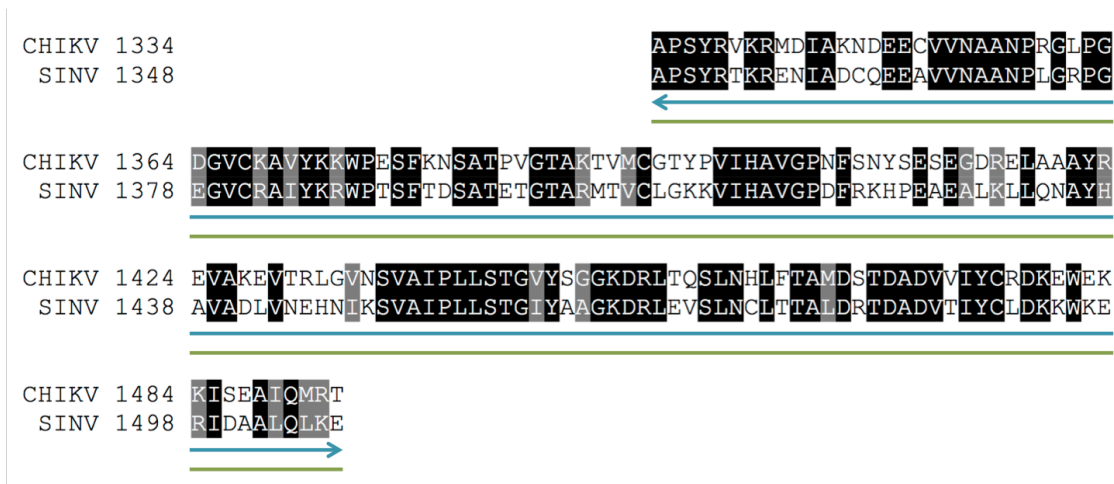


Figure 43: nsP3 macro domain (3GPO in blue, 4GUA in green)

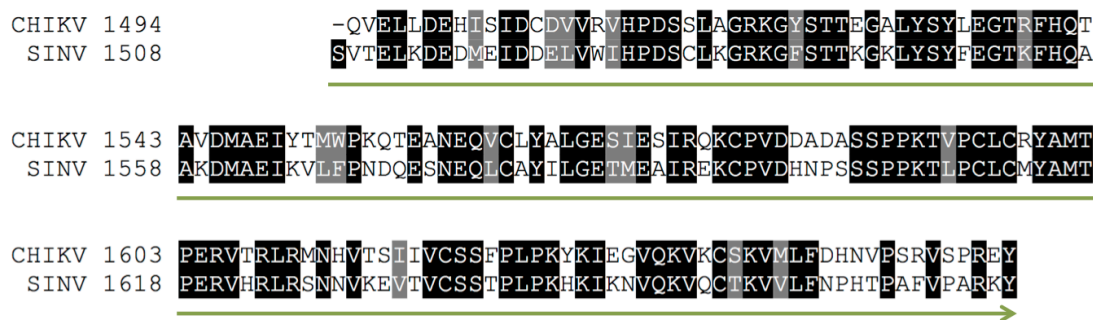


Figure 44: nsP3 alphavirus unique domain (4GUA in green)

The zinc ion present in the structure of SINV was included because it might play a role for the stability of the domain. It was used as environment for induced fit in order to form the relevant interactions with the cysteine residues (Cys 1595,1597,1620,1638) complexing it in the model and in the template.

10 intermediate models were generated and refined with the Generalized Born/Volume Integral (GB/VI) methodology provided in MOE (Labute 2008b). A “medium” setting of refinement was chosen for the intermediate models and no refinement was used for the generation of the final model due to the high similarity to the template. The force field Amber99 suitable for protein modelling was chosen (Wang et al. 2000).

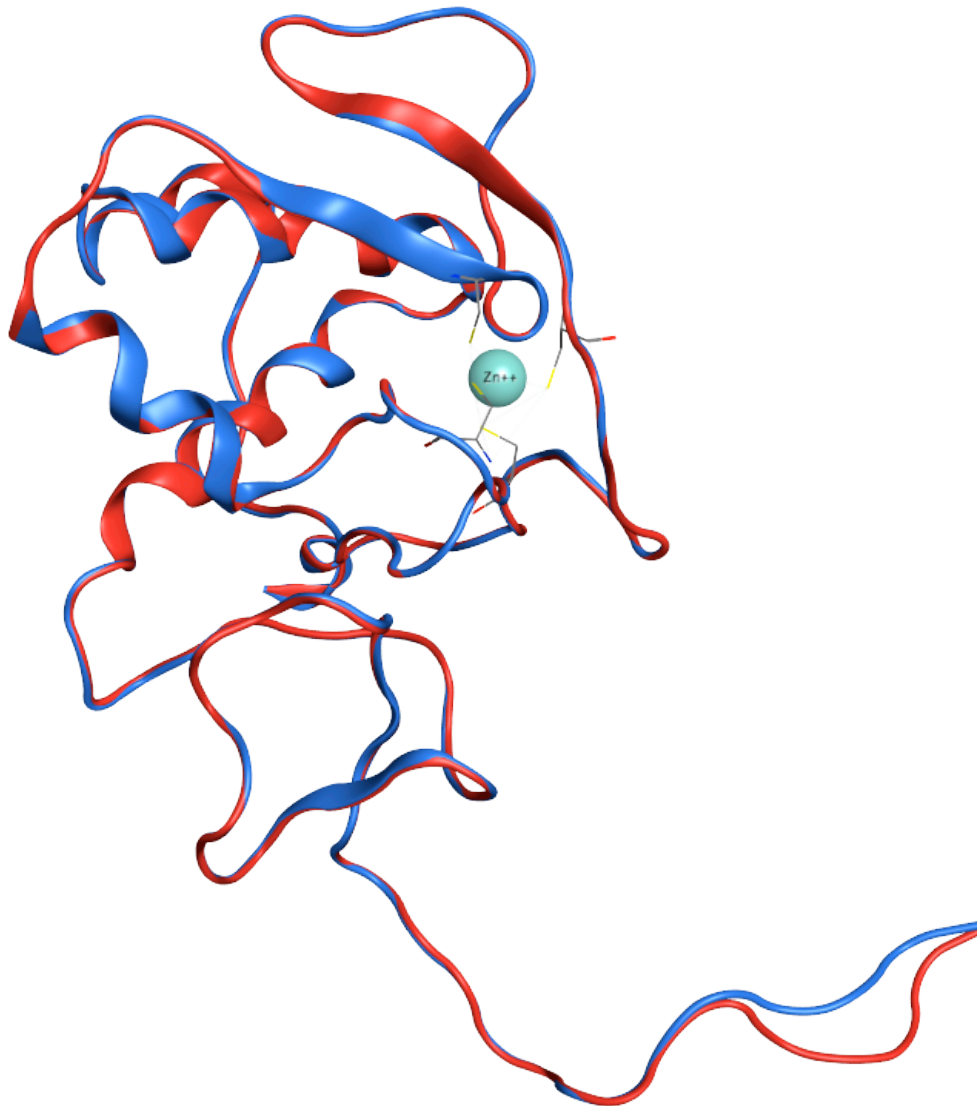


Figure 45: Alphavirus unique domain Homology model in blue and SINV template in red

2.10.1.2 Evaluation of the homology model

After the generation of a model it was investigated in order to eliminate errors in the structure that could influence future work with the model. It has to be noted that crystal structures are analysed and refined with computational models and not all experimental data can be represented and refined correctly based solely on these computational methods. Therefore it is necessary to investigate the crystal structure used as template and to compare the model with the template structures.

First, the present model was visually inspected with special focus on the cysteine residues complexing the zinc ion and the overall similarity of the structure to the template. The RMSD between the model and the template was calculated, a value indicating the mean structural distance between each pair of backbone atoms. Low values indicate that the model is very

similar to the template. The RMSD for this model was 0.29 Å indicating a very close similarity to the template.

Ramachandran or phi/psi plot is mapping the permitted geometries of the dihedral angles in a protein backbone and can therefore be used to assess the quality of a crystallised protein or homology model. Analysis of the Ramachandran plot showed only two amino acid outliers in the model: the first (L1497) being located in the loop at the N-term of the model and the second one is glutamic acid E1570 which is located between two helices on the outside of the domain. For detailed investigation in these two regions the geometry of the backbone will be adjusted but for the moment the two outliers were ignored.

Finally, the model and the crystal structure were loaded into the structure analysis server SAVES. It contains several evaluation tools: PROVE, ERRAT, VERIFY3D and WHATCHECK. Results of the evaluation of the crystal and the template were comparable showing the good quality of the model (Molecular Biology Institute at the University of California 2016).

2.10.2 Expanding nsP3

After the generation and evaluation of the homology model of the AUD the next aim of this project was to assemble the precursor protein P23 for CHIKV. Therefore the available crystal structures for nsP2 protease (3TRK) and nsP3 macro domain (3GPG, 3GPO, 3GPQ) were downloaded from the protein databank PDB. The sequence for CHIKV P1234 was already present from the homology model and was used to align the sequences reported in the crystal structures to create the correct spatial arrangement of the distinct domains. The structure of the homology model was treated like a crystal structure and equally aligned (see above). In order to assemble the complete precursor protein, the loop connecting nsP2 with nsP3, and within nsP3 the connective loop between the macro- and the alphavirus unique domain, had to be built or modelled.

The loop within nsP3 bridging macro domain and AUD did connect directly and only the geometry of the bond had to be properly adjusted for which the loop modeller of MOE was used. The two amino acids Thr 1493 and Gln 1494 had to be connected and were therefore selected from the sequence as query loop. MOE performs a PDB search for the query sequence and gives a list of possible loop geometries found in high-resolution crystal structures. The best loop was then chosen to connect the two domains.

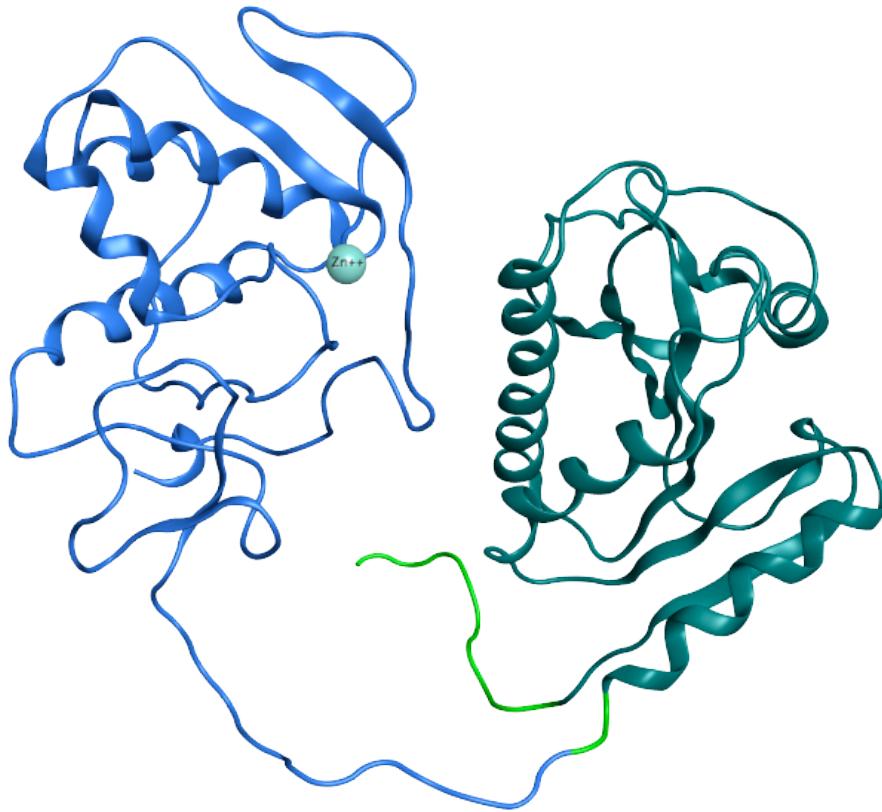


Figure 46: Crystal structure and homology model of CHIKV nsP3 linked

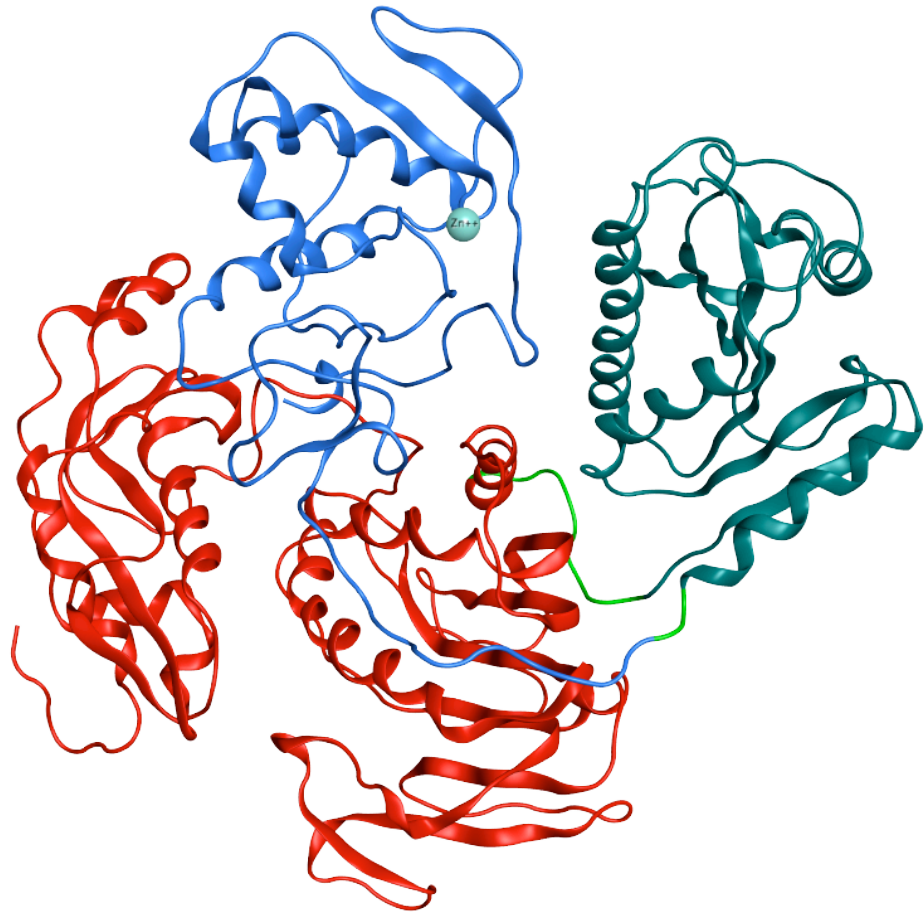
Macro domain in teal; AUD in blue; Connective loops in bright green

2.10.3 Linking the P23 precursor

For the connection between nsP2 and nsP3 a different approach had to be used because the connective loop was too long to serve as query for the loop modeller. Instead, a homology modelling approach was applied. The homology modeller of MOE allows for parts of the query sequence to be overwritten with known structural information. Here, the sequence for nsP2 and nsP3 were overwritten with the crystal structure except for the unknown connective loop: VGQATRAGC (1325-1333).

For that sequence a homology model was built using again the SINV template. Parameters for the homology model were the same as for the model of the AUD. RMSD for the loop is 0.57 Å when compared to the SINV template, which is close to the template and a good value for a loop region.

The final complex was then assembled and is depicted in **Figure 47**. The connective loops are depicted in green, nsP2 in red, nsP3 macro domain in teal and nsP3 AUD in blue.



**Figure 47: Assembly of the precursor P2^{pro}3^{AUD} nsP2 is depicted in red, macro domain in teal and AUD in blue.
Connective loops are shown in green**

2.10.4 Future perspectives

The parts of the precursor presented in this chapter will be investigated in the following section to find new antiviral compounds targeting the precursor.

The helicase domain was very recently crystallised (PDB: 6JIM) (Law et al. 2019) but linking it with the present protein domains and the precursor protein built in this section would be an interesting task towards the elucidation of the structure of the replication complex and the polyprotein precursor.

2.11 The nsP2/nsP3 cleavage site pocket as target

The alphaviral non-structural proteins are synthesised from the open reading frame one (ORF1) of the genomic RNA as one single polyprotein precursor in the form P1234 or P123+4. These polyproteins form the early stage viral replication complex. The internal protease in the nsP2 part of the polyprotein cleaves the polyprotein into the individual non-structural proteins or smaller precursor proteins. The protease cuts the P1/2 site of the same protein strand (*cis*-cleavage) resulting in nsP1+P23+nsP4. The P2/3 cleavage is very efficiently processed by nsP2 protease, so that P23 is nearly undetectable in wild type infections. The cleavage of the P2/3 bond marks also the switch from the early to the late stage replication complex and is likely followed by significant conformational changes and protein rearrangements. It is furthermore necessary to switch RNA template uniquely towards the minus strand template losing the function to synthesise minus stranded RNA and uniquely producing positive stranded genomic and subgenomic RNA (Strauss & Strauss 1994; de Groot et al. 1990; Shirako & Strauss 1994).

Hindering the protease from cleaving the precursor protein might therefore impede the viral replication significantly. Two possible ways to block the cleavage site seem reasonable: First to block the protease itself by inserting a molecule that mimics the natural substrate of the protease but cannot be cleaved and therefore blocks the proteases function as molecular scissors. This approach has been explored previously in our lab and others leading to several antiviral agents against CHIKV (Bassetto et al. 2013; Das et al. 2016).

A second approach is constituted by the rationale of masking the cleavage site and hindering the protease to bind to the loop that it is supposed to cleave. In the previous section the computational construction of the precursor protein of P23 was presented. That molecular complex was then investigated to understand the cross talk between the different domains but also to find new attractive targets for structure based drug design.

2.11.1 The P23 cleavage site

The non-structural proteins nsP2 and nsP3 are lying in the central part of the ORF1 and form the middle piece of the polyprotein precursor P1234. The protease within nsP2 cleaves the polyprotein strand between cysteine 1333 and alanine 1334 in *trans* (which means on a different polyprotein strand) (Shin et al. 2012).

The Figure 48 below show different views of the P23 spanning the protease and methyltransferase-like domain of nsP2 (the N-terminal helicase domain is missing – a crystal structure (6JIM) was published in 2019 after the completion of the practical work on this thesis (Law et al. 2019)) and the macro domain and AUD or zinc-binding domain of nsP3 (the C-

terminal hypervariable domain is still lacking any structural information for CHIKV or any related virus).

The cleavage site between nsP2 and nsP3 lies on a loop between the methyltransferase-like domain and the macro domain. The linker between the macro and the zinc-binding domain contributes to the formation of a deep pocket that might allow the cleaving protease to circumvent the cleavage loop to a certain extent.

This pocket might present a target for drug design as it is big enough to easily accommodate a ligand of drug-like size, it is solvent accessible and it is well on the interface between the two domains, so that a ligand might be able to crosslink the two domains or restrict the flexibility of the contributing loops if the cleavage still occurs or prevent the protease to access the cleavage site by steric hindrance.

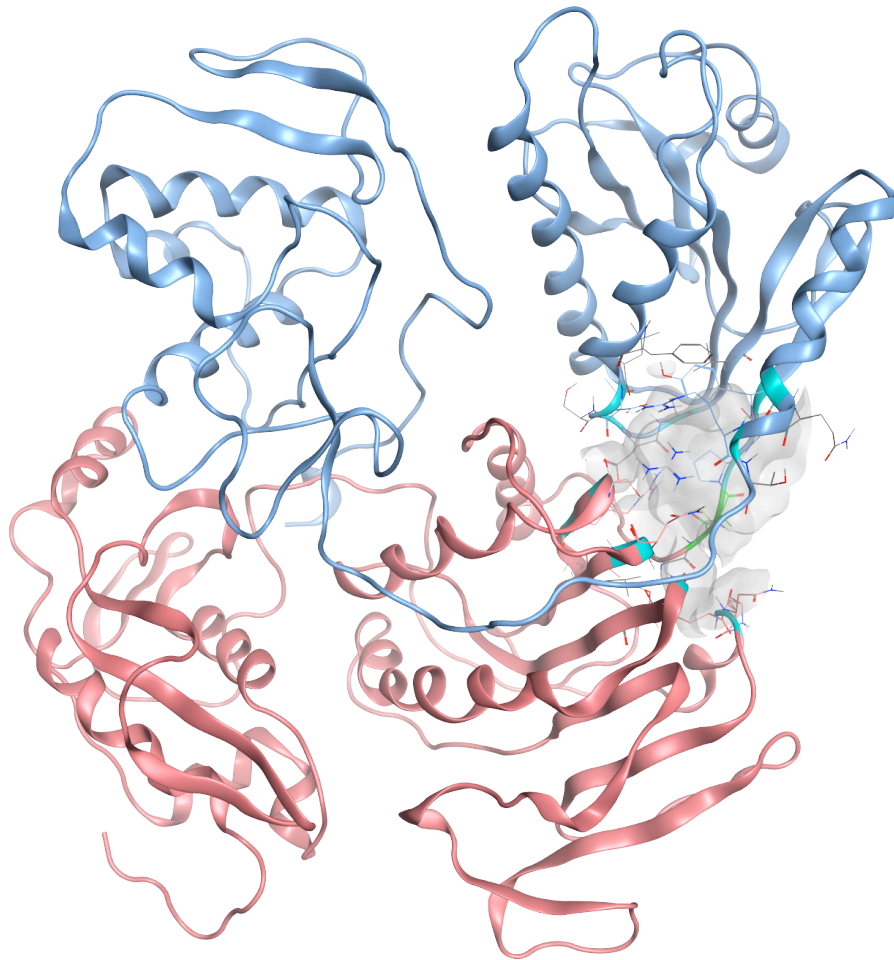


Figure 48: CHIKV P23 precursor (nsP2 part in red, nsP3 part in blue)

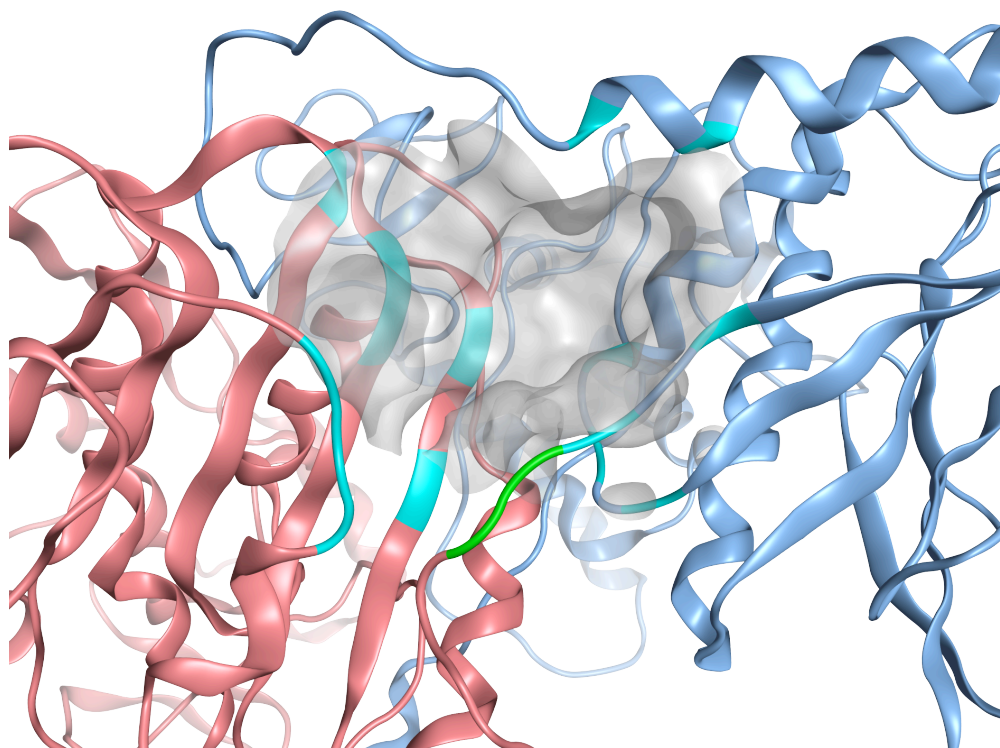


Figure 49: P23 cleavage site pocket

2.12 Aims

The cleavage between nsP2 and nsP3 has proven to be crucial for the complete template switch from positive strand to negative stranded RNA of the virus and therefore to proceed towards the late stage of the viral replication. Interfering with this step of the viral life cycle might lock the RNA replication process at the early stage and halt the generation of new virions.

The cleavage site pocket presents therefore an interesting target. The aim was to investigate this pocket with computational methods and to find new molecules that could interfere with the cleavage but differing clearly from classical protease inhibitors by address the substrate site and not cleaving enzyme.

2.13 Results and Discussion

2.13.1 Site Finder

Site Finder in MOE has revealed several interesting pockets on the P23 precursor of which pocket Nr 3 was chosen (according to the Site Finder ranking). It is defined as depicted in Figure 50. 53 residues form the deep pocket adjacent to the nsP2/nsP3 cleavage site that is depicted as green ribbon. Within the pocket dummy atoms were generated on the centres of the alpha spheres defining the size and shape of the pocket. By visual evaluation one dummy atom in the middle of the binding pocket was selected to become the central point for the grid

box that defines the area where compounds can be positioned in a HTVS or in docking studies. The central dummy atom is coloured in green (see Figure 50).

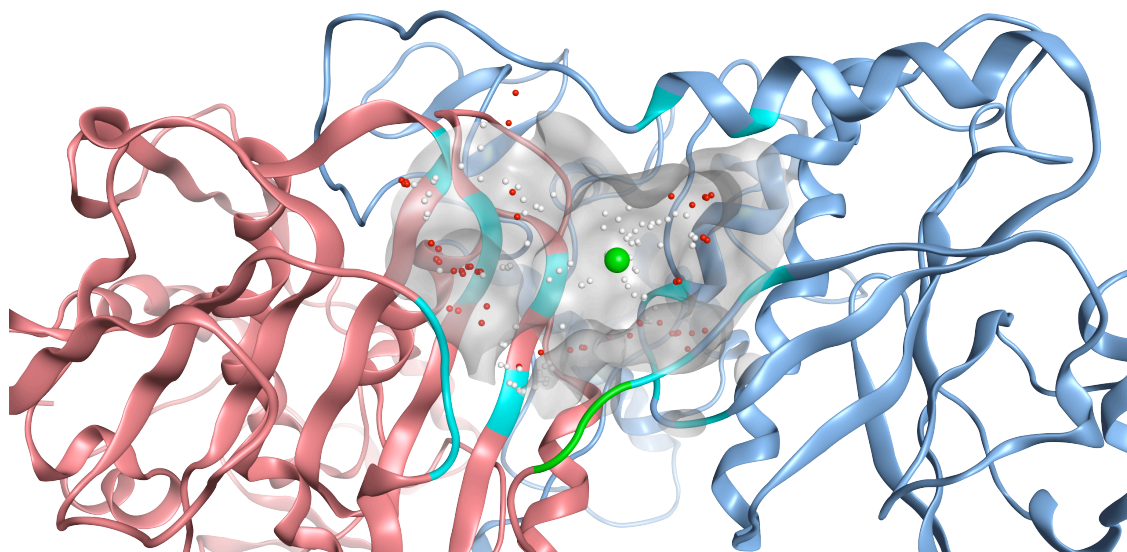


Figure 50: P23 cleavage site pocket with the central dummy atom in green

2.13.2 Virtual Screening workflow

In order to find new molecules to block the P23 cleavage site pocket the SPECS library (www.specs.net) was screened using the high-throughput virtual screening (HTVS) tool in Maestro Glide (Friesner et al. 2004; Schrödinger LLC 2016). The SPECS library was previously prepared in our group and consisted of 389,456 structures in the version used. These structures were then screened, evaluated and selected as described in the workflow below (Figure 51).

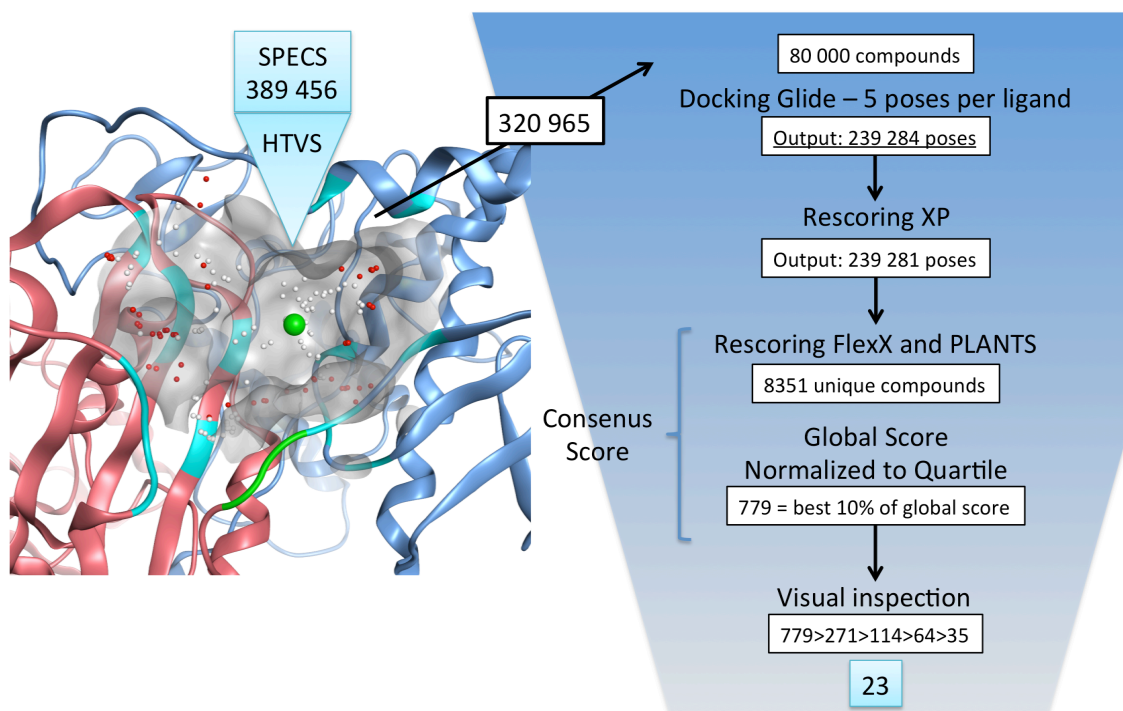


Figure 51: HTVS workflow to select compounds addressing the P23 cleavage site pocket

Of the 389,456 input structures 320,965 were placed successfully in the selected pocket and the best 80,000 compounds were selected (this corresponds roughly to 25%) for further docking studies. The docking was performed with Maestro Glide in the standard precision (SP) mode writing three poses per molecule and allowing a post docking minimization.

239,284 poses were generated in this step and put forward for the scoring evaluation with Maestro Glide in extra precision (XP) mode where each of the docking poses is ranked in comparison with all the other poses resulting from the SP docking. The poses from the SP docking were then also ranked with two other scoring functions: FlexX and the scoring function of PLANTS.

A consensus score was created taking into consideration the Glide XP score, the FlexX score and the PLANTS score. This strategy was previously described in this thesis. All poses that were ranked into the first quartile (numerically calculated) were assigned the number 1, all poses ranked worse than the first quartile were assigned the number -1. The quartile itself receives the number 0. This mathematical operation was carried out for all the three scoring functions. In the end the sum of the three values for each pose was calculated assigning the number 3 to all poses that were ranked above the quartile for each individual scoring function. In any case where not all scores were above the individual quartiles the pose was discarded. Then the poses were filtered to yield only unique molecules in MOE and kept in the order obtained before. 8,351 molecules were kept.

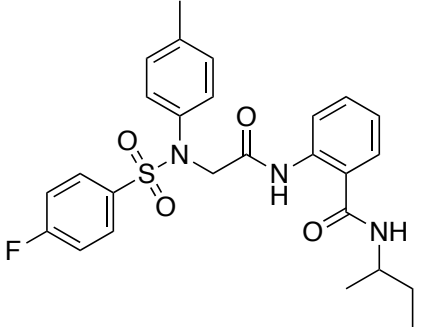
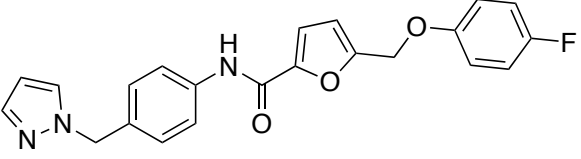
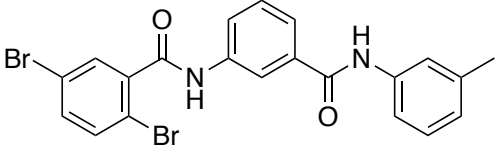
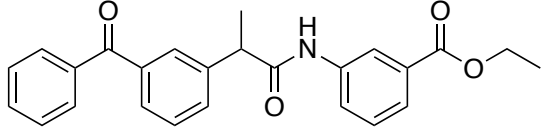
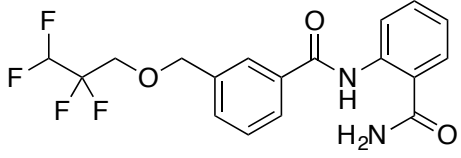
To reduce the number further the Lipinsky druglike descriptor was calculated with MOE. 8,090 molecules were suiting this criterion, which is not surprising as the SPECS library is generally designed to fit the druglikeness criteria. Then, an absolute score was calculated. First, the individual score was normalised to the quartile of each scoring function separately. Then, these values were summed up to result in an absolute score, giving the absolute ranking of the compounds. 779 compounds were selected according to their absolute score (which is roughly 10%).

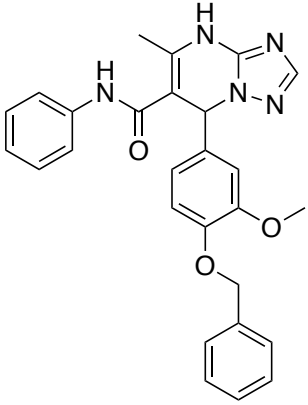
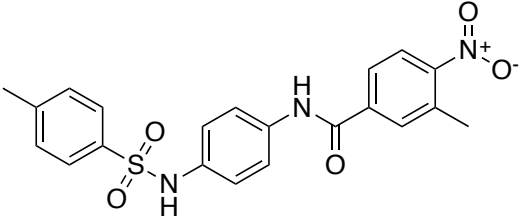
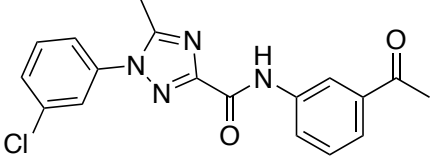
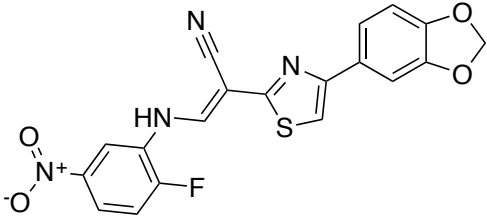
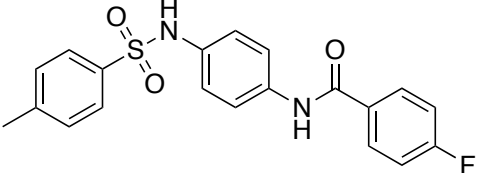
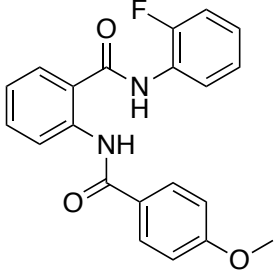
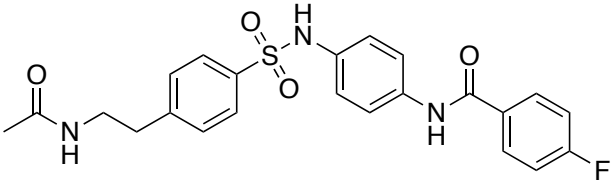
These 779 compounds were then inspected for their pose and their interactions within and with the binding pocket. For the selection of the final compounds mainly the interaction with key residues was taken into account.

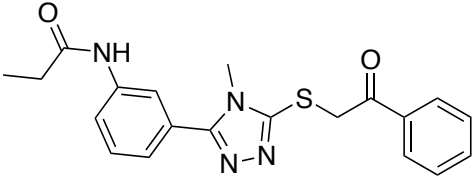
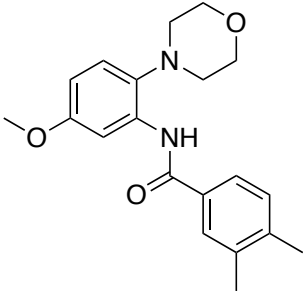
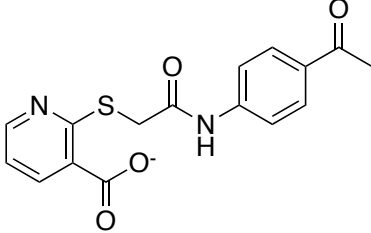
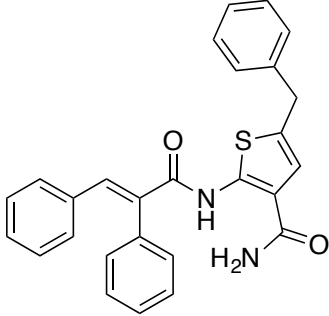
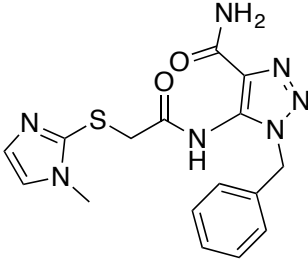
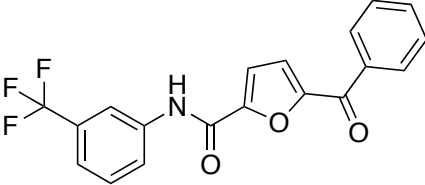
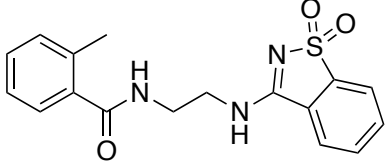
The 23 compounds depicted below were selected, purchased and sent for biological evaluation to our collaborators at the Rega Institute for Medical Research, KU Leuven, Belgium.

2.13.3 Selected compounds

Table 69 Compounds selected for the P23 binding site

Nr	Structure	SPECS code
91		AO-081/41758678
92		AK-968/41170770
93		AN-652/43161813
94		AN-652/42190905
95		AK-968/40709730

96		AF-399/14218098
97		AK-968/41922711
98		AP-501/43409119
99		AE-848/15341257
100		AK-968/11735588
101		AP-064/41684920
102		AO-080/43378516

103		AN-648/15240099
104		AP-970/43481292
105		AO-080/42945061
106		AK-968/41922216
107		AO-476/43417657
108		AG-205/36957217
109		AN-988/41349278

110		AE-848/11420971
111		AK-968/12264015
112		AK-968/11533105
113		AK-968/41922624

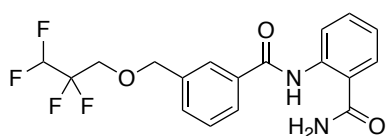
2.13.4 Biological evaluation

The 23 compounds were purchased and sent to our collaborators at the Rega Institute for Medical Research, KU Leuven, Belgium. The CPE reduction assay was performed to assess the antiviral activity and the cytotoxicity of the compounds.

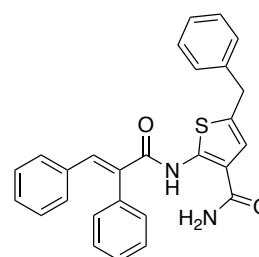
Table 70: Activity and cytotoxicity of the tested compounds

Compound	EC ₅₀ (μ M)	EC ₉₀ (μ M)	CC ₅₀ (μ M)	Comment
91	> 201	> 201	>201	
92	>3,5	> 3,5	3,5 \pm 0,2	data of manual assay not included unreliable data
93	> 68	> 68	>68	
94	> 130	> 130	>130	
95	2,4	>16	16 \pm 13	

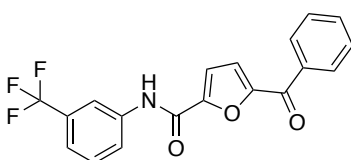
96	> 27	> 27	>27	
97	> 20	> 20	>20	
98	>29	> 29	29 ± 3,9	data of manual assay included
99	> 122	> 122	>122	
100	> 87	> 87	>87	
101	>34	>34	>34	data of manual assay included
102	> 27	> 27	>27	
103	44 ± 7,2	>87	87 ± 22	
104	71 ± 4,6	>121	121 ± 7,8	data of manual assay included
105	> 304	> 304	>304	
106	4,8 ± 1,4	4,65	24 ± 24	data of manual assay included
107	89	> 90	> 90	data of manual assay included
108	< 1,0	< 1,0	1,0 ± 0,7	manual assays failed; crystals
109	204	>259	259	data of manual assay included
110	11 ± 0,8	15 ± 1,5	23 ± 6,5	data of manual assay not included (EC50 = 81 µM in manual assay)
111	250 ± 1,4	>281	274 ± 11	data of manual assay included
112	>39	>39	>39	data of manual assay included
113	8,3	>15	15 ± 1,4	



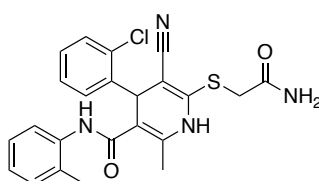
95



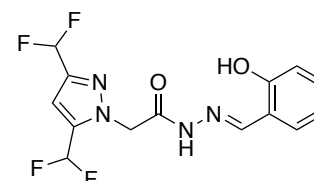
106



108



110



113

In a first assay five compounds showed some antiviral effects but these could not be confirmed by further repetitions. The compounds for which the assay was not repeated did not show a

sufficient activity vs. cytotoxicity ratio for them to be considered for a medicinal chemistry investigation and derivatisation.

2.14 Discussion and Conclusions

The linkage of nsP2 and nsP3 to study the P23 precursor was possible due to the available crystallographic data. Being the last part of the precursor protein to be cleaved and by that marking the template switch to minus strand RNA makes the P23 precursor protein an interesting cleavage product to study. Inhibiting CHIKV nsP2 protease and thereby the cleavage of the polyprotein precursor of the non-structural proteins has proven a good strategy to inhibit the viral replication (Bassetto et al. 2013; Das et al. 2016; Singh et al. 2018).

With the discovery of a pocket in the vicinity of the cleavage site between nsP2 and nsP3 and the rationale that preventing the protease from optimal positioning around this cleavage site the cleavage should be prevented as well, the pocket was used for a HTVS in order to find molecules that could fit perfectly into the pocket.

After the selection of the compounds and the tests in the cell-based antiviral assay they resulted to be inactive. Several factors must be taken into consideration at this point.

First, the selected pocket is found on the interface where the P23 protein was assembled. The connective loops that were modelled are contributing to the pocket. So far there is no crystal structure available for P23 for CHIKV and therefore the model cannot be confirmed by experimental data.

Second, the selected compounds were not active in the CPE assay. Several factors could contribute to this. One of problems often encountered in this work is a problem with solubility in the aqueous media necessary for cell culture. In the microscopic evaluation some compounds were interpreted as cytotoxic even if the compound was precipitating in higher concentrations, because there is no separate score for this. This affects the data and leaves some results to be questioned. We tried to come by this by retesting the compound and giving special attention to any precipitation, but none of the compounds showed any activity.

In the hypothetical case that any of the compound would have been active, activity does not hint that the P23 cleavage site pocket is the target. Biochemical evaluation via binding assays or virological resistance selection could have been a way to investigate the target for the compounds further.

2.15 Experimental

2.15.1 Hardware

All computational calculations were carried out on an 8-core computer with Intel Xeon 1.80 GHz E5-2403 v2 CPUs running Ubuntu 14.04 LTS operating system in the 64-bit version. For the HTVS, Docking and Rescoring tasks only 4 processors were used.

2.15.2 Protein preparation

The structure of the protein was obtained from the previous homology modelling and linking of the nsP2 and nsP3 parts prepared in before in this work. Missing hydrogen atoms and the correct protonation states of the protein were generated with the default settings in the Protein Preparation tool in MOE 2015.10 (Chemical Computing Group Inc. 2016). All water molecules were removed from the structure.

2.15.3 Site Finder

The Site Finder application integrated in MOE 2015.10 was used to identify promising pockets on the P23 precursor. The pockets were inspected in MOE and dummy atoms were generated on the centres of the alpha spheres that define the pocket. The dummy atom in the centre of the desired pocket was then transformed to a carbon atom and its coordinated were used to define the centre of the box in which the selected compounds will be docked. The structure was saved with and without the carbon atom both in .pdb format in order to be used by the different software packages used for the docking and rescoring of the compounds.

2.15.4 SPECS library

As library for the search for new potential molecules that fit into the query pocket the SPECS library was used (www.specs.net). The SPECS 2016 version was used and all the compounds were prepared using the LigPrep function of Schrödinger Maestro and the OPLS_2005 force field (Schrödinger LLC 2016). For each molecule tautomers and the possible ionisation states at pH 7 ± 2 were generated whilst keeping the chirality of the initial input molecules.

3 Enteroviruses

3.1 Introduction

3.1.1 Classification

The family of picornaviruses is one of the oldest and most diversified families of viruses. Enteroviruses belong to the family of *Picornaviridae* within the order of *Picornavirales*. *Picornaviridae* currently consist of 94 species grouped into 40 genera (as of February 2018) (Zell et al. 2017). Among those the genus enterovirus consists of 15 species: enterovirus A-L and rhinovirus A-C. The seven species that contain human pathogens are enterovirus A-D and the rhinovirus species. The classification and names of the members of the *Picornaviridae* are rapidly changing and expanding to accommodate new findings mostly from sequencing data. Furthermore, the old classification via antigenicity or clinical features proved unpractical. Seven new genera and 16 new species have been recently proposed to the ICTV and are currently to be classified (Lefkowitz et al. 2018). This will bring the total number of genera to 47 and the number of species to 110. To reduce the complexity, in this thesis mostly the human pathogenic enterovirus and rhinovirus genera are discussed. The representatives investigated more in detail are printed in **bold** in Table 71.

Table 71: Enterovirus representatives

Genus (nr. of serotypes)	Species
Enterovirus A (25)	Coxsackievirus A2 (CVA2), CVA3, CVA4, CVA5, CVA6, CVA7, CVA8, CVA10, CVA12, CVA14, CVA16, enterovirus A71 (EV-A71), EV-A76, EV-A89, EV-A90, EV-A91, EV-A92, EV-A114, EV-A119, EV-A120, EV-A121 and the simian enteroviruses SV19, SV43, SV46 and baboon enterovirus A13 (BA13)
Enterovirus B (63)	Coxsackievirus B1 (CVB1), CVB2, CVB3 , CVB4 (incl. swine vesicular disease virus 2 [SVDV-2], CVB5 (incl. SVDV-1), CVB6, CVA9, echovirus 1 (E1; incl. E8), E2, E3, E4, E5, E6, E7, E9 (incl. CVA23), E11, E12, E13, E14, E15, E16, E17, E18, E19, E20, E21, E24, E25, E26, E27, E29, E30, E31, E32, E33, enterovirus B69 (EV-B69), EV-B73, EV-B74, EV-B75, EV-B77, EV-B78, EV-B79, EV-B80, EV-B81, EV-B82, EV-B83, EV-B84, EV-B85, EV-B86, EV-B87, EV-B88, EV-B93, EV-B97, EV-B98, EV-B100, EV-B101, EV-B106, EV-B107, EV-B110 (from a chimpanzee), EV-B111, EV-B112 (from a chimpanzee), EV-B113 (from a Mandrill) and the simian enterovirus SA5
Enterovirus C (23)	Poliovirus (PV) 1 , PV2, PV3, coxsackievirus A1 (CVA1), CVA11,

	CVA13, CVA17, CVA19, CVA20, CVA21, CVA22, CVA24, EV-C95, EV-C96, EV-C99, EV-C102, EV-C104, EV-C105, EV-C109, EV-C113, EV-C116, EV-C117 and EV-C118
Enterovirus D (5)	EV-D68 , EV-D70, EV-D94, EV-D111 (from both humans & chimpanzees) and EV-D120 (from gorillas). Human rhinovirus (HRV) 87 has been reclassified as a strain of EV-D68
Rhinovirus A (80)	Rhinovirus (RV) A1, A2 , A7, A8, A9, A10, A11, A12, A13, A15, A16, A18, A19, A20, A21, A22, A23, A24, A25, A28, A29, A30, A31, A32, A33, A34, A36, A38, A39, A40, A41, A43, A45, A46, A47, A49, A50, A51, A53, A54, A55, A56, A57, A58, A59, A60, A61, A62, A63, A64, A65, A66, A67, A68, A71, A73, A74, A75, A76, A77, A78, A80, A81, A82, A85, A88, A89, A90, A94, A96, A100, A101, A102, A103, A104, A105, A106, A107, A108, A109
Rhinovirus B (32)	Rhinovirus (RV) B3, B4, B5, B6, B14 , B17, B26, B27, B35, B37, B42, B48, B52, B69, B70, B72, B79, B83, B84, B86, B91, B92, B93, B97, B99, B100, B101, B102, B103, B104, B105 & B106
Rhinovirus C (56)	Rhinovirus C1-56

3.1.2 Epidemiology and pathogenesis

Enterovirus genera contain many important human pathogens. The best known is undoubtedly poliovirus, but also other types like coxsackieviruses, echoviruses, the numbered enteroviruses and rhinoviruses are a considerable health threat on societies around the globe. Enteroviruses, as their name suggests, are transmitted via the gastro-enteral route but do not normally cause gastro-enteral symptoms (Zell et al. 2017). Rhinoviruses on the contrary are spread via respiratory transmission and are the main cause for the common cold (Jacobs et al. 2013). In immunocompetent adults infections usually remain unnoticed or elicit only mild symptoms that are usually limited to a couple of days. In young children and immunocompromised subjects enterovirus infections can lead to serious complications and even fatal outcomes. The most common serious manifestations are aseptic meningitis, neonatal sepsis-like disease, encephalitis, acute flaccid paralysis, non-specific febrile illness, hand-foot-and-mouth disease (HFMD), herpangina, pleurodynia, pericarditis and myocarditis. Pons-Salort and co-workers summarised the different diseases and their causative agent in a neat table that can be used for quick reference (Baggen et al. 2018; Pons-Salort et al. 2015; Tapparel et al. 2013).

As a representative of Enterovirus A, EV-A71 usually leads to HFMD. It has also been found in patients with neurological complications such as brainstem encephalitis, meningitis and

poliomyelitis-like paralysis. It was first characterized in 1969 in California after an outbreak with neurological complications in the USA. From there it spread to Europe during the mid 1970ies and further to Asia. A detailed and very comprehensive review of the clinical aspects of EV-A71 infections was provided by Ooi and co-workers (Ooi et al. 2010).

Some echoviruses and coxsackieviruses might in addition to the previously mentioned symptoms cause also an inflammation of pancreatic β -cells and therefore lead to type-1 diabetes (Tracy et al. 2010). Apart from the common cold, the group of rhinovirus infections can trigger severe respiratory tract complications like exacerbations of asthma and chronic obstructive pulmonary disease (COPD) as extensively discussed in the review by Jacobs and colleagues (Jacobs et al. 2013).

3.1.3 Transmission

Enteroviruses usually transmit from human to human through close contact of eye, nose and mouth fluids, stool or the secrete that can be found in blisters occurring due to enterovirus infection. Direct contact with the virus is necessary for infection, but also touching contaminated surfaces etc. a subsequent contact with eyes, mouth or nose can lead to infection (CDC 2019). Rhinoviruses can be spread also by droplet infection like sneezing or coughing.

3.1.4 Current treatment

Due to the high diversity of enteroviruses the development of drugs against several of them would be the most desirable option. With the Poliovirus vaccine we are moving towards the eradication of Polio, but achieving this goal is still challenging (GPEI 2017; GPEI 2018). Recently two Enterovirus vaccines were approved in China (Yi et al. 2017). But direct acting antivirals are still lacking. Several compounds have been found to inhibit enteroviruses in vitro and in vivo, some of which were used to treat enteroviruses for the repurposing of the compounds for this new indication but none of them was yet approved for the new indication (Baggen et al. 2018). A list of case studies and reports about off-label or compassionate use of were performed that report the outcomes of these treatments, which in several cases give hope to find an approved drug relatively soon. However the very small sample size and the specific circumstances such as different clinical manifestations, diverse enterovirus strains, and the condition of the patient itself, must be taken into account and therefore do not allow extrapolation to a larger group of people (Gofshteyn et al. 2016; Messacar et al. 2019).

3.1.5 Genome organisation and replication cycle

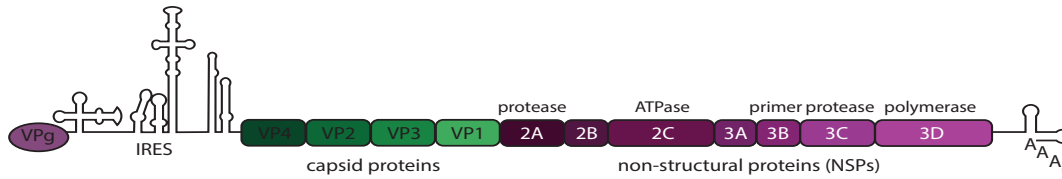


Figure 52: Enterovirus genome
Capsid proteins in green, non-structural proteins in purple
with permission of Denise Seitner

The genome organisation of enteroviruses is schematically depicted in Figure 52. Enteroviruses have a linear single stranded positive sensed genome varying in length from 7.1-7.4 kilobases. In contrast to other (+) ssRNA viruses, which contain a methylated cap structure, the genome contains a viral protein structure called (VPg) at its 5' end. Also at the 5' terminus resides an internal ribosome entry site that is important for the initiation of the translation of the viral polyprotein precursor. The regions encoding structural or capsid proteins are depicted in green and the non-structural regions are coloured in purple. The 3' end contains a polyadenylated tail.

The single open reading frame is translated into one polyprotein, which is cleaved in a stepwise manner into ten individual proteins by the viral proteases 2A^{pro} or 3C^{pro} and autocatalytic processing (Zell et al. 2017).

3.1.6 Replication cycle of enteroviruses

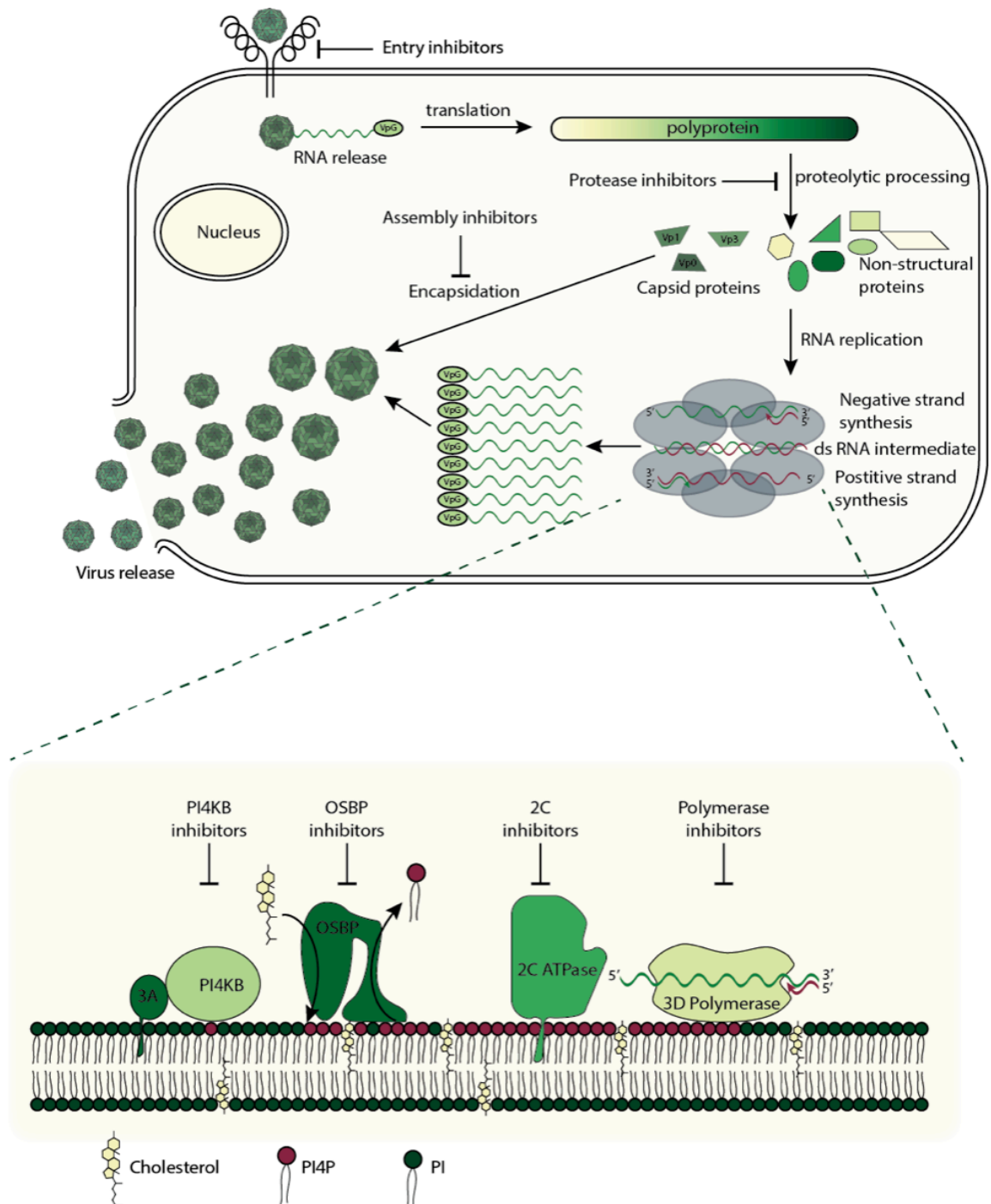


Figure 53: Replication cycle of enteroviruses and a selection of crucial host factors
Image with permission of Denise Seitner

Enteroviruses are around 30 nm in diameter and are not individually enveloped with parts of the host's cell membrane and do not possess glycoproteins. The viral life cycle starts when a viral particle binds to receptors on the membrane of the host cells and these interactions further lead to the endocytosis of the particles (van der Linden et al. 2015). The different enteroviruses are using different cell surface receptors, which are described in more detail in the attachment and entry section. A review by Tuthill and co-workers describes the different

picornavirus entry processes and endocytosis steps in great detail (Tuthill et al. 2010). Depending on the receptors the caveolin, clathrin or not-mediated endocytosis leads to the ingestion of the particle by the host cell. Via receptor binding or pH changes in the endosome the capsid undergoes conformational changes. VP4 is ejected from the capsid and the N-terminus of VP1 is exposed to the surface of the particle resulting in 135S or so-called A particles, which due to an amphipathic helix can fuse with the endosomal membranes. VP4 is also inducing or maintaining a pore within the endosomal membrane, which is crucial for the release of the viral genome into the host cytoplasm. The following steps are very similar for most of the enteroviruses. First, the positive sense genome can be readily translated into one single polyprotein precursor, which is eventually cleaved into ten individual proteins by the viral proteases: into three capsid proteins (VP0, VP1 and VP3) and seven non-structural proteins responsible for the viral replication (2A-C and 3A-D). Some of the intermediate products act in the viral replication process even before being fully cleaved. Already the first polyprotein products are sufficient to replicate the genome and soon after interfere with cellular components to transform the host cell into a virus-producing factory (van der Linden et al. 2015).

For enterovirus replication to successfully take place, the virus induces the rearrangement of intracellular membranes, which then contribute to the so-called replication organelle. Host proteins residing in these membranes are hijacked and utilized by the virus to enhance its replication (Baggen et al. 2018).

To the incoming positive sense genome a complementary negative strand is synthesised by the viral RNA-dependent RNA-polymerase ($3D^{pol}$), which results in a double stranded RNA intermediate. Then the negative strand is used to generate more positive strand RNA, which can serve as a template for translation into viral proteins or as genomic RNA that will be encapsidated into viral capsid proteins to form nascent virions.

The encapsidation and the formation of new virions is a tightly regulated process. First, the three envelope proteins are arranged into protomers. Then the protomers form pentamers, which together with the positive sense genome assemble into provirions. The maturation of the provirions is induced by the viral genome, which induces a rearrangement of VP0 into VP2 and VP4, and the mature virions are usually released via the lytic pathway (Baggen et al. 2018).

3.1.7 Targets and Inhibitors

In order to inhibit virus entry or replication several crucial steps in the viral life cycle or individual viral proteins can be targeted. This approach is preferred as side effects on the human cells can usually be reduced. Another way to disrupt the virus replication cycle is to interfere with essential host factors that enteroviruses rely upon (Baggen et al. 2018). This

section describes the targets under investigation for drug development and some representative inhibitors.

3.1.7.1 Attachment and entry and inhibitors thereof

For the virus to enter the cells there are different pathways. The entry process was reviewed in detail by Tuthill et al. (2010). Usually cell surface receptors are involved as in the case of enteroviruses. Which receptor is used largely depends on the different virus and the availability of the receptor on the cell. Many viruses are rather promiscuous when it comes to their receptor and can utilise different receptor gateways depending on their availability and cell type. The first enterovirus receptor that was discovered was the so-called poliovirus receptor (CD155). Another important molecule at the cell membrane is intracellular adhesion molecule 1 (ICAM-1), the intercellular adhesion molecule 1. Many more receptors were found with the advances in research and many of them can be categorised into the immunoglobulin-like or integrin receptor family. Cellular receptors can be involved in viral attachment or in the uncoating of the viral genome and can therefore be divided into attachment-receptors and uncoating-receptors.

In their review Baggen and co-workers mention several prominent receptors that are involved in the different enterovirus entry processes. The abovementioned receptors CD155 and ICAM-1 are both falling into the category of uncoating receptors as do scavenger receptor class B member (SCARB2), ICAM-5 and coxsackievirus-adenovirus receptor (CAR). In some cases sialic acid, known to be an entry gate for influenza virus, was found to play a role as uncoating receptor, in others it was facilitating the attachment of the viral particles.

Attachment receptors are P-selectin glycoprotein ligand 1 (PSGL-1) (for several coxsackievirus strains), annexin II, dendritic cell-specific ICAM-grabbing non-integrin (DC-SIGN), nucleolin and vimentin (for EV-A71), heparan sulphate (for EV-A71 and echovirus 5). Rhinoviruses use LDL-receptor and VLDL-receptor and the lipoprotein receptor as attachment points. Many echoviruses and several coxsackieviruses use complement decay-accelerating factor for their first cell-contact. Some integrins (integrin $\alpha_v\beta_3$ and integrin $\alpha_2\beta_1$ also very late antigen 2 (VLA2)) were found to be suitable attachment points as well (Baggen et al. 2018).

Capsid binding compounds are already advanced in the development for anti-entero- and anti-rhinovirus application. The viral capsid is composed of 60 repeating units of the capsid proteins VP1-VP4. The myristoylated VP4 lies on the inside of the capsid so that only the other three proteins are exposed to the surface. One of the particularities of the enterovirus capsid is the presence of a cleft called the canyon encircling the five-fold axis where five VP1 components assemble with each other. This surface depression harbours the receptor binding part of the capsid and can be targeted by the so-called capsid binders, that bind to the canyon in a small

hydrophobic pocket usually occupied by a lipid referred to as pocket factor. Pleconaril is the oldest and most extensively studied one of them (Pevear et al. 1999). The canyon is present in many entero- and rhinoviruses, making some capsid binders applicable for broad-spectrum usage. Another capsid binder is the compound Pirodavir developed by Janssen Pharmaceuticals (Andries et al. 1992). It was optimised due to limited chemical stability and the final compound of the series is now developed as Vapendavir (Watson et al. 2003). A drug that gives hope as a potential anti-poliovirus compound is Pocapavir, which has completed Phase II clinical trials as effective (Collett et al. 2017). They all share their mode of action – rigidifying the capsid, ejecting the pocket factor, hindering capsid disassembly – and are trialled against different enterovirus indications. Approval of capsid inhibitors proves difficult because of insufficient efficacy and rapidly evolving resistance (Thibaut et al. 2012; Baggen et al. 2018).

3.1.7.2 Protease inhibitors

The proteases acting during the different cleavage stages of enterovirus polyprotein processing are 2A^{pro} and 3C^{pro} or its precursor 3CD^{pro}. 2A^{pro} catalyses a *cis* cleavage at its own N-terminus, which separates the capsid polyprotein P1 from the non-structural protein precursor P2 and P3. Furthermore it is also involved in the cleavage of components of the cellular cap-binding complex, which leads to a suppression in cellular protein expression (Yu & Lloyd 1991). No inhibitors for 2A^{pro} are in clinical development so far as the 3C^{pro} presents a more promising target due to its higher degree of conservation among the enteroviruses.

3C^{pro} and its precursor 3CD^{pro} are responsible for the remaining cleavages of the polyproteins into their functional proteins during enterovirus protein synthesis. The only cleavage that is not regulated by proteases is the one between VP4 and VP2, which occurs autocatalytically and marks the maturation of the provirion into the mature virus (Zell et al. 2017).

Apart from the processing of the viral polyprotein, the proteases are also involved in disabling the host cell's own protein synthesis. 2A cleaves eukaryotic initiation factor 4G (eIF4G), which is necessary for cap-dependent recognition of mRNAs and protein synthesis. Enteroviruses rely on IRES for recruitment of ribosomes for their protein synthesis and are therefore independent of eIF4G (Novoa & Carrasco 2015). 2A and/or 3C also disable polyadenosine binding protein and several other cellular functions and eventually host protein synthesis comes to halt (Bonderoff et al. 2008). They also abort shuttling between the nucleus and the cytoplasm (Walker et al. 2013) and disrupt functions of the cytoskeleton by cleaving dystrophin (Badorff et al. 1999). Last but not least 2A and 3C are combating mechanisms of the cellular immune response against viruses by cleaving mitochondrial antiviral signalling (MAVS), regulatory factor 7 (Lei et al. 2013), RNA helicase MDA-5 and an interferon beta inducing protein (Lei et al. 2016).

The most prominent protease inhibitor for 3C^{pro} is Rupintrivir, which has completed phase II clinical trials. After this the development was halted mainly because of the poor solubility of the peptidomimetic compound and due to insufficient reduction of disease severity in patients (Hayden et al. 2003). A rupintrivir analogue AG7404 (V-7404) completed phase I clinical trials but the development was halted as well.

DC07090 is a non-peptidomimetic inhibitor efficiently inhibiting EV-A. Non-peptidomimetic compounds might present a better oral bioavailability and could therefore be more promising for clinical development (Baggen et al. 2018).

3.1.7.3 3D^{pol} - RNA-dependent RNA polymerase inhibitors

Non-structural protein 3D^{pro} is the RNA-dependent RNA polymerase of picornaviruses. It is essential for the transcription and replication of the genome and is one of the proteins obligatorily encoded in the viral genome. A particularity of the enteroviruses is the small protein structure VPg at the 5' end of the genome with is the di-uridinylated non-structural protein 3B. The polymerase is also responsible to add the two molecules of UMP to a tyrosine residue on 3B. In this way VPg serves as a primer for the polymerase to initiate the RNA synthesis (Ferrer-Orta et al. 2015).

The enterovirus replication takes place in membranous structures termed replication organelles, which were believed to hide the viral RNA products from antiviral factors in the cytoplasm, but new investigations revealed to rather form the perfect environment for replication and production of new viruses (Charlotte et al. 2017).

Inhibitors of the polymerase target either the activity of the polymerase, essential interactions with the RNA template, interactions or contact between two or more RdRp molecules which undergo oligomerisation or they interact with regulatory proteins required for successful polymerase function (Ferrer-Orta et al. 2015).

Inhibitors are usually classified into two groups. The first are analogues of the nucleosides or nucleotides (NI) and the second are non-nucleoside/nucleotide inhibitors (NNI). NIs exhibit their antiviral activity by mimicking natural nucleosides or nucleotides. Usually they enter the cells as nucleosides and need to be phosphorylated into the active nucleotide by cellular enzymes. They are acting in the active site of the polymerase and are either terminating the elongation of the nascent RNA strand or incorporated into the RNA strand and lead to lethal mutations in the virus. NNIs typically bind in an allosteric way to the polymerase and inhibit conformational changes that take place between the initiation and the elongation phase or they stabilise the inactive conformation of the protein.

Examples for 3D^{pol} inhibitors are gemcitabine, which is a repurposed anti-cancer compound and showed a promising effect in animals, NITD008, which was active in vitro but in vivo toxicity was considered too high for the compound to proceed towards tests in humans.

Furthermore, the known antiviral ribavirin was found effective against enteroviruses and rhinoviruses (Bauer et al. 2017; Baggen et al. 2018).

An example for a non-nucleoside analogue of the 3D^{pol} is the diuretic amiloride, which makes the polymerase more error-prone. This compound can be directly evaluated in clinical studies to see if it is beneficiary in the context of enterovirus infections. Other non-nucleoside compounds were discovered and tested in vitro but none of them has proceeded towards (pre-)clinical development. Representatives are Aurintricarboxylic acid, BPR-3P0128, DTrip-22, Gliotoxin and GPC-N114. The last exhibits a thus far undescribed mechanism of action. It targets the RNA template-primer site (Bauer et al. 2017; van der Linden et al. 2015).

3.1.7.4 2B protein

2B is a small non-structural protein that belongs to the family of viroporins and interacts with lipid membranes, in particular the one of the ER and the Golgi apparatus. It is thought to assemble into oligomers and creates channels or pores within these membranes, which increases their permeability. 2B is not very conserved among the *Picornaviridae* family but enterovirus and rhinovirus 2B seem to be closely related. In its precursor form 2BC it is involved in the enrichment of vesicles and membranous structures in which the viral replication can take place. In the ER and the Golgi 2B is responsible for the reduction of Ca²⁺ levels via transmembrane pores and it inhibits protein trafficking through the Golgi apparatus (Saarnio 2017; Ao et al. 2014). So far no specific antivirals targeting this protein have been identified although there are already a few existing viroporin inhibitors against other viruses. Several calcium channel blockers and the sodium channel blocker amiloride have been tested against enteroviruses and exhibited antiviral activity but a direct connection to the 2B protein could not be established (Gazina et al. 2005; To et al. 2016).

3.1.7.5 2C protein and inhibitors

Information about the 2C protein and its inhibitors is discussed in section 3.2 in this chapter.

3.1.7.6 3A and 3AB proteins

3AB, 3A and 3B/VPg are produced from the common precursor P3 together with 3CD, 3C and 3D. Proteolytic cleavage first separates 3AB from 3CD and then 3CD^{pro} cleaves 3A and 3B. 3A and also 3AB is membrane bound. 3A is mainly responsible for the recruitment of cellular factors to the replication sites and to introduce morphological changes in the membranes of the ER and Golgi. During this process the trafficking of cellular proteins between the ER and the Golgi is halted, likewise are secretory cellular pathways.

Functions of 3AB might be the delivery of 3B (VPg) protein to sites of viral RNA synthesis, it is necessary for the correct cleavage between 3C and 3D and stimulates the activity of the viral polymerase and the 3CD-RNA binding. Furthermore it might be involved to link the replication

complex to suitable membrane structures (Tracy et al. 2008). An NMR structure of 3A found that it occurs in homodimeric units (Strauss et al. 2003). Furthermore, it was found for several enteroviruses that 3A can interact with ACBD3, which then recruits PI4KB (Lyoo et al. 2019). The importance of PI4KB is discussed in a specific section. It can be concluded that 3A serves as an important hub for protein-protein interactions and complex formation for concerted viral replication.

Some chemical compounds are reported to cause mutations in 3A but several of them were found to interact with cellular components and evoke only indirect compensatory mutations in 3A. Further studies are needed to elucidate 3A's modes of (inter)action during the viral life cycle and a way to use this non-structural protein as a target for drug development.

3.1.7.7 3B/VPg protein

VPg (viral protein genome-linked) can exist either as VPg or as VPg-pUpU. The latter is synthesised by the 3D^{pol} using Tyr3 as template for the attachment of two UMP units. As the complementary strand a stem loop structure in the coding region of the P3 used and a single adenylate residue is utilised twice to complement the uridine. Negative and positive strand possess the VPg at their 5' end and it serves as a primer for negative and positive strand RNA synthesis. Although it lies at the 5' end of the genomic RNA it is not used like the cap-structure of mRNA as an initiation for translation. This role is fulfilled by an IRES structure between the VPg and the coding region of the positive stranded genome. Structure and mechanism of VPg in complex with 3D^{pol} are partially elucidated by a model but these insights are so far only used to gain better insights into possible target sites on the 3D^{pol} (Wessels et al. 2006).

3.1.7.8 Assembly inhibitors

The exact steps by which the capsid proteins assembly and the role of proteins associated with this process are not yet fully understood. A cellular chaperone heat shock protein 90 (Hsp90) binds to myristoylated P1 protein, which is the polyprotein precursor part that contains the capsid proteins. It is assisting the cleavage into three capsid proteins VP0, VP1 and VP3. These self-assemble into protomers, which subsequently form pentamers. Twelve pentamers build the procapsid. VP1 and VP3 together with the non-structural protein 2C coordinate the insertion of the genomic RNA into the procapsid to generate the provirion. To form mature and infectious viral particles VP0 need to be processed into VP2 and VP4. This occurs when RNA is binding to the capsid proteins and introduces conformational changes (van der Linden et al. 2015).

Geldanamycin and its derivative 17-AAC are both inhibitors of Hsp90 that are used in the treatment for the Hsp90 mediated degradation of proteins that are overexpressed in cancers. Blocking Hsp90 in its function as chaperone protein for P1 has therefore an interesting antiviral

effect. So far the compounds have not been tested in clinical trials. Problems with repurposing of geldanamycin also arise from the toxicity that is tolerated in a cancer drug but not very likely to be accepted for a medicine against common cold (Tsou et al. 2013).

Furthermore, glutathione is involved in the capsid formation. Depleting the glutathione pools in the cell might be a strategy to target enterovirus capsid assembly. Two inhibitors should exemplarily be mentioned here: Buthionine sulfoximine and TP219. The former is an inhibitor of the synthesis of glutathione and the latter is a glutathione scavenger. Both have shown to interfere with the capsid formation and therefore inhibit the viral assembly. Glutathione targeting is the less promising strategy, as not all enteroviruses depend on glutathione (Bauer et al. 2017).

3.1.7.9 Host factors as antiviral targets

Picornavirus infection relies on a series of host proteins or host factors that are crucial for the viral replication. Host factors are an interesting antiviral strategy because several enteroviruses often rely on the same host factors regardless of the species and resistance against the treatment might take a long time to develop. Targeting a host factor that carries out essential functions in the human cell in physiological conditions might lead to adverse effects on the human body. The development of compounds against host factors, although desirable remains challenging.

Two well-known and commonly targeted host factors are PI4KB, OSBP. Phosphatidylinositol 4-kinase type III β (PI4KB) is a cellular protein that catalyses the phosphorylation of phosphatidylinositol at the position D4 to yield phosphatidylinositol-4-phosphate (PI4P). PI4P is a molecule involved in signalling and regulation of trafficking through the Golgi apparatus and the trans Golgi network (Reid et al. 2015). PI4KB interacts with a Golgi adaptor protein named ACBD3 and helps to create PI4P enriched membranes. Enterovirus 3A protein recruits PI4KB to the membranes, which enriches them in PI4P lipids to create the necessary environment for enterovirus replication (Lyou et al. 2019). Therefore enteroviruses utilise PI4KB for membrane remodelling and the generation of the replication organelle. All enteroviruses use PI4KB as a host factor but inhibitors of the protein are lethal in mice, therefore a development for clinical use of PI4KB inhibitors does not seem likely. In vitro it has been shown that even if the inhibitors are targeting a host factor, compensative mutations in protein 3A might be able to overcome the inhibitory effect (Bauer et al. 2017).

Another promising host target is oxysterol-binding protein (OSBP). In a drug-repurposing screen Itraconazole, an anti-fungal compound, was identified as a broad-spectrum enterovirus inhibitor. It exerts its inhibitory effect by blocking the OSBP mediated exchange of PI4P lipids with cholesterol at ER-Golgi contact membranes. OSBP binds to PI4P lipids (enriched in the replication organelle membranes through PI4KB) and the exchange of PI4P and cholesterol is

essential during viral replication (Strating et al. 2015). Inhibitors of OSBP are the cholesterol analogue 25-hydroxycholesterol, AN-12-H5, T-00127-HEV2 (Arita et al. 2013), TTP-8307 (Albulescu et al. 2017) and the natural compound OSW-1 (Albulescu et al. 2015). Itraconazole might be a promising candidate for the treatment of rhinovirus infections. Mouse models have found a prophylactic effect of the compound (Shim et al. 2016).

3.2 2C protein

3.2.1 Introduction to the non-structural protein 2C

The 2C protein is undoubtedly an interesting protein target for the development of broad-spectrum antiviral agents against picornaviruses, especially due to its high sequence conservation within the family (Xia et al. 2015). Until 2017 no crystal structure was available for 2C but sequence analysis, biochemical and virological characterisation revealed several important features of the protein.

First, it is an NTPase, more specifically it possesses ATPase and GTPase activity. ATP binding and cleavage is carried out by the residues of the walker motifs; Walker A aa 129-136 is binding the nucleotide, Walker B motif spanning the residues 172-177 is binding Mg^{2+} , motif C is located at residues 217-223, and the residues R240 and/or R241 as R-finger(s) from a second monomer (Guan et al. 2017). Helicase activity was long predicted (Gorbalenya et al. 1990) and eventually assessed by Xia et al. but the assay could not be repeated, thus far (Xia et al. 2015). Furthermore, 2C contains an N-terminal membrane-binding motif that was predicted as an amphipathic helix and was mapped to the amino acids 21-54 (Teterina et al. 1997). Direct interaction with RNA was demonstrated. The residues involved are 21-45 and 312-319 (Rodriguez & Carrasco 1995). Sequence analysis revealed a cysteine-rich region from amino acid 269 to 286 that the crystal structures revealed as zinc binding motif (Guan et al. 2017). Xia and co-workers experimentally demonstrated that 2C acts as an ATP-dependent RNA helicase that can be classified into the SF3 helicase superfamily (Xia et al. 2015).

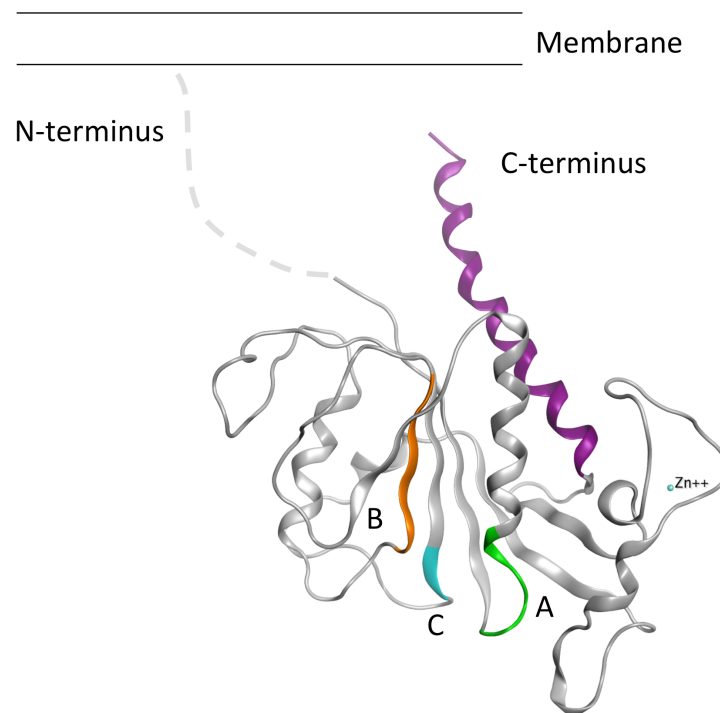
Several functions within the viral lifecycle have been identified. 2C is supposedly involved in uncoating where residues K259, M293, K295, C272 and H273 of Poliovirus 1 proved to be important (Li & Baltimore 1990; Wang et al. 2012; Wang et al. 2014; Asare et al. 2016). It participates in cellular membrane rearrangement (Cho et al. 1994). It is associated with RNA replication and immune evasion (Lei et al. 2016). Furthermore, it plays a role in encapsidation of RNA into new viral particles for which the residues K279 or R280, C272, H273, N252, K259, Q65, L125 and V218 seem to be important, as viruses with mutations in these residues fail to properly encapsidate the RNA into the nascent virion (Wang et al. 2012; Liu et al. 2010; Asare et al. 2016). It is also hypothesised to function as a helicase and as a chaperone molecule although experimental data is inconclusive (Papageorgiou et al. 2010; Xia et al. 2015). Despite the manifold processes in which 2C participates, the details at molecular level remained elusive.

In 2017, Guan et al. deposited the first crystal structure of the truncated 2C protein of EV-A71 in high resolution (PDB: 5GRB, 5GQ1). The structure confirmed most of the virological data and

finally opens the door for targeted rational drug design against this protein. In late 2018, the same group published a truncated structure of 2C of poliovirus. The second structure confirmed the conservation of the fold and is very similar to the EV-A71 one (PDB: 5Z3Q) (Guan et al. 2017; Guan et al. 2018).

3.2.2 Characterisation of the 2C structure

The crystal structure by Guan et al resolved the EV-A71 2C protein from aa 115-329. The N-terminal residues that are involved in the interaction with the cell membrane and responsible for oligomerisation were not resolved in this structure. 2C is also thought to act in the early stage replication complexes in its precursor form 2BC, before the cleavage between the two individual proteins occurs. A structural resolution of the precursor would also provide valuable information about the two non-structural proteins.



**Figure 54: Semi-schematic depiction of 2C protein
Walker A (green) and B (orange) motifs, motif C (turquoise) and the C-terminal helix (purple)**

In the following section a detailed description of the crystal structure is given. This description is a summary of the findings by Guan et al. (2017). The crystallised EV-A71 2C protein was resolved from residue 116 on to the C-terminal residue 329. It reveals the predicted ATPase domain, the structure of a zinc finger or zinc binding domain and a long C-terminal α -helix. The classical α/β canonical Rossmann fold is presented by five parallel β -sheets and flanked on one side by one α -helix α_1 and on the other side by two α -helices α_2 and α_3 . The Walker A motif

spans the residues 131-136, Walker B comprises 172-177 and the small motif C is composed by the residues 222 and 223. The two arginines R240 and R241 reside on the opposite side of the protein and are forming the arginine finger part of the bipartite ATP binding site when the protein oligomerises. The cysteine-rich zinc-binding domain lies between the residues 270-286. The zinc finger does not fall into a known class of zinc finger folds, as the third cysteine residue is not present as it would be for the CCCC-type zinc finger (Krishna et al. 2003). Instead the zinc is complexed only by three cysteines: 270, 281 and 286 via their S_γ atoms. The three cysteines form a triangular plane with the zinc atom. The backbone carbonyl oxygen of S282 builds one of the pyramidal tips of the bipyramidal structure and a structural water molecule is positioned on the other one. E272 and K288 form a salt bridge to stabilise the zinc-binding site even further.

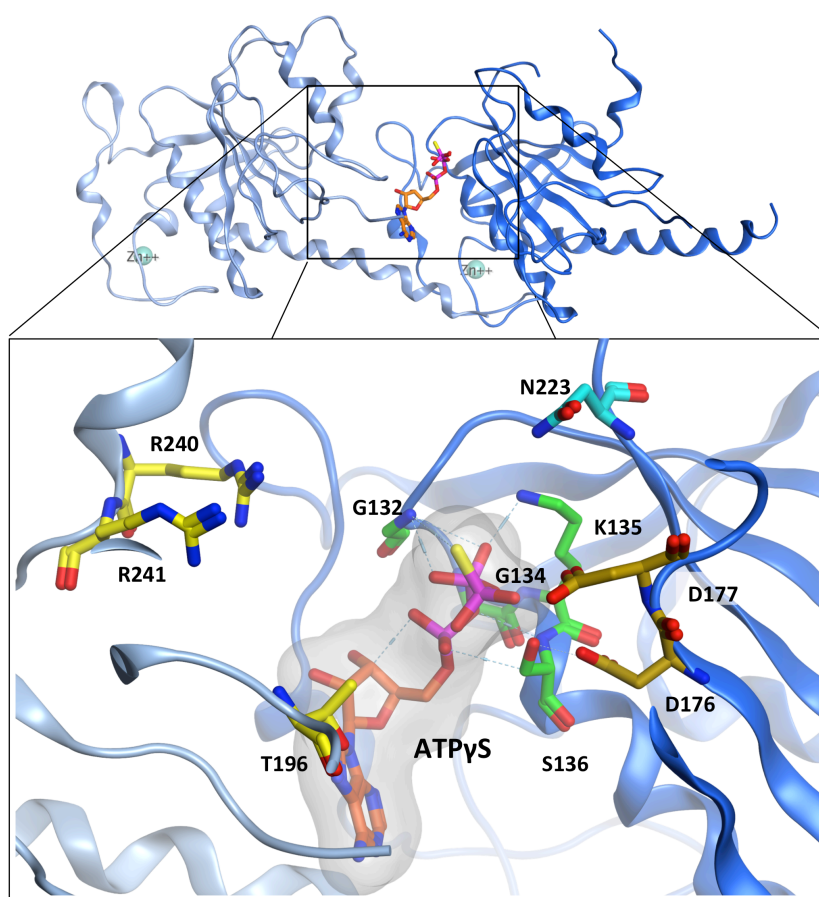


Figure 55: Dimer interface and close-up of the ATP binding pocket

Not only the N-terminal part is necessary to form oligomers but also the C-terminal helix is a crucial interaction point. The residues 320-329 are interacting with a shallow pocket on a different monomer especially with R144, which was found to abolish oligomerisation if mutated. Small angle X-ray scattering (SAXS) curves with a constant concentration performed by Guan et al. revealed that 2C protein was shown to elute in complexes corresponding to the 4-fold mass of the monomers. They showed a rod-like arrangement. In previous studies TEM-

196

analysis revealed a hexameric ring structure for 2C of echovirus 30 and FMDV (Papageorgiou et al. 2010; Sweeney et al. 2010). These 2C proteins adopt a ring shaped structure and their ATP-binding sites lie on the interfaces of two monomers. Crucial for the interaction between the monomers is also the C-terminal helix, more specifically the side chains of residues T323, I324, L327 and F328. In the different configurations revealed by the crystal structure, the terminal part of the C-terminal helix has a hinge region (aa 318-319) that makes the most C-terminal part flexible to adapt to a pocket between the ATP binding site and the zinc finger allowing the resulting complex to adopt a hexameric shape. Guan and co-workers superimposed the structure of EV-A71 2C onto the D2 domain of the cryo-EM structure of human p97 (PDB: 5FTK) to model the hexameric structure and it fitted the D2 domain ring well. Therefore they hypothesise that also EV-A71 2C might act as a hexameric ring-shaped helicase of the SF3 type. Contrasting to the DNA helicases that are crystallised the inner core of the ring is negatively charged; therefore, it is not thought to bind RNA within this channel. Additionally, RNA binding sites have been identified on the rim of the hexamer in the middle of the C-terminal helix. One RNA molecule might be bound to the hexameric 2C protein on more than one RNA-binding site, thus further stabilizing the ring-shaped structure.

The functional implications of oligomer formation of 2C protein were explored by mutations of the crucial interaction sites. First, if regions involved in oligomer formation in the C-terminus (320-329) are deleted, the ATPase activity is partially or fully abolished (Guan et al. 2017), the same holds true for mutations in the pocket binding domain and the mutation of the arginines in the supposed R-finger (Guan et al. 2017). Mutation E325A impairs the oligomer formation resulting in dimeric structures, which retain some ATPase activity in the biochemical assay but no viral production can be observed for viruses carrying this mutation. E325 is involved in forming a salt bridge with R144 on a second 2C molecule stabilising the oligomer structure (Guan et al. 2017).

Taken together the revelations that were provided by the crystal structure served as starting point to investigate the 2C structure of EV-A71 and several other 2C proteins like the one of CVB3, EV-D68 and, polio in greater details, especially with the focus on drug discovery and targetability.

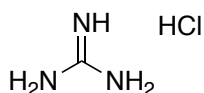
3.2.3 Inhibitors of 2C protein

A second approach to target the 2C protein of human enteroviruses would be a ligand based one. In the past several enterovirus inhibitors were discovered and the following ones were found to target the 2C protein according to resistance mutations arising in the protein. In the light of the resolution of the crystal structure the mutations caused by the different enterovirus inhibitors that can be mapped to the 2C sequence, can now be investigated for

their location on the protein. Therefore, it might be achievable to uncover the mode of action of some of the known inhibitors, by identifying suitable pockets on the crystal structure.

Most of the 2C inhibitors reportedly exhibit activity against CVB3 and several resistance mutations were reported for this virus. Therefore, as described later, homology models of CVB3 and other human enteroviruses were created to have a better overview of the pockets and residues in the potential regions of activity.

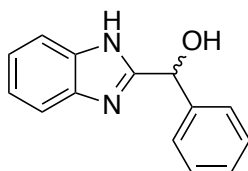
3.2.3.1 Guanidine hydrochloride



Guanidine hydrochloride

Despite the lack of a crystal structure in the past, several 2C inhibitors were discovered and 2C as their target was confirmed by mutation studies. Guanidine hydrochloride (GuaHCl), which is a small salt molecule, inhibits 2C protein of several virus strains (poliovirus, several coxsackieviruses, some echoviruses and FMDV)(Thibaut et al. 2012). It was the first 2C inhibitor discovered and it is still in use as standard for activity measures of other compounds (Rightsel et al. 1961). It inhibits the negative strand RNA synthesis, the binding of 2C protein to host membranes as well as the ATPase activity of the protein (Barton & Flanagan 1997; Pfister & Wimmer 1999). The dose to see the inhibitory effect can be as high as within the millimolar range but effects are consistent and GuaHCl has even been tested in vivo (Eggers 2004). It is not suitable for the use in human subjects because of toxicity concerns and the high dose that would be required. Guanidine is still used as a control for the evaluation of new enterovirus inhibitors and as a tool compound to differentially investigate the function of picornavirus 2C proteins.

3.2.3.2 2-(α -Hydroxybenzyl)-benzimidazole (HBB)

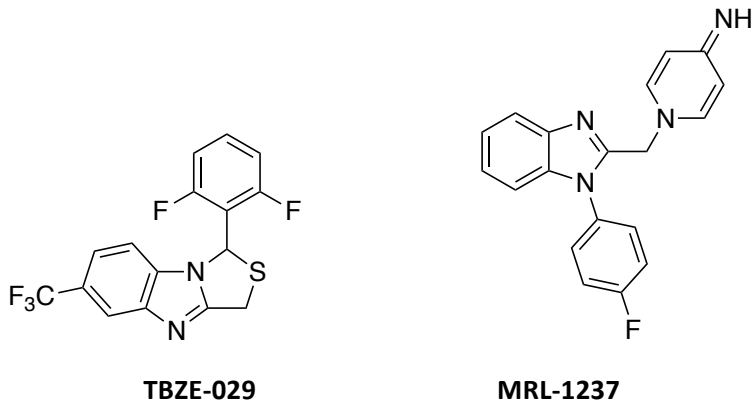


HBB

2-(α -Hydroxybenzyl)-benzimidazole (HBB) was found to inhibit several picornaviruses by blocking its RNA replication (Eggers & Tamm 1962). It was discovered in the 1960ies and is like GuaHCl very well studied. It was the first of a series of benzimidazole derivatives that was then

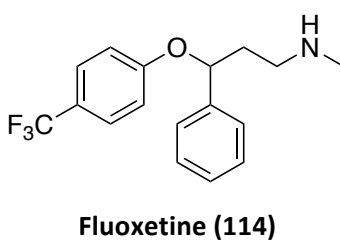
developed into enterovirus inhibitors. HBB shows activity against poliovirus a number of coxsackieviruses, echovirus 9 (Eggers & Tamm 1961).

3.2.3.3 TBZE-029 and MRL-1237



In following studies the benzimidazole backbone proved successful in several other enterovirus inhibitors. TBZE-029 and MRL-1247 are the two compounds depicted above. TBZE-029 was described with a series of analogues and proved active against CVA9 several coxsackie B viruses, echoviruses 9 and 11 but neither against rhinoviruses or poliovirus (De Palma et al. 2007). MRL-1247 is another benzimidazole derivative that was tested by Shimizu and shows activity against poliovirus coxsackie B viruses (Shimizu et al. 2000). Resistance mutations and cross-resistance of viruses with mutations against guanidine or HBB demonstrated that the target for these compounds was as well the 2C protein (De Palma et al. 2008).

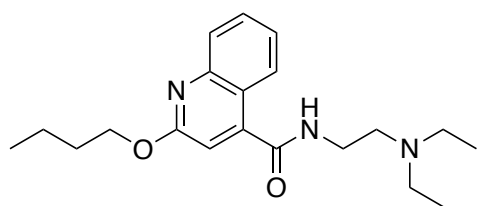
3.2.3.4 Compounds from repurposing screenings



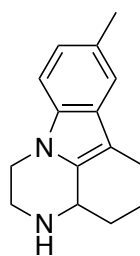
Fluoxetine was identified as an inhibitor for enterovirus B and D species during drug repurposing screens (Zuo et al. 2012; Ulferts et al. 2013). The compound marketed as Prozac[®] is a selective serotonin reuptake inhibitor and is used in the treatment of anxiety disorders and depression (Wong et al. 2005). Since the discovery as anti-enterovirus compound case reports and a small clinical trial have been published. In the case report a child with chronic enterovirus infection was saved by the use of fluoxetine (Gofshteyn et al. 2016). In the second study patients with acute flaccid myelitis were given fluoxetine in different doses. This study did not find a benefit in the use of fluoxetine for this particular case (Messacar et al. 2019). The

study has several limitations and it should not be excluded that fluoxetine might have beneficiary effects when given early enough and only when the symptoms are evoked by a susceptible enterovirus strain.

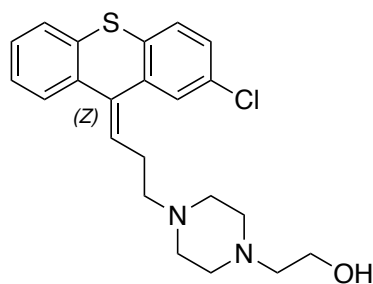
Zuo and co-workers identified also other compounds that inhibit CVB3 replication during their library screening including Mefloquine, Lycorine, Dibucaine, Cycloheximide, Zuclopenthixol and Flupentixol (Zuo et al. 2012). In parallel, Ulferts and co-workers analysed in depth the activity of fluoxetine on different enterovirus strains and identified the ²²⁴AGSINA²²⁹ loop as a hotspot for resistance mutations against fluoxetine (Ulferts et al. 2013). In 2016, Ulferts et al. performed a screening similar to the one of Zuo in 2012, as well on the Prestwick Chemical Library. They identified several active compounds against CVB3, and thereafter set out to characterise the activities of pirlindole, dibucaine, zuclopenthixol and fluoxetine in greater details. These compounds all evoked mutations in the 2C protein (Ulferts et al. 2016). In section 3.7 of this chapter these compounds were used for docking studies in 2C.



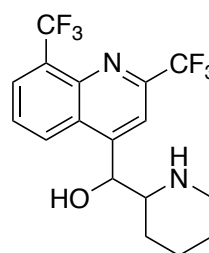
Dibucaine



Pirlindole

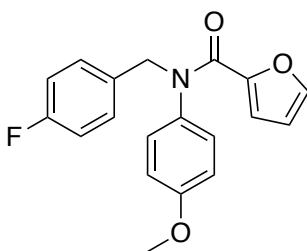


Zuclopenthixol



Mefloquine

3.2.3.5 Novel 2C targeting compounds



N-(4-Fluorobenzyl)-N-(4-methoxyphenyl)furan-2-carboxamide (115)

In a second screening the Zuo et al. identified a series of 2C inhibitors with a broader activity on different enterovirus strains in a second screening approach. Of particular interest to this thesis is compound 2 (**115**), which will be further discussed in section 3.8.3 of this chapter (Zuo et al. 2016). It is depicted above.

Table 72: Mutations of 2C inhibitors
(Table adapted from supplementary material of Guan et al. 2018)

Name	Virus	Mutations	References
Guanidine HCl	PV-1	I142V, A143G; N179G, M187L, S225T, I227M, A233T/S	(Pincus et al. 1986; Tolskaya et al. 1994; Shimizu et al. 2000)
	E-9	E64G, A133T	(Klein et al. 2000)
	CVB3	A224V	(De Palma et al. 2008)
HBB	PV-1, E-9, CVA9, CVB3	I227L, A229V	(Klein et al. 2000)
MRL-1237	PV-1	I120V, F164Y	(Shimizu et al. 2000;
	CVB3	N179A/G, I227V	De Palma et al. 2008)
TBZE-029	CVB3	A224V, I227V, A229V	(De Palma et al. 2008)
Fluoxetine (114)	CVB3	A224V, I227V, A229V	(Ulferts et al. 2013)
Pirlindole	CVB3	A224V, I227V, A229V	(Ulferts et al. 2016)
Zuclopenthixol	CVB3	A224V, I227V, A229V	(Ulferts et al. 2016)
Dibucaïne	CVB3	A224V, I227V, A229V	(Ulferts et al. 2016)
115	CVB3	S58N, C179F, I227V, N257D	(Zuo et al. 2016)

3.3 Aims

After the release of the first crystal structure of EV-A71 2C by Guan et al. in 2017 Lisa Bauer (LB), Roberto Manganaro (RM) and myself (BZ) initiated a new research collaboration among the ANTIVIRALS ETN to investigate 2C as a target for rational drug discovery.

The main aims of this collaboration were:

- to analyse the crystal structure of EV-A71 2C and use it as a template for homology modelling to expand the structural information towards other enteroviruses

- to model the ring-shaped hexameric helicase structure that was suggested in the publication of the crystal structure and potentially draw conclusions on the mode of action of that protein
- to find potential druggable pockets on the protein and use them as sites for a high throughput virtual screening approach to discover new molecules that could inhibit the replication of the virus
- to analyse the data of known 2C inhibitors and to deduct a binding mode and to chemically optimise the compounds with structure-based and medicinal chemistry-based methods

The project was organised between Cardiff (BZ and RM) and Utrecht (LB), with Cardiff handling the computational and chemistry-related parts of the project and Utrecht playing the virological counterpart.

We planned to use molecular modelling tools in order to build the hexameric model of the 2C protein and homology models for the different viruses and strains, to find target sites on the protein and to perform the HTVS on different sites. We used docking tools to investigate the fit of known inhibitors in the hypothesised pockets and molecular dynamic simulations to analyse the movements of the protein over time under the different conditions and with different inhibitors.

RM and BZ both worked on the modelling equally and RM together with Moira Lorenzo Lopez (ML), an Erasmus student, synthesised the molecules discussed during this thesis. Some novel compounds were also synthesised by Thamir Rahman and Marcella Bassetto later in the project.

Biological tests of the compounds and mutation studies were performed by LB. Later, biochemical assays on the purified 2C protein were performed in the laboratory group of Bruno Coutard (BC) in Marseille, by himself or his student Pricilla Kazzi (PK).

Four different parts of the project will be discussed in this thesis:

- High-throughput virtual screening on the C-terminal domain binding pocket
- Homology models of different enteroviruses and higher order complexes of 2C
- Investigation of the mode of action of known 2C inhibitors
- Design and Synthesis of new promising (pan-)enterovirus inhibitors

With the availability of the crystal structure of enterovirus 2C protein, the drugability of this target has increased. The fact that 2C is the most conserved protein within the family of picornaviridae gives hope to find new inhibitors that might act as pan-inhibitors and investigations on this protein might be promising to find a broad spectrum treatment for enteroviruses.

The first goal was to investigate the structure and function of 2C protein in detail. Homology models for the different picornaviruses were generated to investigate the common features and the differences of the protein, specifically towards drugable pockets. Sequence alignment and superimpositions of the models should lead these investigations.

2C protein acts as a homooligomer and is believed to form a hexameric ring shaped structure. This has been investigated in several structural biology studies but a three-dimensional structure of the hexameric arrangement on molecular level is still lacking. In order to investigate this the aim was to create a hexamer structure based on the symmetry and subject the result to molecular dynamics simulations. For the sake of time and to limit the amount of data generated, molecular dynamic studies serving drug design purposes were carried out on monomers or dimers.

The third pillar in this investigation relies on the activity data of different compounds that are known to be 2C inhibitors. Several of them were discovered in the process of (high throughput) screening of compound libraries. Some of the known 2C inhibitors are approved drugs, which makes repurposing a possible option as already demonstrated in the case of Fluoxetine. These compounds are profiled and tested to cause resistance mutations in 2C protein, which likely makes 2C the target protein. The known mutations and investigations in close collaboration with our biological partners drove the aim to elucidate the mode of action of known and newly designed 2C inhibitors.

Eventually, based on the information acquired for the binding mode of fluoxetine, other compounds were designed, synthesised and tested in several rounds, in order to obtain new inhibitors for enteroviruses. With the hypothesis of a binding pocket structure-based design and classical medicinal chemistry approaches were combined to result in compounds with improved activity on a broader spectrum of viruses.

3.4 Investigations on the EV71 2C crystal structure

Two different crystal structures were published in by Guan and co-workers in 2017. The access codes of the PDB are 5GQ1 and 5GRB. In this work the 5GRB structure became the most important starting point and is extensively used throughout this thesis, mostly as a template for homology models or as reference structure for molecular dynamics simulations. It contains ATPγS and gives therefore a more complete picture of the 2C protein.

Several pathways were chosen to target the picornavirus 2C protein. The first steps were taken towards testing the drugability of the protein. Some points were already addressed in the publication. One of which is the “pocket binding domain”, the last 6 residues of the C-terminal helix that is thought to insert into a lipophilic surface pocket of another monomer of 2C. In this way the 2C protein is able to assemble to differently structured oligomers (Guan et al. 2017).

The most relevant form of oligomerisation is thought to be a hexameric ring structure. Sequence analysis predicted 2C protein to be a member of the SF3 superfamily of ring-shaped hexameric helicases. This hypothesis is also explored by the authors of the study, where they were able to superimpose the 2C monomers to domain D2 of p97 (5FTK) a cellular hexameric ring-shaped hexameric ATPase of the AAA (ATPases Associated with different Activities) family (Guan et al. 2017; Guan et al. 2018).

3.5 Pocket for the C-terminal domain

In the publication by Guan et al. 2017 the authors explored quite in depth the oligomerisation motifs of the C-terminal helix and the corresponding pocket on a second monomer. This interaction is also reflected in the orientations of the monomers in the crystal structure (Guan et al. 2017). It seemed therefore to be a promising target site for the development of oligomerisation inhibitors. It was reported that oligomerisation is crucial for several functions of the 2C protein during replication, especially if it acted as a helicase (Tolskaya et al. 1994).

For this purpose the crystal structure dimer between chain A and chain C was investigated where chain C fits into a pocket-like surface region on chain A. The target site is shown in Figure 56. It is in close proximity to both the zinc-binding domain and the ATPase pocket. The last six residues of the C-terminal helix are stretching roughly the core of the pocket. Residues important for the interaction between the two monomers in this region are depicted in Figure 56.

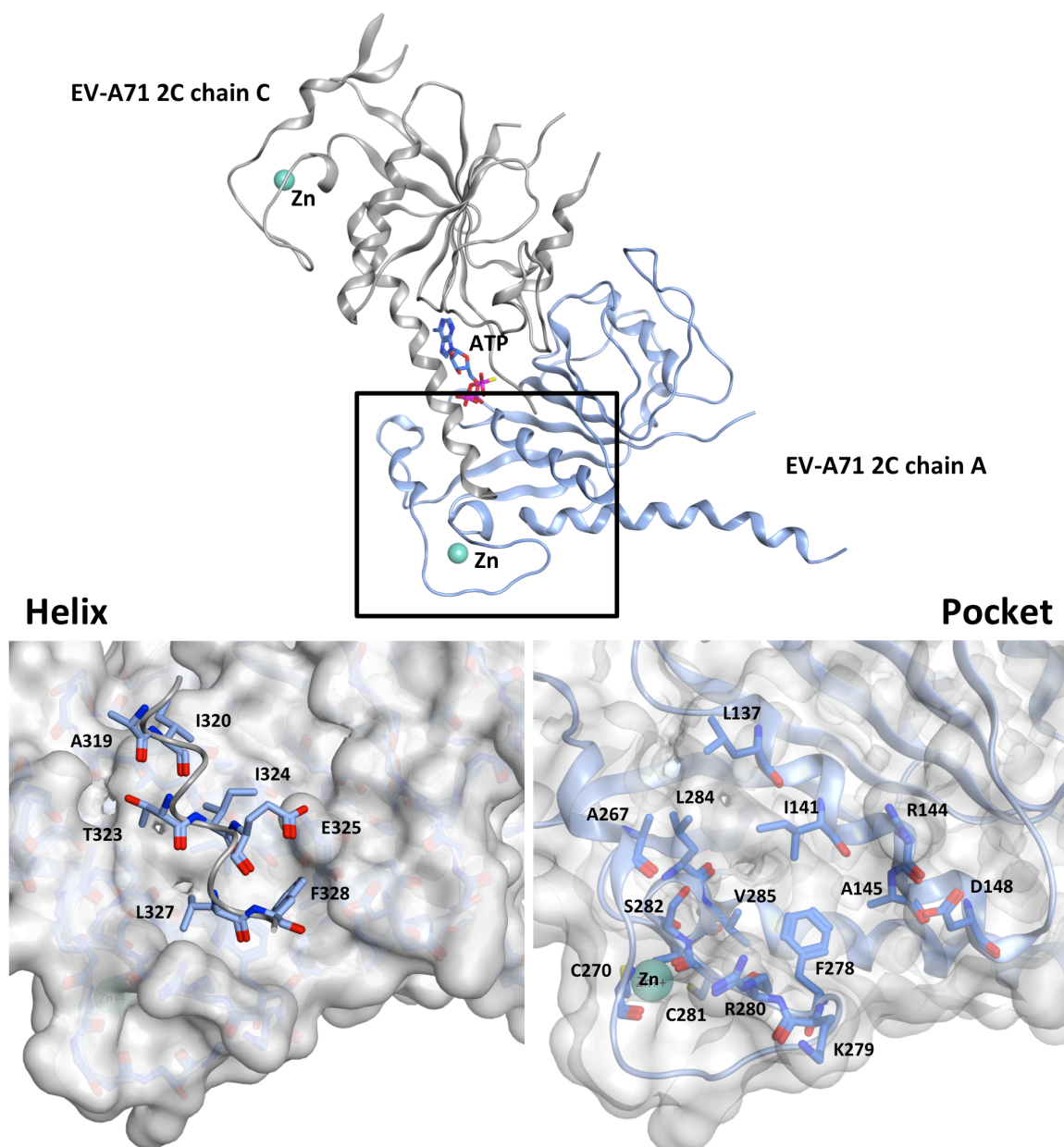


Figure 56: 2C Dimer oligomerising with C-terminal helix (top); C-terminal helix (bottom-left) and the corresponding pocket (bottom-right)

3.5.1 HTVS

The protein pocket was targeted using a high-throughput virtual screening (HTVS) approach. A schematic view of the HTVS workflow that was used to target the pocket of the C-terminal helix is depicted below.

HTVS workflow

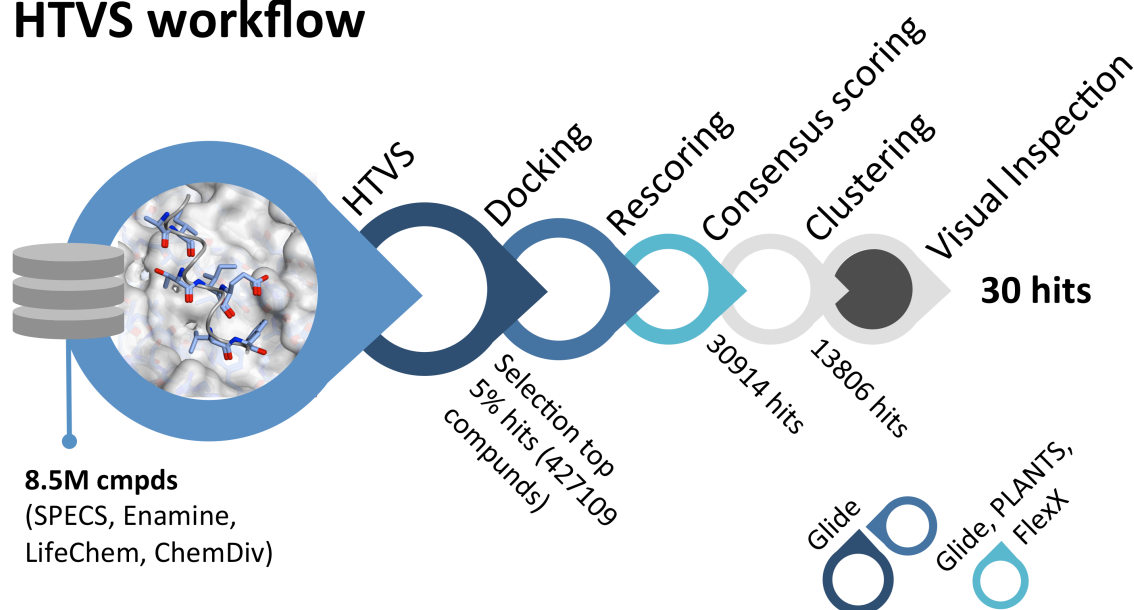


Figure 57: High-throughput virtual screening workflow (Design by Roberto Manganaro)

The described pocket was used as target site for a high-throughput virtual screening. First, a centre point for the position of the compounds was chosen ($x = -26.86$; $y = 6.13$; $z = -40.34$). Then a grid box of 10 Å with that centre coordinates was created for the HTVS procedure, which is similar to the docking described in previous chapters.

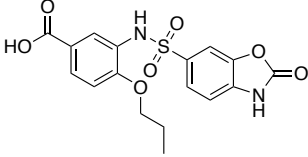
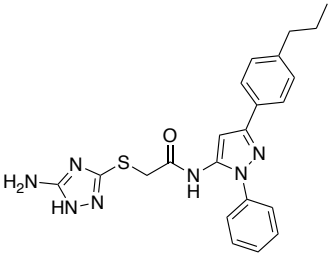
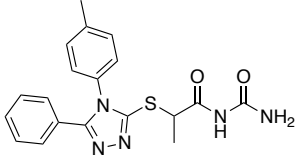
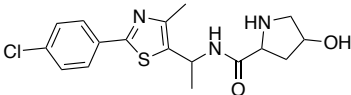
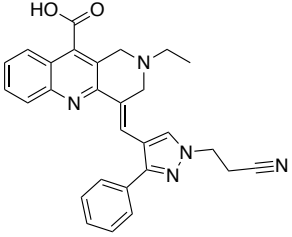
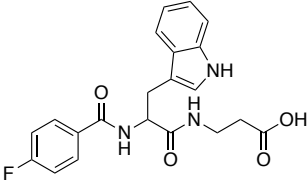
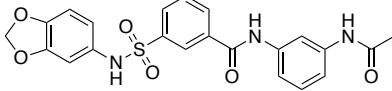
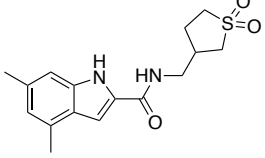
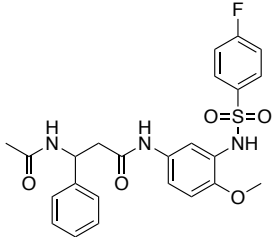
Compounds from four different libraries (SPECS (www.specs.net), Enamine (<https://enamine.net>), Chemdiv (<http://www.chemdiv.com>) and Lifechemicals (<https://lifechemicals.com>)) were screened first for a quick assessment of their fit in the pocket with the HTVS setting in the docking tool Glide. In total 8,542,178 compounds were docked. Then, 5% of the compounds (427,109) were selected for a standard precision docking with Glide. The ligands were treated fully flexible and three different poses per compounds were created and kept after post-docking minimisation. The receptor was kept rigid as per default for Glide SP docking. The resulting 1,233,636 docking poses were to be rescored with Glide XP, FlexX and PLANTS to reach a consensus score. For the handling of the output the data needed to be incorporated into an Excel file, which has a maximum of 2^{1024} rows. For this reason only 1,048,576 structures were further assessed with the three scoring functions. 30,914 compounds that scored within the first quartile of the compounds in each of the three scoring functions were selected. A fingerprint-clustering step was carried out and several descriptors were calculated that would facilitate a decision during the visual inspection of the 13,806 compounds. Eventually, both my colleague Roberto Manganaro, and myself, selected 30 final compounds, each, in independent visual inspections. Five of the final compounds were chosen by both of us, highlighting a certain consensus among the important criteria in the

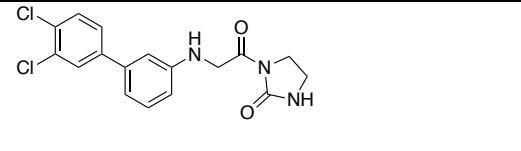
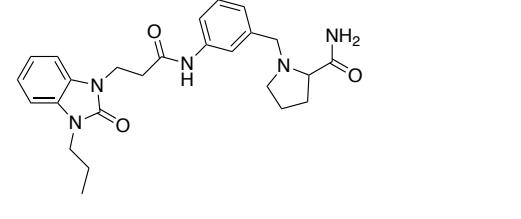
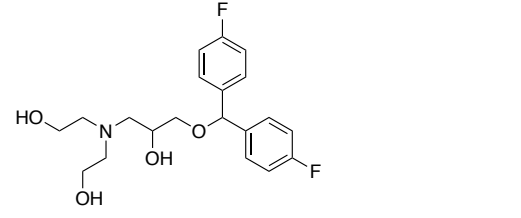
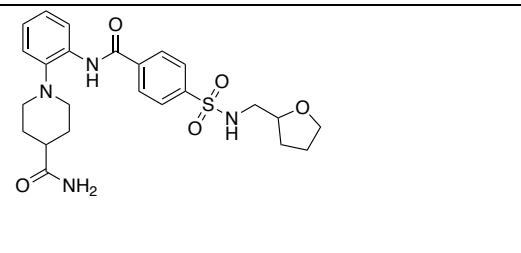
visual inspection. Of the 30 compounds we then agreed upon, 28 compounds were bought and tested (Table 73).

Table 73: Compound selection after HTVS

Nr	Structure	Name	Vendor	Activity
116		AM-807/13614233	SPECS	Not tested
117		D058-0287	ChemDiv	NA
118		D-426-0932	ChemDiv	NA
119		E518-0596	ChemDiv	NA
120		F2652-0108	LifeChemicals	Not tested
121		F290-0231	ChemDiv	NA
122		F292-0192	ChemDiv	NA
123		F447-0028	ChemDiv	NA

124		G629-0288	ChemDiv	NA
125		L268-0831	ChemDiv	NA
126		P072-0797	ChemDiv	NA
127		V030-1740	ChemDiv	NA
128		Y040-2631	ChemDiv	NA
129		Y041-5172	ChemDiv	NA
130		Z1102033654	Enamine	NA
131		Z1238812198	Enamine	NA
132		Z1328600108	Enamine	NA

133		Z1381333285	Enamine	NA
134		Z18356009	Enamine	NA
135		Z19860729	Enamine	NA
136		Z2242893803	Enamine	NA
137		Z234753724	Enamine	NA
138		Z235335563	Enamine	NA
139		Z31734866	Enamine	NA
140		Z319058038	Enamine	NA
141		Z33333716	Enamine	NA

142		Z739822092	Enamine	NA
143		Z786010070	Enamine	NA
144		Z94601670	Enamine	NA
145		Z972378562	Enamine	NA

3.5.2 Biological evaluation

30 compounds were finally selected, but the compounds from SPECS and LifeChemicals were not purchased. 28 compounds were then sent for testing to Lisa Bauer, PhD student in the laboratory of Frank van Kuppeveld at the university of Utrecht (NL). The compounds were evaluated in a multicycle CPE reduction assay. None of the compounds was able to inhibit the virus induced cytopathic effects at any concentration between 0.01-100 μ M.

3.5.2.1 Protocol for the multicycle CPE-Reduction Assay

The protocol was provided by Lisa Bauer and reported here as received: Subconfluent monolayer of Hela R19 cells were seeded in a 96-well and treated with serial dilutions of the respective compounds and infected with EV-A71 (strain BrCr) at the lowest multiplicity of infection (MOI) that resulted in full CPE within 2-3 days. The medium contained 10% fetal bovine serum. The cells were incubated for 2-3 days until full CPE in the infected and untreated cells was observed. Cell viability was determined in parallel using the AQueous One Solution Cell Proliferation Assay (Promega) according to the manufacturer's protocol. The optical density at 490 nm was determined using a microplate reader. Raw optical values were converted to percentage of untreated and uninfected cell controls after the subtraction of the background. Cytotoxicity was assessed in a similar set-up.

3.5.3 Discussion and Conclusions

Although the pocket accommodating the C-terminal helix was validated by mutation studies to be important for the oligomerisation and function of the protein (Guan et al. 2017), the compounds that were selected in the HTVS workflow, and finally by visual inspection, failed to inhibit EV-A71. Reasons for this can be manifold. For once, the pocket accommodating the ultimate residues of the C-terminal helix is relatively shallow. Furthermore, it does not provide any crucial enzymatic function for the protein and the residues involved in the interaction between the two 2C molecules were not systematically mutated, thus the pocket was not validated as a druggable target. Of highest importance seems to be the salt bridge between the residues E325 and R144, but this was the only indicative and validated interaction that could potentially be targeted. In a personal communication Bruno Coutard presented some results from the PISA server (Protein interfaces, surfaces and assemblies' service PISA at the European Bioinformatics Institute. (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html))(Krissinel & Henrick 2007)) that predicts interaction interfaces on proteins. The interface scored 0.000 in complex formation score, which ranges from 0-1, with 1 being highly likely to form complexes and 0 not likely. Although the evaluation of this interface in the publication of Guan et al. seems reasonable, the score by the PISA platform might give some indication that this interaction and pocket are not of relevance.

On the other hand, given the vast amount of compounds screened in this HTVS, the selection of only 30 compounds might be very restrictive and it would have proven useful to test more compounds. As a second approach, a shape-based screening based on the last residues of the helix and their pharmacophore features was planned but not conducted due to limited time. It was planned to choose compounds that would both be among the best compounds after the consensus scoring and among the best ranked compounds of the shape-based screening. These compounds would then be subjected to visual inspection, reducing the human bias and perhaps yielding results that would be even better founded, computationally. In the end this would still not guarantee active compounds with certainty.

In conclusion, the seemingly low-hanging fruit of simply targeting a pocket that is crucial for the oligomerisation might raise more questions that it was able to answer. Maybe a pharmacophore screening instead of an HTVS might reduce the compounds in a more meaningful way and could potentially provide active oligomerisation inhibitors in the future. This is of course only after a deeper evaluation of the pocket/interface from the biological point of view.

3.6 Investigations on 2C models for different enteroviruses

2C protein is among the most conserved proteins within the enterovirus family. The spatial arrangement of the features revealed by structural biology studies is therefore believed to be similar for all human enteroviruses. The resolution of the crystal structure of the EV-A71 2C protein thus presented a steppingstone to structural insights into all human enterovirus 2C proteins. Retrospectively, this logic proved reasonable, as the poliovirus 2C crystal structure released in 2018 is essentially very similar to the one of EV-A71 (Guan et al. 2017; Guan et al. 2018).

In order to compare the resistance mutations reported for known enterovirus inhibitors of different viruses, homology models were generated. These homology models were also used for docking studies for the inhibitors in order to discover their potential binding pockets. Furthermore, structures of higher order were modelled because the similarity to SF3 helicases would indicate a hexameric ring-shaped structure of 2C protein. Also the rod-like structures reported by Guan and co-workers for the EV-A71 2C protein required at least dimers that could contribute both sides of the bipartite ATPase pocket (Guan et al. 2017).

3.6.1 Homology models of 2C monomers

3.6.1.1 Procedure

The sequences of several different enterovirus strains were provided by Lisa Bauer based on the availability in their research laboratory. For the computational investigations these sequences were retrieved from UniProt (www.uniprot.org). Sequence alignment with the crystal structure (PDB: 5GRB chain A; Uniprot entry: B9VUU3_9ENTO position: 1227-1440) was performed using the alignment tool in the MOE sequence editor (Chemical Computing Group Inc. 2016). It is of note that the first homology model was made using the EV-A71 BrCr strain (Uniprot entry: Q66478) because the crystal structure strain had several differences. EV-A71 BrCr was also used as the enterovirus reference strain in the laboratory in Utrecht, therefore a correct homology model was of great value.

The homology model of the monomer was modelled using the same settings for all the homology models created for the project. The only change was the sequence of the desired virus for which the model should be created.

The sequences of several viruses of interest were retrieved under the Uniprot entry numbers listed in Table 74.

Table 74: Sequences of viruses retrieved for homology modelling

Virus	Entry ID	Name	Sequence Position
EV-A71	Q66478	POLG_HEV71B EV-A71 strain BrCr	1112-1440
CVB3	P03313	POLG_CXB3N CVB3 strain Nancy	1101-1429
Polio	NP_740473* P03301	NTPase 2C Enterovirus C POLG_POL1S (Polio type 1 strain Sabin)	1128-1456
EV-D68	NP_740744.1*	NTPase 2C Enterovirus D	1114-1443
Human Rhinovirus 14	P03303	POLG_HRV14	1101-1429
Human Rhinovirus 2	P04936	POLG_HRV2	1088-1409
EMCV	P03304	POLG_EMCV Encephalomyocarditis virus	1193-1517
Aichi-Virus		From Collaborator Linda Visser	

* these sequences were retrieved from <https://www.ncbi.nlm.nih.gov/protein/> using the identifier for the search

The workflow was then continued as follows: The structure template 5GRB chain A was protonated in MOE with the Protonate3D function (Labute 2008a). The sequence of interest was pasted into the sequence editor in MOE and aligned with the MOE-Align tool (Multiple sequence alignment, Chemical Computing Group Inc. 2016). The N-terminal part of the sequence was deleted to fit the length of the crystal structure sequence. The full 2C protein ranges from position 1112 to 1440 for EV-A71. The crystallised part ranges from position 1227 to 1440. Two differences were noted comparing the sequences of the crystallised 2C part and the sequences of the EV-A71 BrCr strain: E207A and K209A. These two differences would not need a homology model and could easily be mutated to the correct residues but it was a good practice to build the homology model to already have a roadmap for the following models.

In the homology model function of MOE, first a name was given to the current system, and the output .mdb file (database). To view the models after the process finished and to monitor the progress the box 'open database viewer' was ticked. The sequence and the structural template were selected. In the options for the modelling 'C-terminal and N-terminal outgap modelling'

was disabled, no atoms were selected to be included as environment for induced fit and automatic disulphide bond detection was disabled.

10 Models were generated each sampling the sidechains once at a temperature of 300K (default settings). The intermediate models were refined with medium refinement to a RMS Gradient of 1 using the GB/VI score for the models (Labute 2008b). The final model was not refined but Protonate 3D was applied. The forcefield was set to Amber12:EHT (Case et al. 2012; Gerber & Müller 1995) allowing for the modelling of proteins, nucleic acids and small molecules, using the R-Field for the calculation of electrostatics (Tironi et al. 1995).

The sequences of the viruses in consideration for homology models were truncated to the corresponding length of the EV-A71 2C protein resolved in the crystal structure, which lacks the N-terminal 115 residues. The sequences of the 2C proteins differ in length by one amino acid containing either 329 or 330 residues. For all remaining homology models the same procedure was applied as described above.

3.6.1.2 Results

The 2C monomer models were successfully generated for the abovementioned viruses. For the quality of the models the Ramachandran plot on MOE was inspected and the outliers were investigated. Furthermore, the SAVES server is a web-based platform for structure quality evaluation that is mainly designed for crystal structures, but it is also a useful tool to investigate homology models (<http://servicesn.mbi.ucla.edu/SAVES/>).

Table 75: Analysis of the monomer models on the SAVES server

Monomers								
	1	2	3	4	5	6	7	8
	EV71 5GRB.A	CVB3	EV68	HRV-14	HRV-2	Polio	EMCV	AichiV
	95.2	92.8	91.9	91.4	90.8	90.9	86.9	84.8
Ramach	3.8	6.2	6.2	6.7	5.8	5.7	8.9	9.5
	1.0	1.0	1.9	1.9	3.4	3.3	4.2	5.7
Errat	85.075	96.95	89.055	88.776	85.279	84.925	94.203	88.78
Verify3D	80.09	79.15	80.19	89.10	84.69	82.46	86.11	83.57
Prove	5.1	4.7 out	4.4 out	5.0	7 out	6.0 out	8.7 out	6.5 out
All values in %								

Ramachandran expected: 98% in favoured region, 2% in allowed region, no outliers
 ERRAT: high resolution expected quality factor 95% or higher, 2,5-3 Å resolution 91%
 Verify3D: % of aa \geq 0.2 in the 3D/1D profile; $>80\%$ pass

The resolution of the crystal structure used as the template for the homology models is 2.8 Å. Therefore, the results from the SAVES server account for the lower resolution in the evaluation of the template crystal structure, which is also reflected in the generated homology models. Only high-quality high-resolution structures can yield optimum homology models that are comparable with a crystal structure. With this consideration in mind the monomers were used for further studies without further processing.

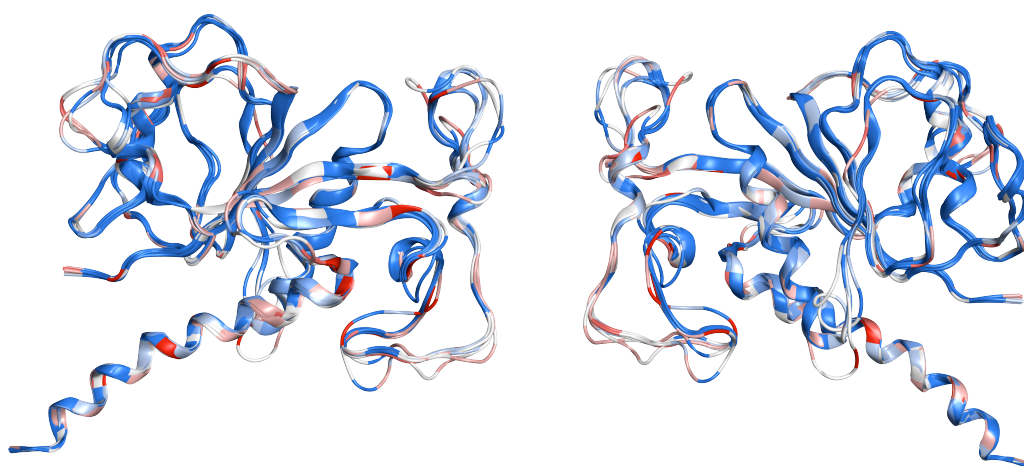


Figure 58: Superimposition of the homology models coloured by conservation of residues

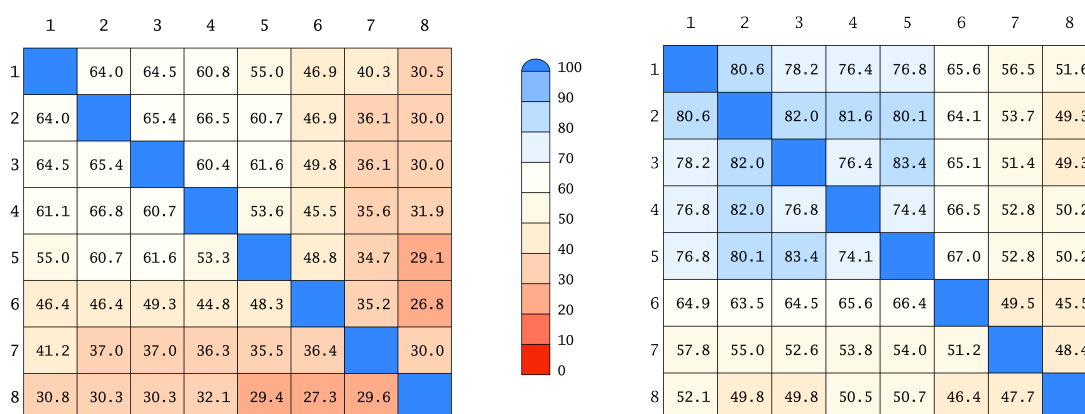


Figure 59: Identity (left) and similarity (right) between the homology models numeration according to Table 75

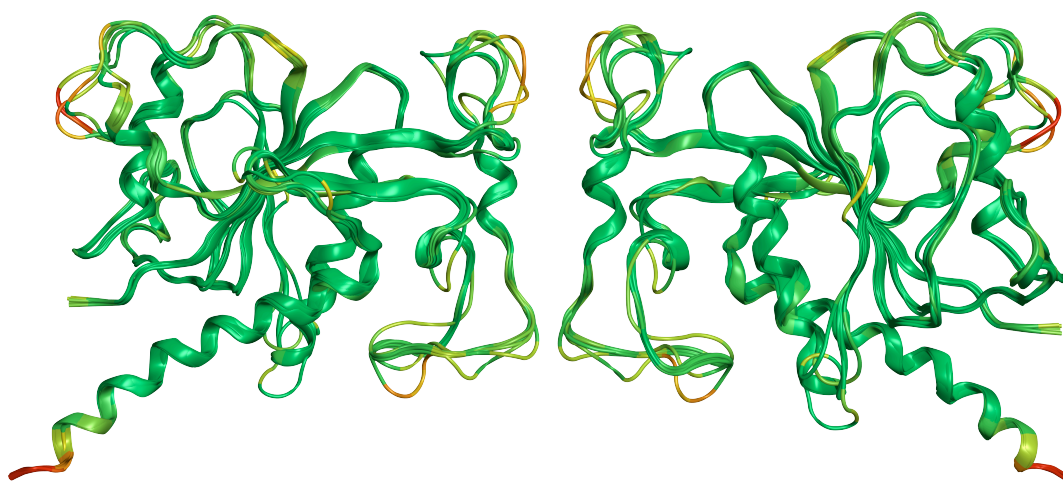


Figure 60: Superimposition of the homology models coloured by RMSD

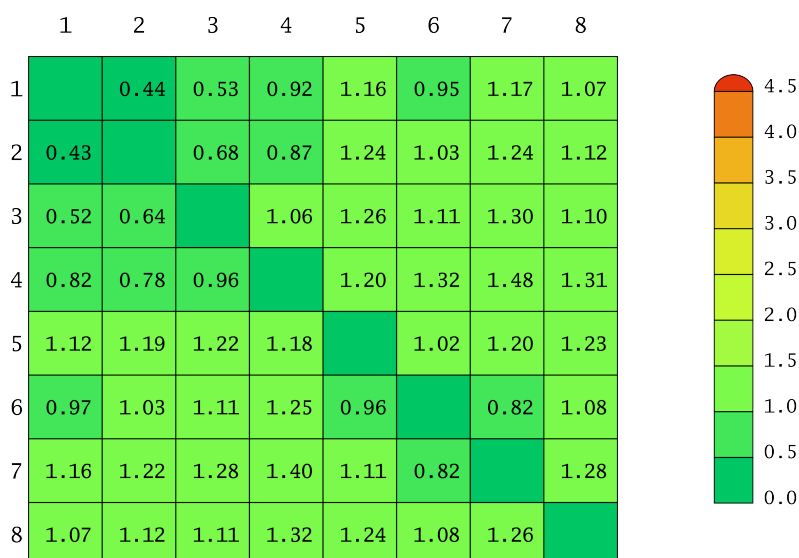


Figure 61: RMSD to the template (numeration according to Table 75)

3.6.2 Structural models of higher order

3.6.2.1 Dimer

After the homology modelling of the monomers and the investigation of the protein in greater depth, it became clear that a more meaningful model for the pursuit of drug discovery against 2C, but also for the understanding of the functions of the protein and the mode of actions of known inhibitors, would be a complex of higher order, at least a dimer. As described in the introduction to 2C, in 5GRB the chains A and F are arranged in a configuration that is similar to the one found in hexameric SF3 helicases. The hinge region of the C-terminal alpha helix allows for a movement that would close up the ATP binding site that is composed of residues from

two different monomers. Effects that go beyond one monomer, such as loop movements affecting more distant regions in the oligomer, could be investigated. The dimer is still sufficiently small as a system to subject it to molecular dynamics simulations that would give predictions of interactions between the monomers forming the bipartite ATP binding site, saving considerable computing time in comparison with a hexameric model.

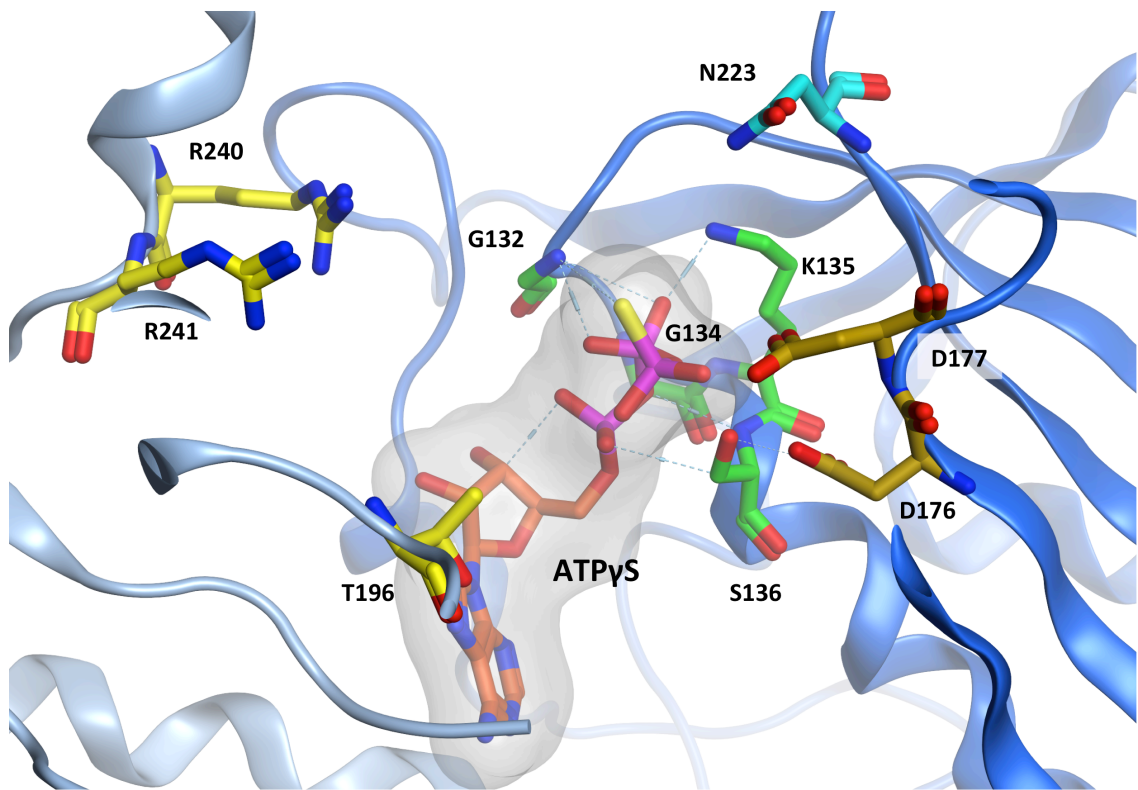


Figure 62: ATP binding site between chain A and F of 5GRB

In green the residues of the Walker A motif, Walker B in gold, C motif in turquoise on chain A in darker blue; R-finger(s) and residue T196 in yellow on chain F in lighter blue. ATP γ S in orange surrounded by a light grey surface.

3.6.2.1.1 Homology models

As described before the sequences for the relevant viruses were already downloaded for the monomer models. In order to model the dimers, the sequence for the to be modelled virus was duplicated. One sequence chain was paired with chain A the other with chain F of the structure template. The rest of the procedure is identical to the monomer models. In the model chains A and F were used as templates because they are the two chains where the C-terminal helix of chain A fits into a cavity of chain F. Furthermore, interactions between the ATP molecule and both chains are described in the publication of the crystal structure. In particular, T196 is to mention, which forms a hydrogen bond to the ATP molecule Figure 62. In the crystal structure the configuration is not optimal to show the interaction of the R-fingers

from the second monomer with the ATP but it is conceivable that in a hexamer they are interacting with the ATP molecule.

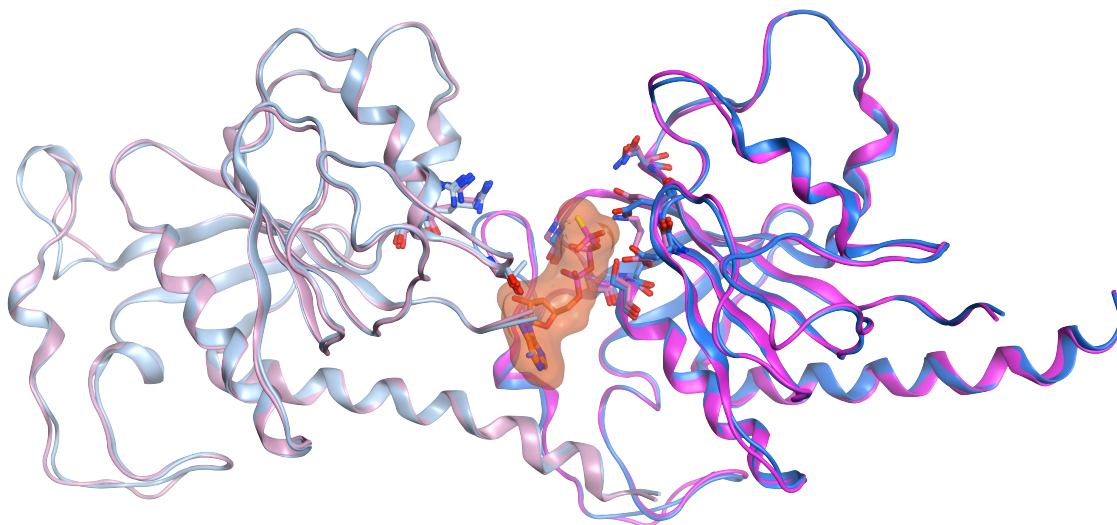


Figure 63: Dimer formed by chain A and F of EV-A71 and CVB3 superimposed
EV-A71 5GRB (blue) CVB3 model (purple) ATP γ S (orange); chain A dark coloured, chain F pastel

3.6.2.1.2 Molecular dynamics

In parallel to the homology modelling, the complex of chain A and F of the crystal structure 5GRB was subjected to molecular dynamics simulations. The protein was first prepared with the protein preparations wizard in Maestro (Sastry et al. 2013). Then the system for the molecular dynamics simulation with Desmond needed to be set up (Bowers et al. 2006; Schrödinger LLC & DE Shaw Research 2017). The complex was placed in a cubic box with at least 20 Å buffer distances from the protein at any point on the protein. The box was solvated using the TIP3P water model (Jorgensen et al. 1983). As a force field OPLS3 was selected (Harder et al. 2016). Charges on the protein were neutralised by adding Na⁺ or Cl⁻ ions and in order to simulate conditions also found in the biochemical assays with the protein, 0.01M MgCl₂ was added.

To start the MD simulation the simulation time was set to 100 ns, the number of frames to 625. For the main simulation the NPT ensemble was used at a temperature of 300K and a pressure of 1 atm. The system was relaxed with the automated relaxation protocol before the simulation. The initial speed for the simulation was chosen using a random assigned by the software.

The relaxation protocol executes several restrained MD simulation steps with different ensembles at a temperature of 10K. First, the system is equilibrated with a Brownian Dynamics NVT ensemble for 112 ps. The system is simulated for 48 ps at a constant pressure of 1 atm in

the NPT ensemble, 12 ps 10K restrained heavy atoms, 12 ps 300K restrained heavy atoms, 24 ps 300K unrestrained. After this, the actual simulation for 100ns starts with the parameters set during the launch.

3.6.2.1.3 Results

The 2C proteins of the viruses mentioned in Table 74 were modelled as dimers. These models were not used during the course of the project, because the monomers yielded sufficient insights and only molecular dynamics simulations and the selection of several representative frames would provide a suitable picture for further drug discovery efforts. In terms of quality of the models there were no significant differences between the results for the monomers and the dimers. Exemplified are the superimposed crystal structure dimer and the CVB3 dimer in Figure 63. Instead of the homology models, the molecular dynamics simulations of the crystal structure dimer of EV-A71 were later investigated for drug discovery purposes.

An ensemble docking approach on representative frames from the molecular dynamics on the EV-A71 dimer yielded a small molecule with activity against EV-A71, supposedly targeting the ATPase active site. These studies were performed by Roberto Manganaro based on the molecular dynamics simulations described above.

3.6.2.2 Hexamer

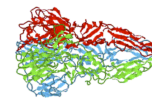
For the modelling of the geometric arrangement of the hexameric 2C protein the SymmDock server was used (<http://bioinfo3d.cs.tau.ac.il/SymmDock/php.php>) (Schneidman-Duhovny et al. 2005a; Schneidman-Duhovny et al. 2005b). This is a webserver that arranges a .pdb structure file with monomers or dimers into a specific higher order geometry depending on the input parameters. It gives 20 solutions as output, several of which might not correspond to biochemical and virological description or findings but only represent a reasonable geometry for the program.

3.6.2.2.1 SymmDock

The chain F monomer of EV-A71 was prepared in MOE for the submission to the SymmDock server (<http://bioinfo3d.cs.tau.ac.il/SymmDock/php.php>) (Schneidman-Duhovny et al. 2005a; Schneidman-Duhovny et al. 2005b) and saved in .pdb format. The first try was using dimers of chain A and F prepared for the input but this did not yield any reasonable geometry. For another try residues 318-329 were deleted, as an incorrect angle of the C-terminal helix at the kink region might result in clashes. Eventually, the desired geometry was obtained with the

complete chain F as monomer and the symmetry order of 6. The submission form of SymmDock is shown in Figure 64.

SymmDock



Prediction of Complexes with C_n Symmetry by Geometry Based Docking

[\[About SymmDock\]](#) [\[Web Server\]](#) [\[Download\]](#) [\[Help\]](#) [\[FAQ\]](#)

PatchDock - An Algorithm for Rigid Unbound Docking of Molecules

Unit Molecule: the asymmetric unit of the multimer complex (PDB format)

Type the PDB code (PDB-code:chain Id, e.g. 1f23:A)

or upload a file

Symmetry Order: (any number from 2 to 100: 3 for trimer, 4 for tetramer etc.)

e-mail address (the results are sent to this address)

Binding Site: binding site file (optional)

Distance Constraints: distance constraints file (optional)

Figure 64: Screenshot of the SymmDock website

An email address is required for the collection of the results. 20 solutions were returned as output. Not all the results corresponded to the biological findings and the mode of action of SF3 helicases, but the best ranking solution with the desired geometry was selected for subsequent work. In Figure 65 the hexamer output is shown. The ATP and ADP molecules were arranged to mimic a circular stepwise cleavage of ATP. This arrangement was an attempt to create different binding sites with different configurations of ATP, ADP and the empty pocket for the molecular dynamics simulations and might not conform with the actual mechanism the 2C helicase employs. Enemark and Joshua-Tor describe a mechanism of loading and cleavage of ATP in their review but other mechanisms might exist (Enemark & Joshua-Tor 2006).

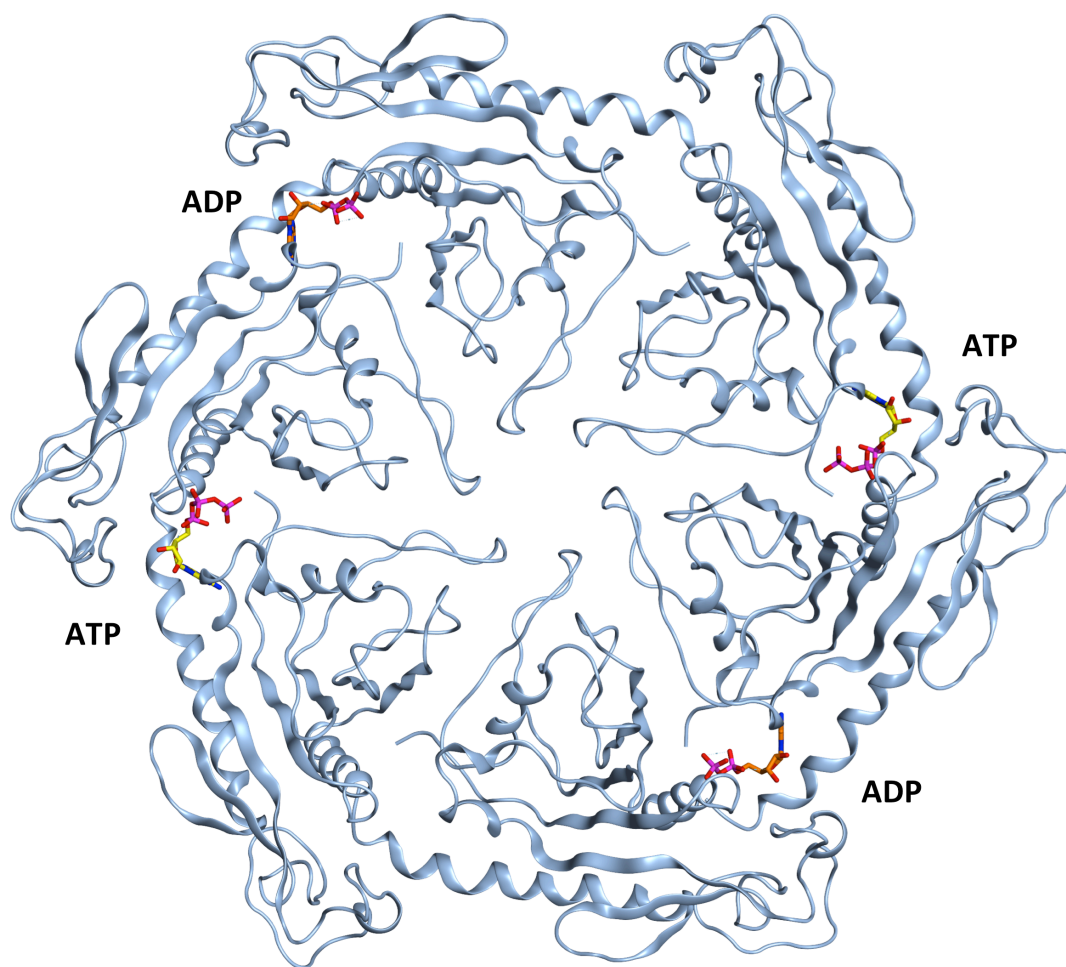


Figure 65: Hexamer output from SymmDock prepared for molecular dynamics simulations

3.6.2.2.2 Molecular Dynamics

The hexamer model was subjected to molecular dynamics in order to even out any artificial geometry that might have been due to the imposed symmetry. Near clashes and steric problems were resolved with the protein preparation wizard in Maestro (Sastry et al. 2013) before a valid molecular dynamics simulation could be started. The complex was then placed into a cubic water box 10 Å bigger than the extremities of the complex. TIP3P was used as water model to solvate the box (Jorgensen et al. 1983). As the force field for the simulation OPLS3 was chosen (Harder et al. 2016). Na⁺ atoms were added to the system in order to neutralise the charges on the complex and magnesium chloride (0.01 M) was added to the box to simulate physiological conditions. The relaxation protocol was run by default as described for the dimer.

The equilibrated system was then simulated for 100 ns at constant temperature (300 K) and pressure recording snapshots every 160 ps.

3.6.2.2.3 Results

One snapshot of the MD simulations shows an ATP molecule and a magnesium ion in the ATP binding site. Interactions with R240 are present as well. This suggests that the 2C protein forms indeed hexameric complexes to exert its functions. This geometry is favourable for the ATP cleavage and the representation in Figure 66 corresponds to the conformation reported for other SF3 AAA+ helicases (Enemark & Joshua-Tor 2006).

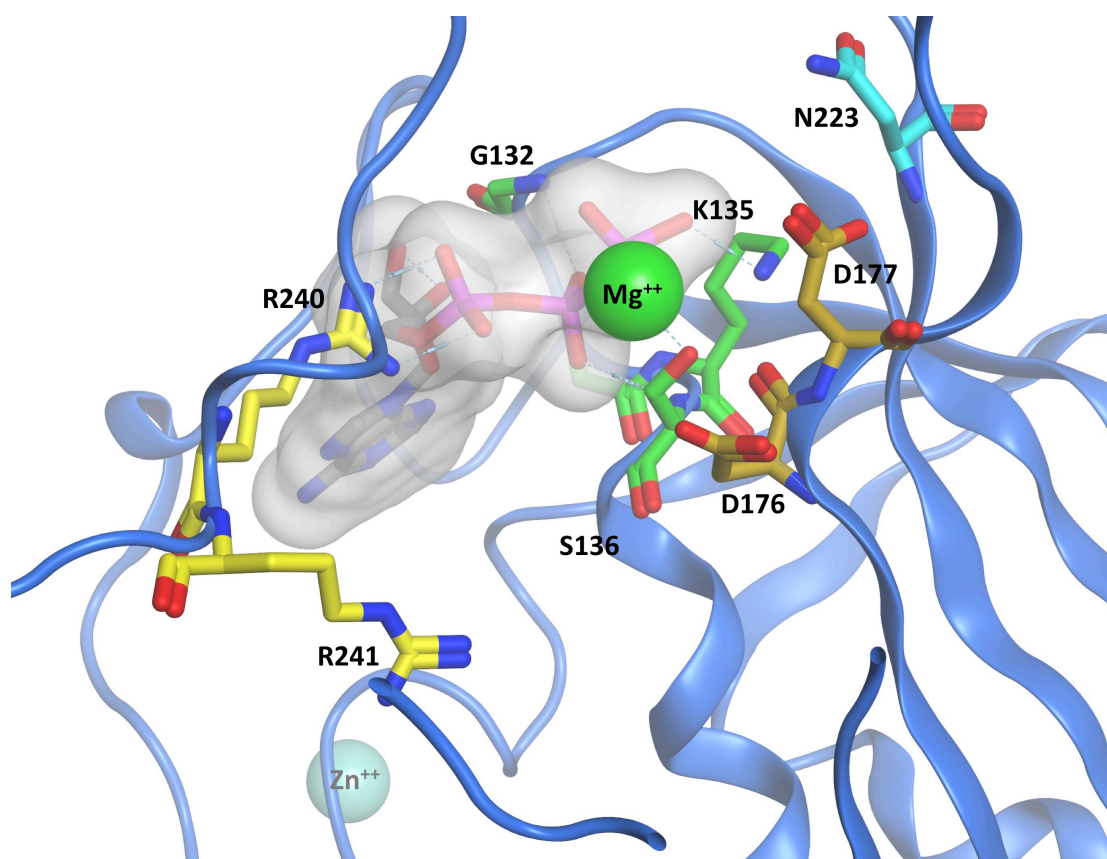


Figure 66: Configuration of ATP between chain A and F after MD simulations

3.6.2.2.4 Discussion

Helicases uses ATP to generate energy in order to exert a motor function (unwinding or moving along a nucleic acid strand). In the crystal structure the ATP analogue ATPyS was used in order to prevent a quick cleavage of ATP. For the modelling this was corrected although the first attempt to run the molecular dynamics simulations was still using ATPyS. How many of the bipartite ATP binding sites are occupied by ATP at a given time point and if there are more than one site occupied at the same time is still not known for 2C. In order to have a realistic model it would be necessary to allow for the cleavage of the terminal phosphate group of the ATP. This process usually requires divalent cations such as magnesium. Standard molecular dynamics simulations do not allow for bond breakage, thus this process could not be

monitored in our model. It was also not clear how a nucleic acid strand would be coordinated by the hexameric 2C complex. Therefore, the hexamer model was eventually only used to get a better understanding of the coordinating residues constituting the ATP binding site and other interactions between the monomers.

3.7 Mode of action of known 2C inhibitors

The next step in the project was aimed at detecting possible binding sites for the known 2C inhibitors and to understand and prove the interactions of the compounds within those pockets. In the beginning, the intention was to investigate as many 2C inhibitors as possible for their mode-of-action, but due to a lack of time and resources eventually in depth research was only possible for fluoxetine.

3.7.1 Site Finder

For further drug discovery and mode-of-actions studies on known 2C inhibitors, suitable binding sites on the protein needed to be detected. A tool designed to find potentially druggable cavities on protein structures is Site Finder that is incorporated in the MOE software suite. A detailed description of the methodology and procedure is found in the MOE manual (Site Finder, Chemical Computing Group Inc. 2014). It is based on the exclusion of regions on a protein that are tightly packed. An electrostatic- and tautomerisation-independent methodology also classifies the lipophilicity and hydrophilicity of the protein cavity described by so-called alpha spheres. The clustering of the spheres results in the identification of sites, which are then evaluated by the environment of amino acids and receive a propensity for ligand binding (PLB) score (Soga et al. 2007).

Site Finder was run on the EV-A71 dimer and on the CVB3 dimer homology model. Sites were selected for their rank in Site Finder but especially for their vicinity to mutation hotspots reported for the known 2C inhibitors. One region that features a particular density of mutations is the ²²⁴AGSINA²²⁹ loop in CVB3. Next to this loop Site Finder identified several pockets that are numbered according to their rank. The two pockets that were eventually used on EV-A71, CVB3 they will be termed site A and site B. They correspond to site 4 and site 5 in EV-A71 and to site 2 and 6 in CVB3 (see Figure 67 and Figure 68). The sites were compared for their accessibility on both EV-A71 and CVB3. Differences between the monomers and the different viruses are depicted below.

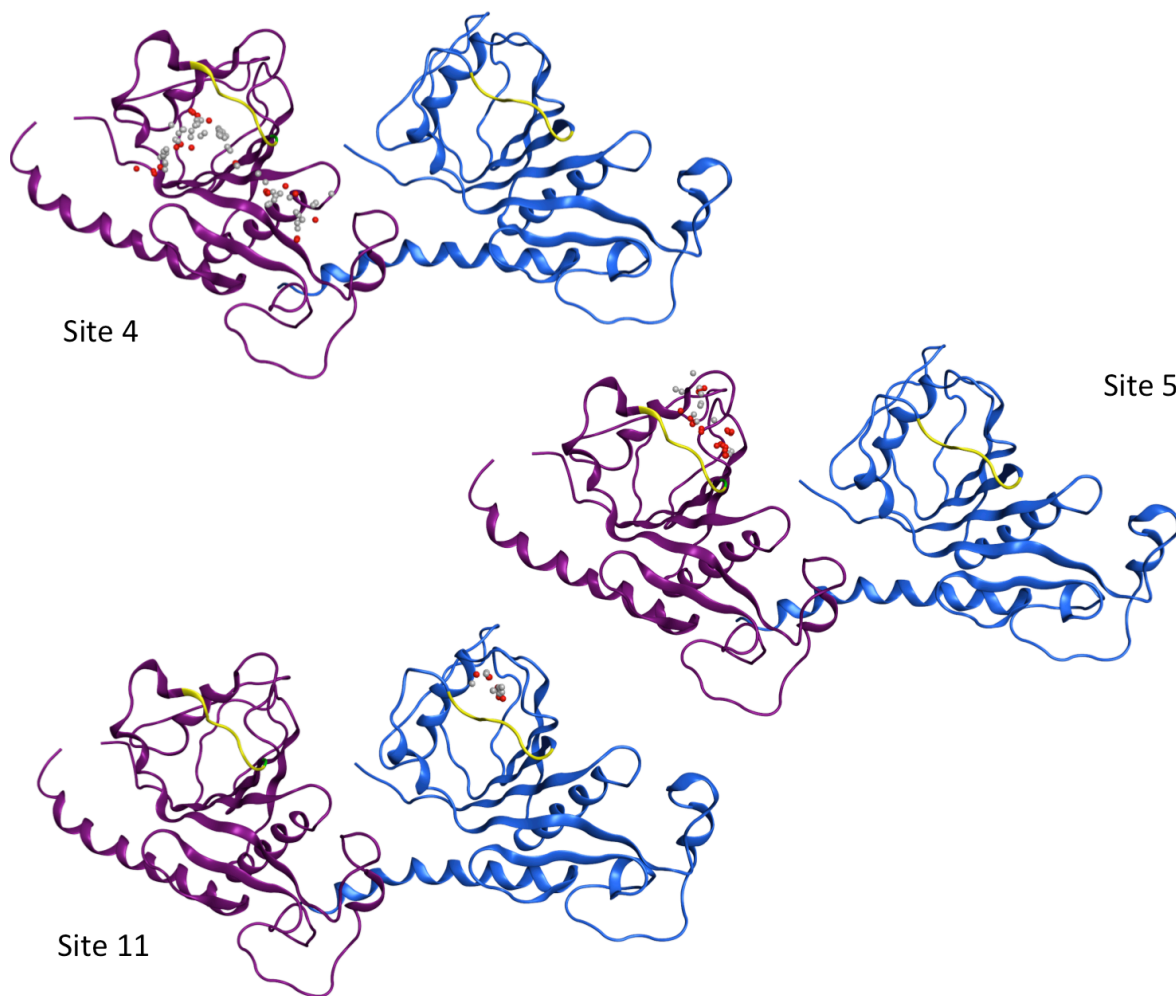


Figure 67: Site Finder results on the dimer of EV-A71

224-229 loop in yellow, sites as white and red spheres, site 4 and 11 are corresponding sites on different monomers

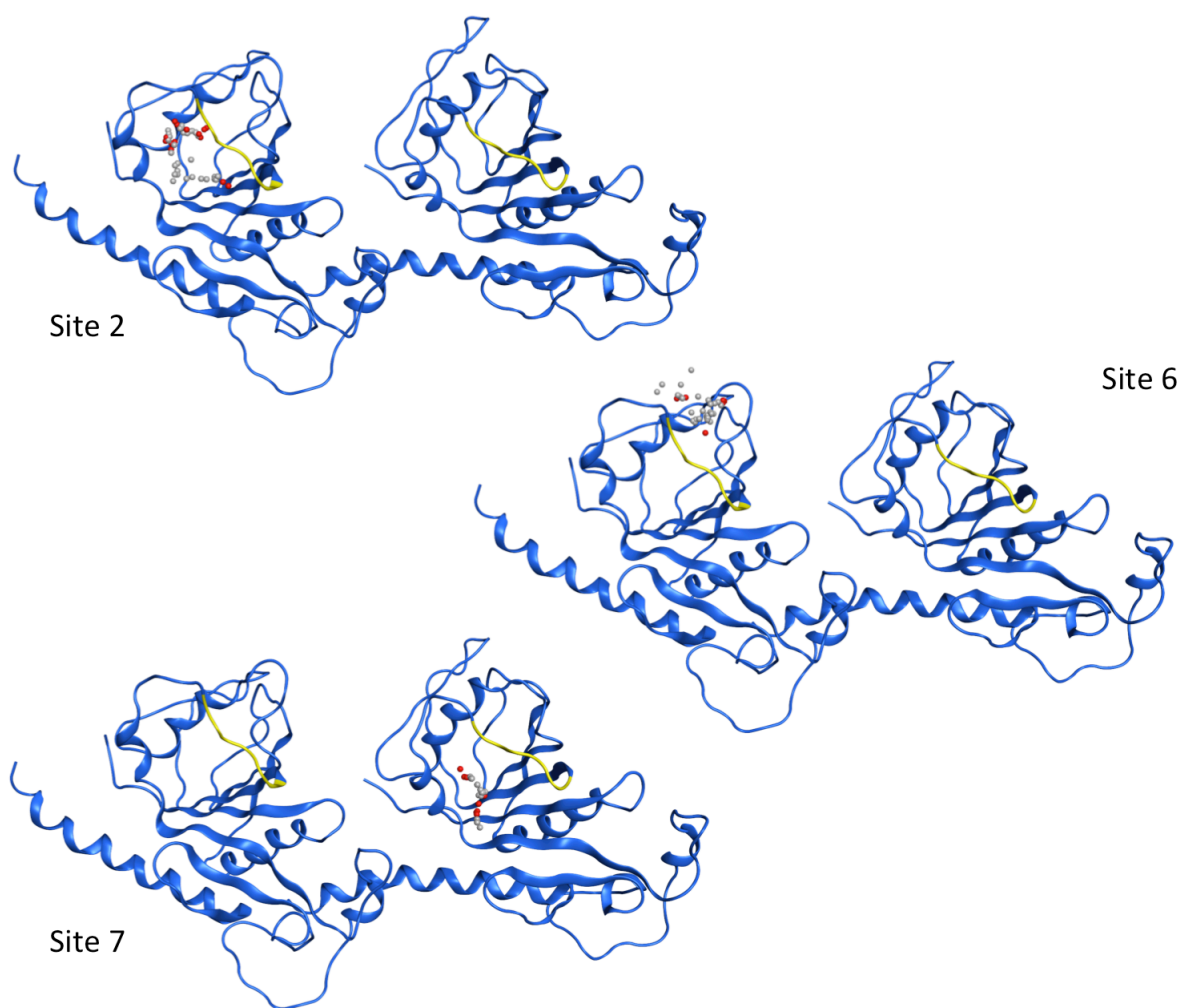


Figure 68: Site Finder results on dimer of CVB3 homology model
²²⁴AGSINA²²⁹ loop in yellow, site 2 and 7 are corresponding sites on the different monomers

3.7.2 Docking

In the next step the known 2C inhibitors (see section 3.2.3) were collected in a database and prepared for the docking. The compounds were docked into the binding site of CVB3 as this virus is more susceptible to the inhibition of several of the known compounds whereas EV-A71 is not sensitive to the known 2C inhibitors. Two sites were chosen for the docking: Site A in the region of site 4 with a centre of $x=-59.369$, $y=15.100$, $z=11.151$ and Site B was positioned at $x=-59.369$, $y=21.604$, $z=1.817$ for the docking on CVB3. These coordinates represent the centre points for each receptor grid that is required for the docking with Glide. Receptor grids were generated for both sites. Then, the compounds that were docked needed to be prepared with the ligprep protocol in Maestro. Default settings were used and up to 32 conformations were generated for each compound. All obtained conformations were then docked in both site A and site B with Maestro Glide in standard precision mode. The poses were visually inspected for their fit in the binding pocket and their interactions with the protein. The best protein-ligand complex for each site was then saved and prepared for molecular dynamics simulations,

mainly because of the potential flexibility of the loops that cannot be accounted for in docking studies with a rigid protein. Further studies were performed only on (R)- and (S)-fluoxetine. In Figure 69 the poses for both fluoxetine enantiomers in both sites on CVB3 are depicted.

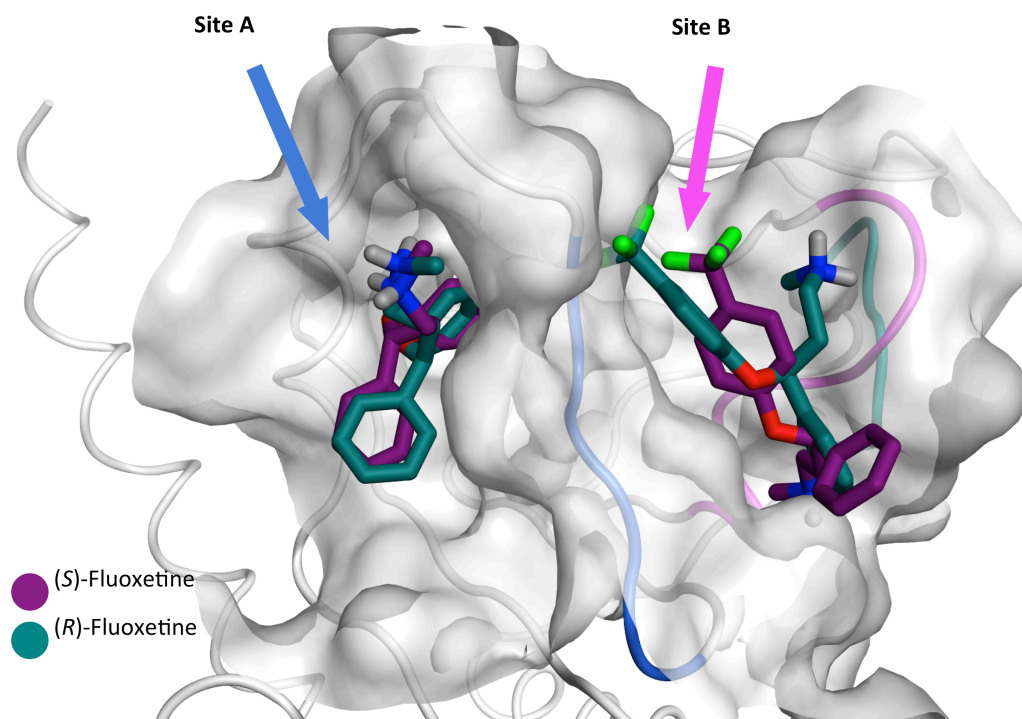


Figure 69: Docking poses of (R)- and (S)-fluoxetine into CVB3 into site A and B

3.7.3 Molecular dynamics of (R)- and (S)-fluoxetine

The four different complexes were subjected to molecular dynamics simulations. In Figure 69 the ²²⁴AGSINA²²⁹ loop is coloured in blue and it is thought that this loop is flexible. Therefore, both site A and site B might change in size and shape depending on the loop movement.

The monomers of 2C in which the compounds were docked were prepared with the protein preparation wizard in Maestro. Importantly the protein termini were capped. Then system for the molecular dynamics simulation with Desmond was set up. The protein ligand complex was placed into a cubic box that was then solvated using the TIP3P water model. The distances of the box from the extremities of the complex was set to minimum 20 Å. The default force field OPLS3 was used, the charges were neutralised with by default with Na⁺ or Cl⁻ ions and 0.01M MgCl₂ was added to simulate experiment conditions. The simulation was relaxed using the automated relaxation protocol running default settings: 112 ps of equilibration at 10 K in an NVT ensemble, simulation of 48 ps at constant pressure of 1 atm in the NPT ensemble. The main simulation was run for 100ns keeping the temperature constant at 300 K, saving snapshots/frames ever 160 ps.

The command line script `thermal_mmgbsa.py` was used to calculate the predicted $\Delta G_{\text{binding}}$ energies, slicing the trajectory into frames and calculating the average binding energies for the ligand at a given snapshots. Each complex was simulated three times creating different runs by assigning a random seed value for the initial velocity.

Table 76: Calculated binding energies (kcal/mol) of the protein ligand complexes during molecular dynamics simulations

	Compound	MD1	MD2	MD3
Site A	(R)-fluoxetine	-29.71*	-42.70	-19.85*
	(S)-fluoxetine	-41.63	-42.28	-41.54
Site B	(R)-fluoxetine	-29.59*	-42.95	-27.38
	(S)-fluoxetine	-34.79	-29.88	-30.83*

The results from the $\Delta G_{\text{binding}}$ values averaged over 100 ns of simulation for each independent repetition of the molecular dynamics simulation for each site and each compound are depicted in Table 76. The asterisk indicates that the compound detached from the binding site during the simulation and such an event is reflected in the shift of the binding energy towards higher values. This leads to the conclusion that the complex (S)-fluoxetine in site A seems to be the most favourable complex.

Visualising the interactions of the ligand in the binding site after the simulation supports this conclusion. (S)-fluoxetine forms a series of interactions with polar sidechains at the rim of the pocket, while docking deeply into the hydrophobic pocket with its trifluoromethylbenzene moiety. A snapshot of one of the late stages in the molecular dynamics simulation of (S)-fluoxetine in site A is shown in Figure 70.

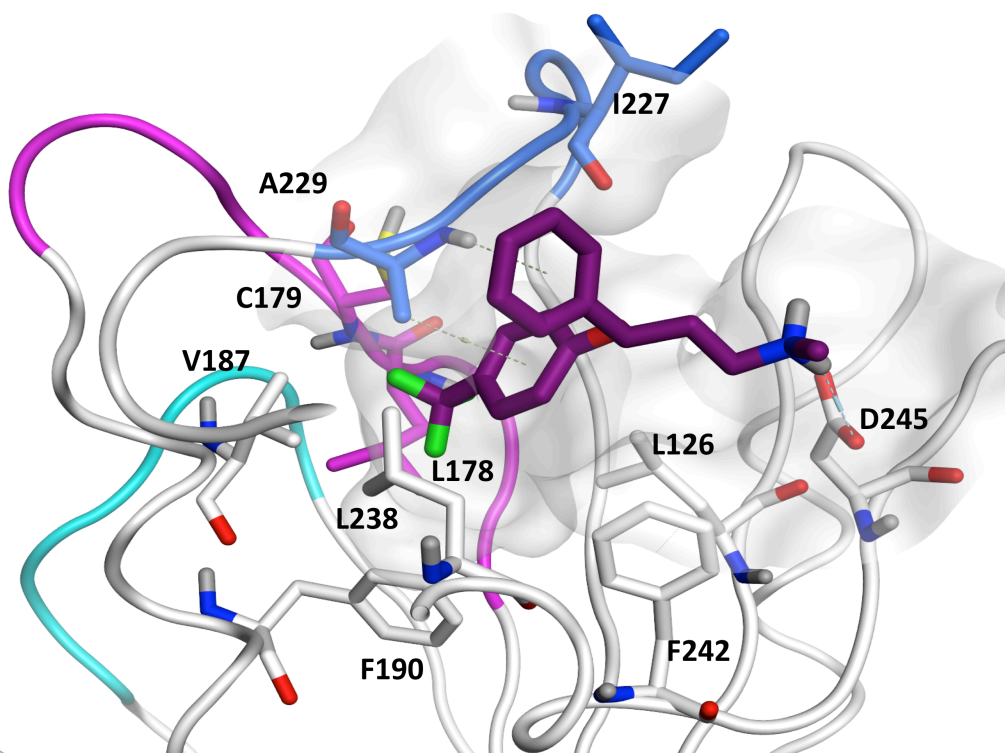


Figure 70: (S)-fluoxetine snapshot during molecular dynamics simulation in site A

3.7.4 Confirmation of the modelling results

The molecular dynamics simulations clearly suggest a difference in binding energies of the two enantiomers of fluoxetine. To further investigate this hypothesis, both enantiomers and the racemic compound were provided to our collaborators in Marseille and Utrecht, for an in depth virological and biochemical evaluation of the enantiomers compared to the racemic mixture.

3.7.4.1 Biological evaluation

The two enantiomers of fluoxetine were tested individually for their activity against CVB3 and in comparison to the racemic mixture. Ulferts et al. reported an activity for the racemic mixture of $3.36 \pm 0.47 \mu\text{M}$ on CVB3 in Vero cells (Ulferts et al. 2013), which could be confirmed in the assays performed by Lisa Bauer in a CPE-reduction multicycle assay with CVB3 where the racemic mixture resulted in an EC_{50} of $3.2 \pm 0.95 \mu\text{M}$ and a CC_{50} of $29.32 \pm 0.35 \mu\text{M}$ in HeLa R19 cells (Figure 71 left).

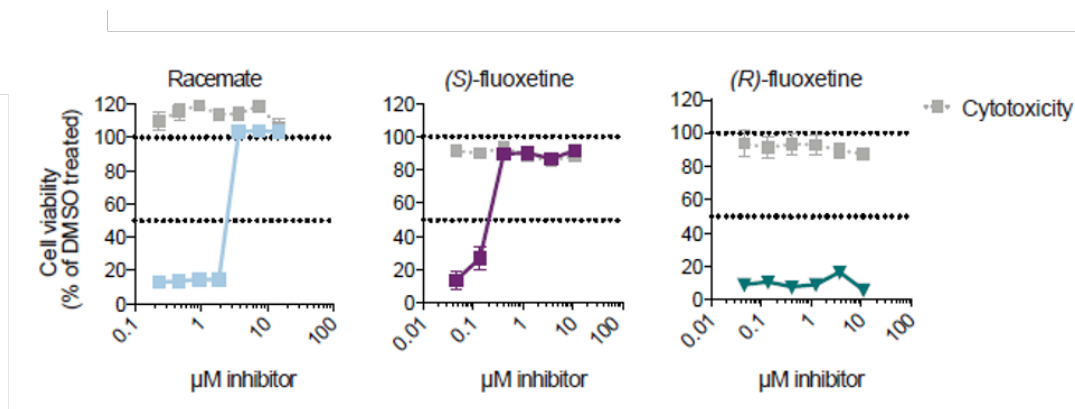


Figure 71: CPE reduction assay of racemic, (S)- and (R)-fluoxetine against CVB3
Figure provided by Lisa Bauer

Expectedly, the S-enantiomer of fluoxetine inhibited CVB3 more potently with an EC_{50} of $0.4 \pm 0.15 \mu\text{M}$, which is one log unit lower than the racemic mixture. The cytotoxicity is comparable with the racemic mixture and resulted in a CC_{50} of $28.63 \pm 1.40 \mu\text{M}$. The R-enantiomer did not inhibit the virus. (Figure 71 centre and right)

The results were validated in single cycle assays using an RLuc-CVB3 reporter virus treating the cells with serial dilutions of the different compounds for 7h. After the lysis of the cells the luciferase activity could be detected and was quantified. The cytotoxicity was determined in an MTS assay and in HeLa R19 cells it ranges between 23 and 28 μM . The activity of the S-enantiomer ($EC_{50} = 0.42 \pm 0.17 \mu\text{M}$) was 5-fold higher than the racemic mixture ($EC_{50} = 2.02 \pm 0.94 \mu\text{M}$). The R-enantiomer did not exert any antiviral effect. Data is plotted in Figure 72.

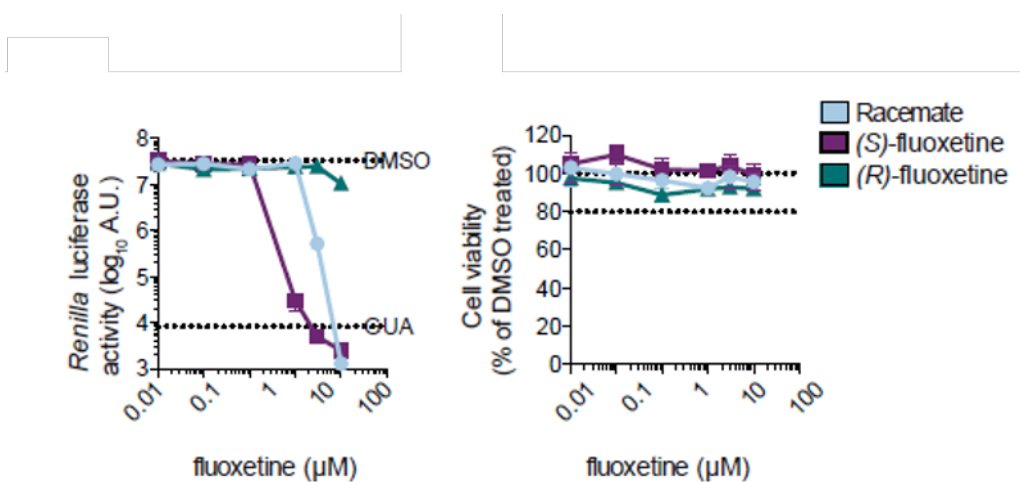


Figure 72: Single cycle assay of racemic, (S)- and (R)-fluoxetine against RLuc-CVB3
Figure provided by Lisa Bauer

3.7.4.2 Evaluation of six fluoxetine fragments for necessary chemical features

Then, fluoxetine was dissected into six fragments that present smaller molecules with a reduced set of features. The reasoning behind this was to see if a less complex molecule could

exert a similar effect like fluoxetine and to get an indication, which might be the most important features of fluoxetine. The resulting fragments are depicted in Figure 73. Roberto Manganaro and his Erasmus student Moira Lorenzo Lopez synthesised fragments **F1**, **F2** and **F4**, whereas the other three were commercially available and therefore purchased. Fragments **F3** and **F6** were obtained from Sigma Aldrich and fragment **F5** from Alfa Aesar. For fragment **F3** the racemic mixture was used. All six fragments were at least 95% pure.

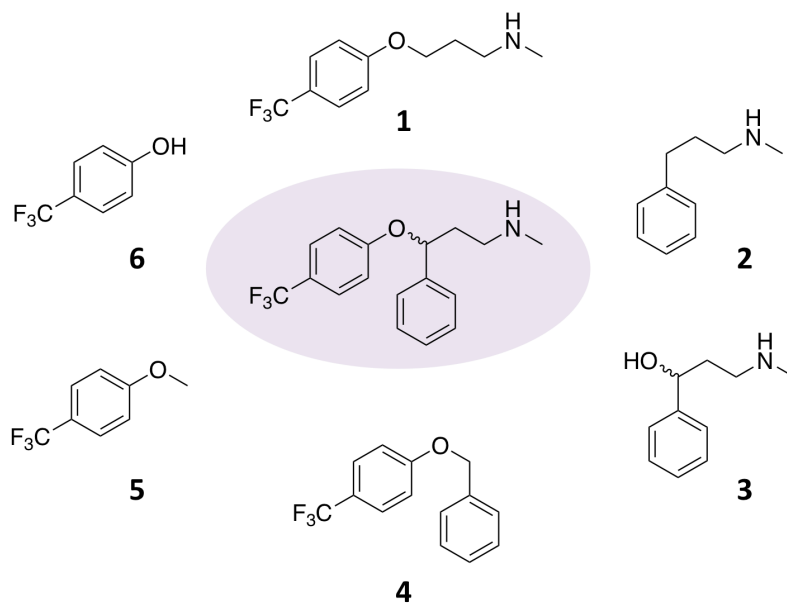


Figure 73: Fluoxetine and its fragments (F)

Biological evaluation of the fragments showed some activity in the multicycle assay only for fragment **F1**. According to the graph in Figure 74 it is questionable if the compound exerts its activity through a specific antiviral effect or by cytotoxic effects on the cells.

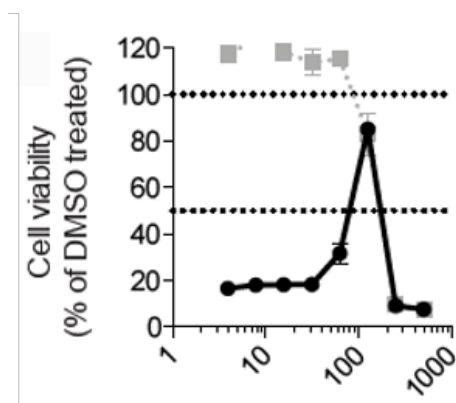


Figure 74: CPE-reduction assay for fragment F1 (activity in black, cytotoxicity in grey)

To exclude an antiviral effect due to toxicity, the fragment was tested in a single cycle assay against the RLuc-CVB3 virus. The cells were treated with 100 μM and 10 μM of fragment **F1**. An antiviral effect at 100 μM was indeed detected. To exclude any unspecific effects, fragment **F1** was also tested on fluoxetine insensitive EV-A71 and *Renilla* luciferase-expressing encephalomyocarditis virus (RLuc-EMCV). The viruses were not inhibited, thus the observed effect can be attributed to a specific antiviral effect against CVB3, although weak.

3.7.4.3 Biochemical evaluation

3.7.4.3.1 Thermal shift assay

In the thermal shift assay (TSA) the binding of the two fluoxetine enantiomers and the racemic mixture to the purified CVB3 2C protein was evaluated. Full-length 2C protein proved to cause problems during protein production and purification. Deletion of 36 amino acids from the N-terminal of the protein provided a homogenous monomeric protein preparation, which is necessary for binding assays like TSA and isothermal titration calorimetry (ITC).

During the TSA the melting temperature of the protein T_m is measured and shift towards higher temperatures in presence of the compound indicate binding and stabilisation of the complex. The racemic mixture induced a dose-dependent increase in T_m over a concentration range of 10 μM to 250 μM after which the melting temperature decreased. (S)-fluoxetine alone stabilised the protein in a purely dose-dependent manner without the destabilising effect at higher concentrations. (R)-fluoxetine seemed to be responsible for the destabilising effect, as can be observed in Figure 75.

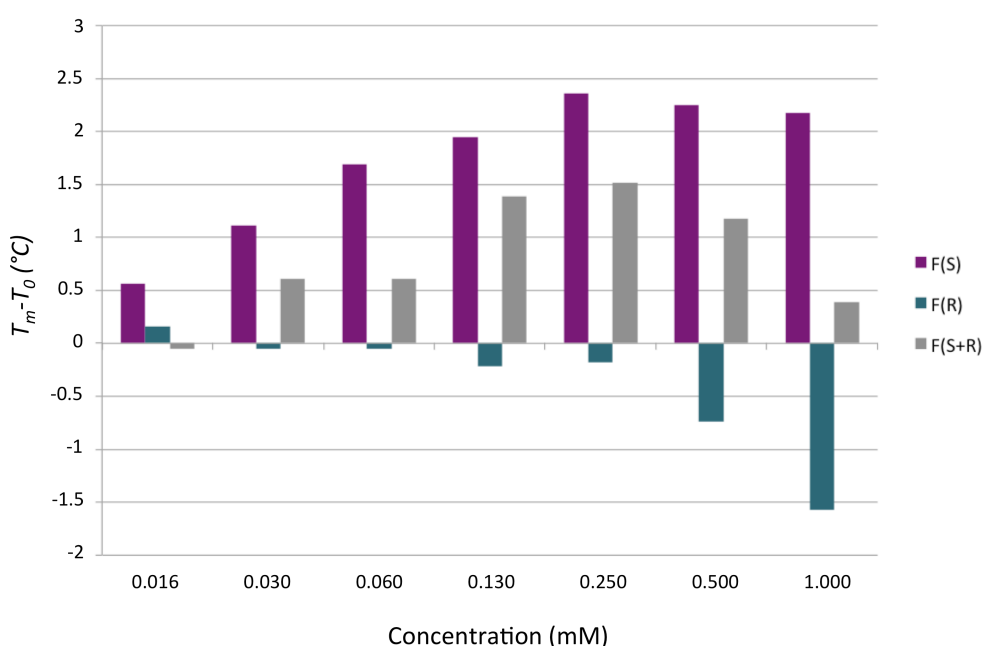


Figure 75: TSA of the fluoxetine enantiomers and the racemic mixture

The fluoxetine fragments were also evaluated for their binding to 2C protein. Due to the low molecular weight that is inherent to fragments the effect on the melting temperature is usually less pronounced and normally in the range of maximum 1°C. Furthermore, high concentrations are required to observe a stabilising effect (Coutard et al. 2014). In the present assay setting the compounds were tested at concentrations between 100 and 400 µM, but none of the fragments showed any effect on the melting temperature or stabilisation of the protein.

3.7.4.3.2 Isothermal titration calorimetry

To confirm the indicative results of the TSA an isothermal titration calorimetry assay was performed with the goal to determine the dissociation equilibrium constant (K_d) for the two enantiomers of fluoxetine. (S)-fluoxetine bound with a K_d of 9.5 µM, whilst the determination of K_d for (R)-fluoxetine was hampered by the fact that 2C aggregated partially during the titration with the compound. This would be again indicative for the destabilising effect of (R)-fluoxetine on the 2C protein of CVB3. K_d curve needed to be fitted to a different model and the resulting dissociation equilibrium constant was calculated to be around 200 µM. The data is found in Figure 76. The results strongly suggest that only (S)-fluoxetine is responsible for the binding to 2C protein.

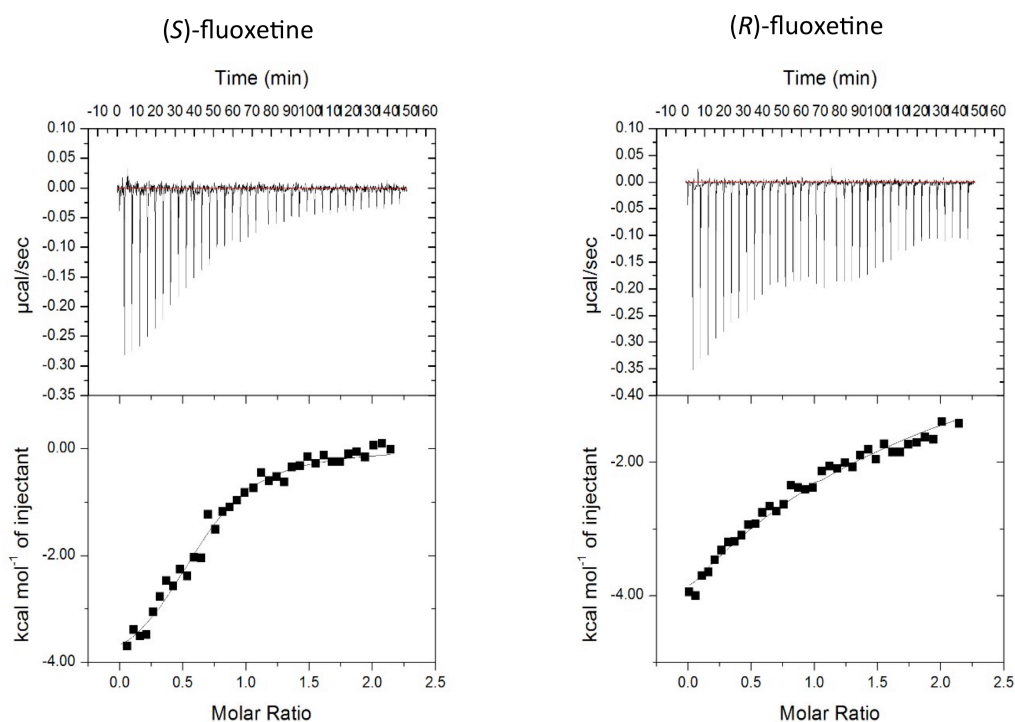


Figure 76: ITC measurements for (S)-fluoxetine and (R)-fluoxetine

3.7.4.3.3 Evaluation of (S)-fluoxetine against a panel of enteroviruses

In the next step, the antiviral activity of (S)-fluoxetine was determined for a panel of different enteroviruses. Activity of the racemic mixture was previously only demonstrated against CVB3 (Nancy) and EV-D68 (Fermon) (Zuo et al. 2012; Ulferts et al. 2013). The enantiomerically pure (S)-fluoxetine was not only more active on the already assessed viruses, but some activity could also be detected against rhinovirus 2 (HRV-2) and rhinovirus 14 (HRV-14) at concentrations of 7.95 ± 0.39 and 6.34 ± 1.02 μM , respectively. A summary of the results is reported in Table 77. (S)-fluoxetine failed to inhibit EV-A71 (BrCr) and poliovirus (Sabin). Concentrations up to 30 μM were tested, beyond this fluoxetine proved cytotoxic.

Table 77: Evaluation of (R)-, (S)- and racemic fluoxetine against a panel of enteroviruses (values in μM)

Virus	Species	Strain	114	(S)- fluoxetine	(R)- fluoxetine	SI _{racemic}	SI _{(S)- fluoxetine}
EV-A71	EV-A	BrCr	NA	NA	NA	NA	NA
CVB3	EV-B	Nancy	2.02 ± 0.52	0.42 ± 0.17	NA	14.51	71.56
PV-1	EV-C	Sabin1	NA	NA	NA	NA	NA
EV-D68	EV-D	Fermon	1.85 ± 0.10	0.67 ± 0.22	NA	21.72	42.73
HRV-2	RV-A		NA	7.95 ± 0.39	NA	NA	3.60
HRV-14	RV-B		NA	6.34 ± 1.02	NA	NA	4.52
CC ₅₀			29.32 ± 0.35	28.63 ± 1.40	23.63 ± 1.4		

3.7.4.3.4 Mutations in the predicted pocket render CVB3 resistant to (S)-fluoxetine

At the beginning of this study several mutations on the 2C protein were already reported for the different identified 2C inhibitors. For the racemic fluoxetine the triple mutant A224V-I227V-A229V (also AVIVAV) was reported to arise in CVB3. This mutant also provides resistance to other 2C inhibitors. Treating the triple mutant with (S)-fluoxetine alone confirmed the expected resistance. To probe whether all three mutations are needed to provide this resistance or not, single mutants were tested. The A224V single mutant did not render the virus resistant to (S)-fluoxetine. The I227V mutation, on the contrary, seemed to be the most important contributor to the resistance. For A229V dependence on the antiviral compounds was reported for known 2C inhibitors such as GuaHCl, HBB, TBZE-029 and MRL-1237 (De Palma et al. 2008), so that this mutant virus evolved to require the said compounds for efficient replication. This dependency did not extend to (S)-fluoxetine and the A229V mutation could not provide resistance against (S)-fluoxetine either.

For an evaluation of the binding pocket that was predicted as previously described, yielding a stable complex with (S)-fluoxetine, several mutations were introduced at crucial residues within the pocket. Two mutations at the deep hydrophobic pocket were introduced: C179F and F190L. Both viruses were highly resistant to (S)-fluoxetine. The C179Y mutation was previously raised by other 2C inhibitors as well as the F190L mutation was described by Thibaut et al. for a novel 2C inhibitor (Zuo et al. 2016; Thibaut et al. 2012). Then V187M was introduced because of sequence comparison between CVB3 and EV-A71 that contains a methionine at this position. No resistance against (S)-fluoxetine was observed. During the molecular dynamics simulations the methylamino moiety of (S)-fluoxetine interacted several times with D245 as it is also depicted in Figure 70. Mutation of D245 to asparagine did not render the virus resistant to (S)-fluoxetine either. The two last mutations were introduced into the RLuc-CVB3 virus. (S)-fluoxetine was compared to the controls GuaHCl and BF738735, which is a inhibitor of enterovirus replication that acts independently of the 2C protein as a host-directed antiviral.

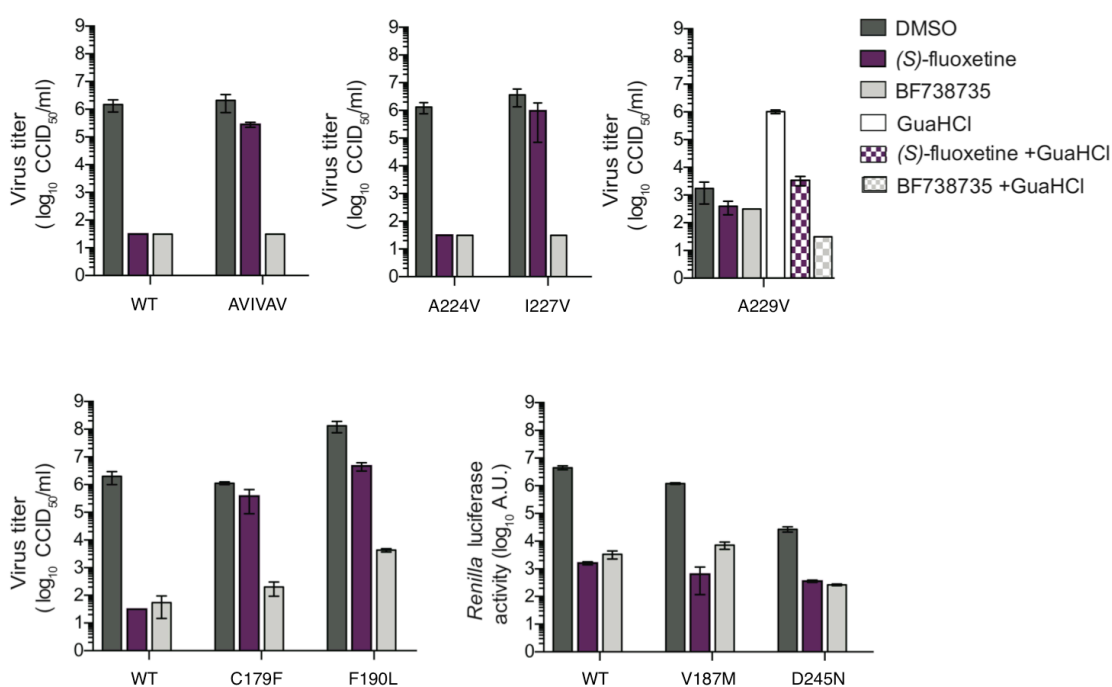


Figure 77: Mutation studies for (S)-fluoxetine to validate the potential binding pocket
Figure provided by Lisa Bauer

The recombinant 2C proteins carrying the A227V single mutation, the C179F or the A229V substitutions were expressed and purified for an evaluation in the TSA together with (S)-fluoxetine. As expected, none of the substitutions was able to form a complex that was thermally more stable in the presence of (S)-fluoxetine suggesting that the compound is not

binding to the protein. Interestingly, C179F was more sensitive to thermal denaturation in the presence of (S)-fluoxetine.

In conclusion, mutations in position 227, 229, 179 and 190 resulted in resistant viruses against (S)-fluoxetine. The location of these residues suggests that the interaction might actually not take place in site A but potentially in site B. During the visual inspection of the trajectory it was observed that the fluoxetine adopted a position that resembles the one in site A, burying the trifluoromethyl moiety deep in the hydrophobic pocket and exposing the methylamino moiety to the solvent Figure 79.

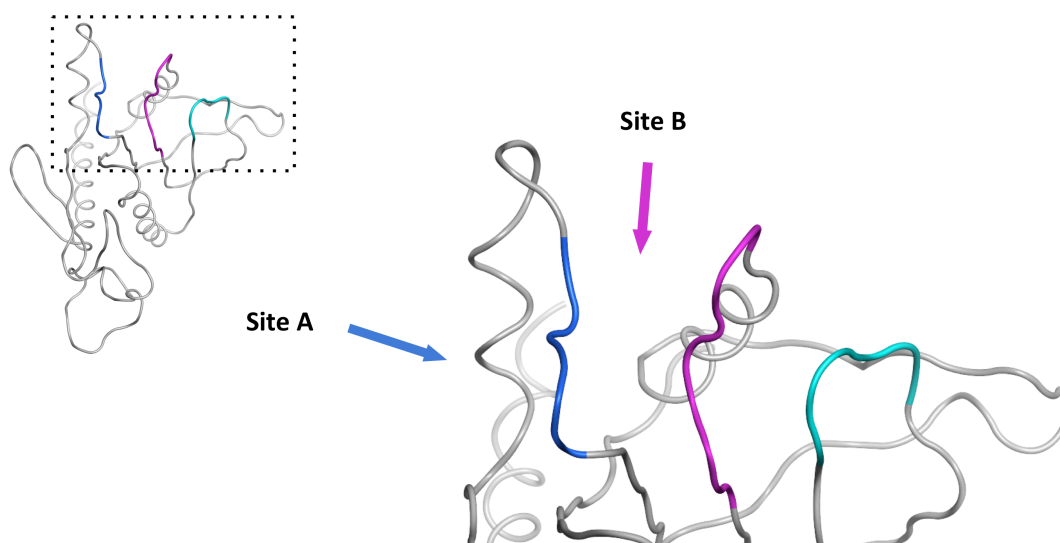


Figure 78: 2C protein with focus on site A and site B
²²⁴AGSINA²²⁹ loop blue, loop 175-183 pink, loop 158-163 turquoise

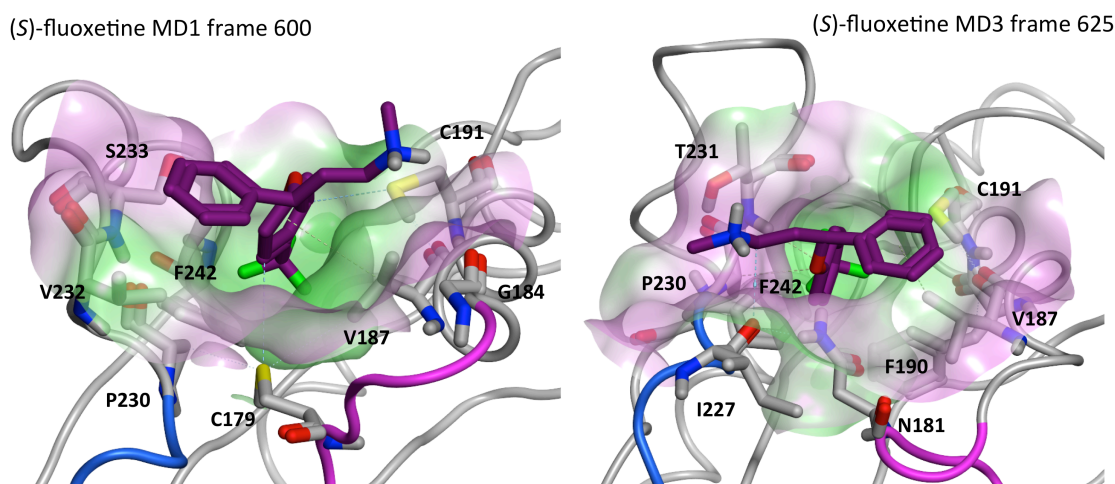


Figure 79: Close-up on site B with (S)-fluoxetine in MD1 frame 600 and MD3 frame 625
 Surface colour lipophilicity (green lipophilic, pink hydrophilic), Ribbons coloured as in Figure 78

3.7.5 Preliminary crystal structure

While the mutation studies, testing and modelling was on-going, Professor Bruno Coutard also attempted to crystallise the CVB3 2C protein. The protein alone did not yield any crystals but in co-crystallisation with (S)-fluoxetine a CVB3 2C structure was obtained. Like the one of Guan et al. it was truncated. This work is still unpublished and under refinement and validation, but the preliminary structure revealed a different binding site for (S)-fluoxetine compared to the models.

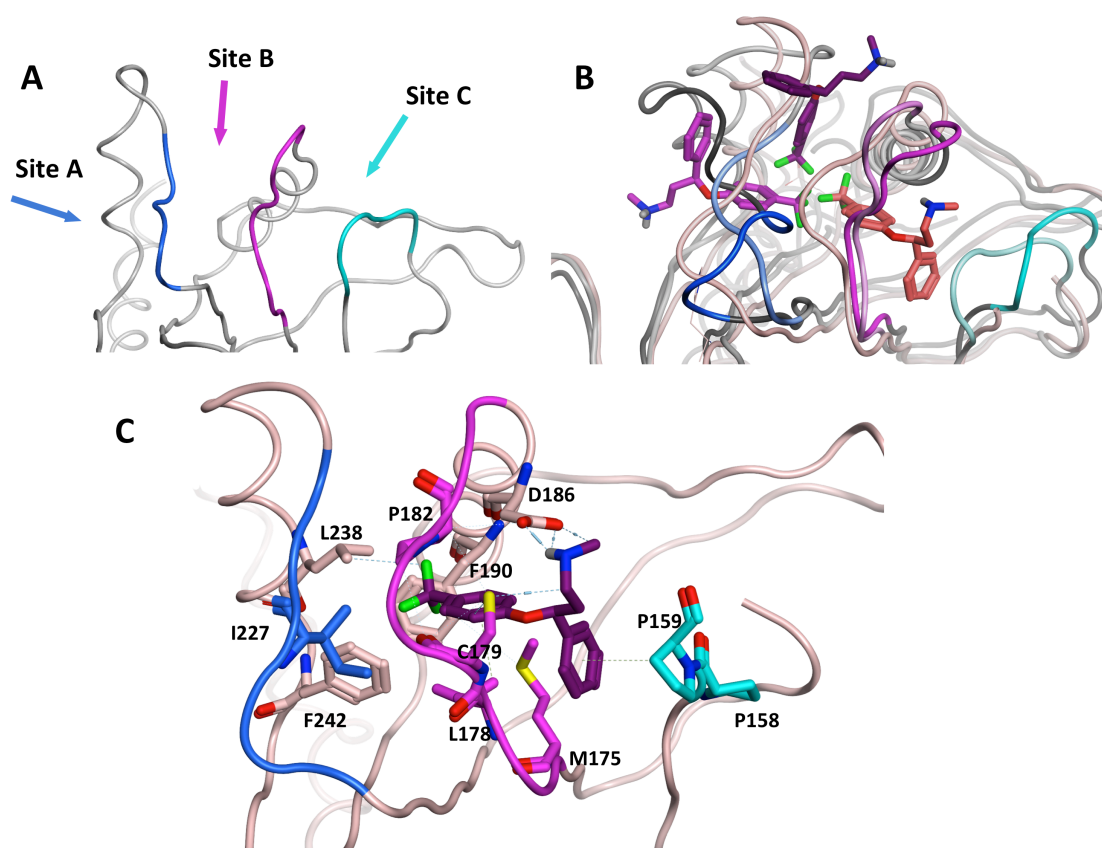


Figure 80: (A) Three potential binding sites in 2C (B) Superimposition of MDs on site A and site B with the preliminary crystal structure (C) Pocket view of the crystal structure with relevant residues

Panel A in Figure 80 shows the view of the three potential pockets for (S)-fluoxetine based on the hypothesis that the three coloured loops show a certain flexibility. Panel B depicts a superimposition of the molecular dynamics simulations on site A and site B for (S)-fluoxetine with the preliminary crystal structure obtained from Prof. Bruno Coutard (unpublished data). The trifluoromethyl moiety is positioned nearly on the same spot for the three entrance directions. In panel C the most important pocket residues of site C are depicted and draw a picture on why the mutations I227, C179 and F190 are responsible for resistance against fluoxetine.

3.7.6 Summary and Conclusion

In this part of the work, starting from the newly released EV-A71 2C crystal structure, first a homology model was created. In order to identify pockets that might be suitable for known inhibitors to bind to, the tool Site Finder was used and two sites flanking the ²²⁴AGSINA²²⁹ loop were identified. The reasoning for choosing the ²²⁴AGSINA²²⁹ loop as a first point for investigation was the report of several inhibitors including fluoxetine that induce mutations among this loop, the most important of which turned out to be I227V. Fluoxetine was previously identified and tested as racemic mixture. For the docking and computational modelling the enantiomers were specified and all further work was performed on each enantiomer separately. Docking of (R)- and (S)-fluoxetine seemed more favourable in site A which was confirmed by the molecular dynamics simulations where at least (S)-fluoxetine was able to interact in a more stable way. Overall, the S-enantiomer seemed to bind stronger to the protein.

The virological data revealed that (S)-fluoxetine exerted a 5-fold more potent antiviral effect than the racemic mixture, whereas (R)-fluoxetine did not show any antiviral activity in the CPE assays. To further evaluate whether (S)-fluoxetine was binding directly to 2C protein the compound was tested in TSA and ITC assays for its direct binding to CVB3 2C protein. These assays conclusively demonstrated a direct interaction for (S)-fluoxetine with 2C.

Fragments of fluoxetine that were synthesised or purchased to elucidate which chemical features are responsible for the antiviral effect. In the CPE assay only fragment **F1** showed a slight but specific antiviral effect. This effect could not be mirrored in biochemical assays, probably due to the fact that only small changes can be measured for fragments and high concentrations are required to see an effect. If binding occurred it might have evoked effects that were below the detection limit of these methods.

The mutation studies were guided by the visualisation of interactions of (S)-fluoxetine with amino acids on 2C. Mutation of residue D245 that remained without effect and the fact that A224V and A229V did not provide resistance to a virus with only the single mutations gradually led to the hypothesis that the predicted site A might not actually be the binding site for (S)-fluoxetine. Mutations C179 and F190 could affect site A as well but are also located in site B. Due to hydrophobic nature of the bottom of both binding sites the position of the trifluoromethyl moiety was believed to occupy more or less the same space regardless of the entry channel for both site A and site B.

When Prof Bruno Coutard present the preliminary crystallisation data of CVB3 2C in complex with (S)-fluoxetine, it surprisingly revealed a third entry channel to reach the hydrophobic region in the vicinity of F190. The identified mutations and all the other biological data were still inconclusive. From the modelling point of view the chance of identifying site C was quite

low, as the crystal structure does not present any cavity in this region of the protein. Although the flexibility of the loops was already taken into consideration the lack of a cavity or a clear opening of between loop 158-163 and 175-183 during molecular dynamics simulations on the apo-protein led to the belief that the proline-rich 158-163-loop might stay in a rather stable conformation compared to loop 175-183 and therefore not allow for compounds to induce a pocket. Only the collaborative efforts and the immediate feedback between all participants in this project allowed for such a quick and extensive progress towards the elucidation of the mode of action of (S)-fluoxetine as 2C inhibitor.

3.8 Design of novel 2C inhibitors with improved and broad-spectrum activity

3.8.1 Fluoxetine analogues

The aim to elucidate the mode of action of known 2C inhibitors and the in depth studies on fluoxetine led to the design of chemical compounds that could serve as molecular probes for the identified binding sites. The rationale behind the different modifications were driven by the pose of (S)-fluoxetine after the molecular dynamics simulations.

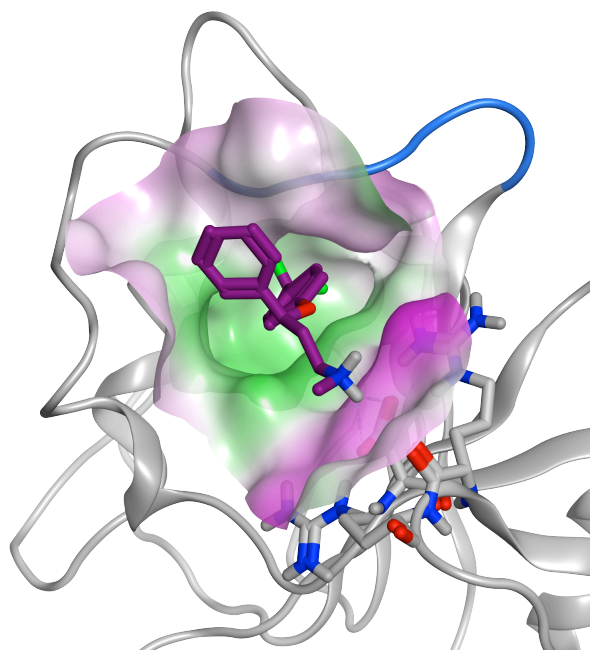
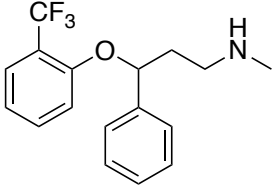
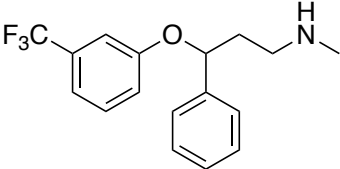
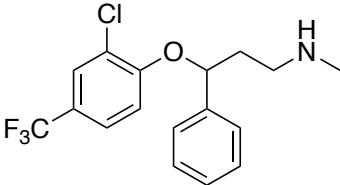
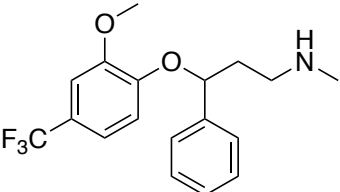
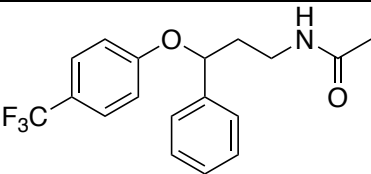


Figure 81: (S)-fluoxetine in frame 625 of MD2 interacting with D245

In Figure 81 it can be observed that on the pocket occupied by the trifluoromethylbenzene ring might provide some space for additional substituents in *ortho* or *meta* position. Therefore, the modifications presented in Table 78 are mainly *ortho*, *para* di-substituted or present a *ortho* or *meta* substitution on the trifluoromethylbenzene ring, only. From a medicinal chemistry point of view it was also interesting to see if the position of the substitution is relevant or if it can

even improve the binding and antiviral activity. This design was not based on docking studies of the newly designed compounds but solely on rational substitution planning after the investigation of the pose depicted in the figure above. Furthermore, modification of the methylamino part of the molecule was achieved by acetylation resulting in molecule **151**. The compounds in this part were synthesised by Roberto Manganaro and tested as racemic mixture in the multicycle CPE assay on CVB3.

Table 78: Fluoxetine analogues

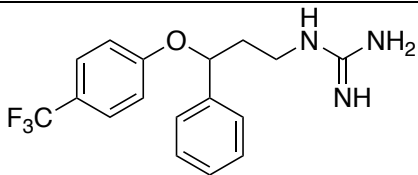
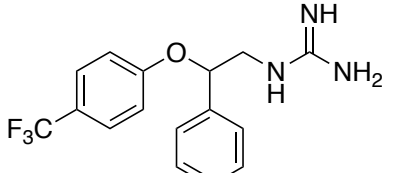
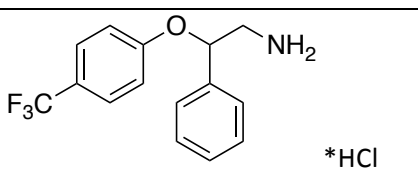
Nr	Structure	EC ₅₀ (μM)	CC ₅₀ (μM)	SI
114	Fluoxetine (racemic)	3.2 ± 0.95	29.32 ± 0.35	9.1
146	(S)-fluoxetine	0.4 ± 0.15	28.63 ± 1.40	71.6
147		NA	28.86 ± 0.86	-
148		NA	12.56 ± 1.79	-
149		NA	3.14 ± 0.07	-
150		NA	18.41 ± 1.26	-
151		NA	>30	-

Unfortunately none of the compounds designed to fill the binding pocket with different substitutions on the trifluoromethylbenzene ring of fluoxetine were active. Neither was the molecule with the modified amino chain. This might not be surprising because the pocket that is occupied by (S)-fluoxetine in the crystal structure is narrower than site A.

3.8.2 Guanidine analogues of fluoxetine

In a next step of modifications the rational was to combine the known 2C inhibitor Guanidine HCl with fluoxetine. In the figure from the molecular dynamics simulations the interaction of the amino group with D245 is flanked by two arginine residues. Introducing the guanidine group instead of the amino group might yield a better interaction with D245 and/or displace one of the arginine residues. Roberto Manganaro therefore synthesised molecules **152-154**. As the guanidine group would prolong the chain a second compound was synthesised reducing the chain by one carbon atom resulting in molecule **153**. Both compounds were synthesised as racemic mixtures. Compound **154** was a synthesis intermediate that would also contribute to a more complete structure activity relationship study therefore it was tested as well.

Table 79: Guanidine analogues of fluoxetine

Nr	Structure	EC ₅₀ (μM)	CC ₅₀ (μM)	SI
114	Fluoxetine (racemic)	3.2 ± 0.95	29.32 ± 0.35	9.1
146	(S)-fluoxetine	0.4 ± 0.15	28.63 ± 1.40	71.6
152		0.41 ± 0.27	>50	>121.9
153		1.22 ± 0.15	>50	>40.9
154	 *HCl	4.20 ± 0.93	32.26 ± 0.10	7.7

The combination of these two known 2C inhibitors resulted in the active compounds with a 5-10-fold improved activity compared to the racemic mixture of fluoxetine. It is of note that these compounds were only synthesised as racemic mixtures due to time constrains. Separation of the enantiomers or enantiomeric pure synthesis would be interesting for these compounds as a stereo selective effect similar to fluoxetine could also be expected in this case.

3.8.3 *N*-Benzyl-*N*-phenylfuran-2-carboxamide compounds their analogues

The publication of Zuo et al. in 2016 described a series of new enterovirus inhibitors that mainly targeted the 2C protein. While looking for fluoxetine analogues that would not possess a stereo centre thus avoiding enantiomerically pure synthesis or purification of chiral compounds, compound 2 (**115**) of this publication seemed an interesting candidate for further development. The compound (Figure 82) possesses a *N*-benzyl-*N*-phenylfuran-2-carboxamide backbone and induced mutations in the same positions as fluoxetine.

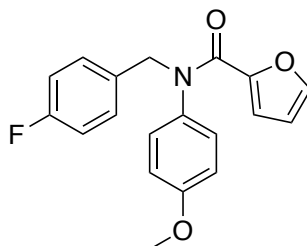
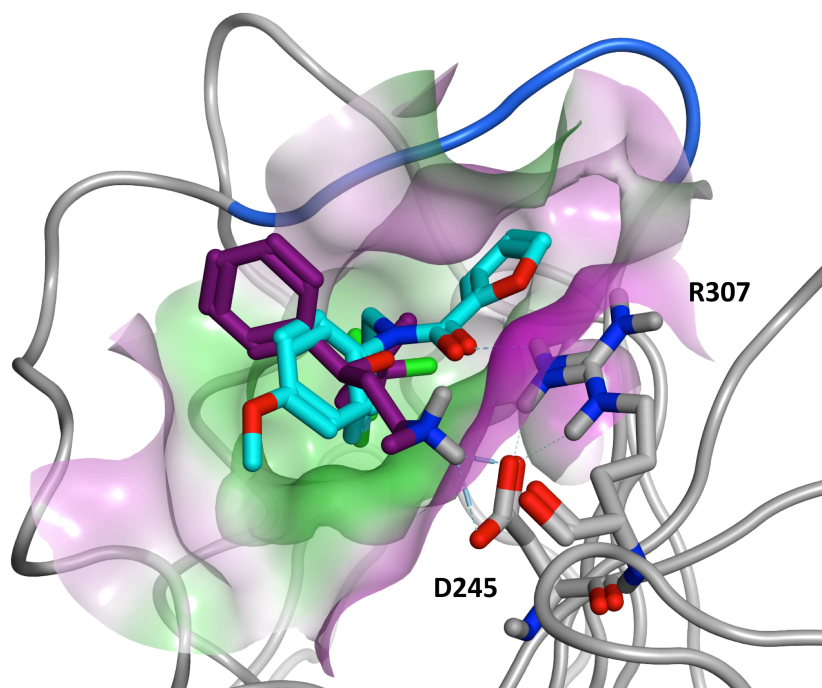


Figure 82: Compound 115
***N*-(4-Fluorobenzyl)-*N*-(4-methoxyphenyl)furan-2-carboxamide**

At this stage of the project the mode-of-action studies on fluoxetine were in their molecular dynamics phase and site A seemed a very promising binding pocket. Therefore, the compound was docked into site A. The last frame of the MD simulation was extracted and saved as .pdb file. The TIP3P solvent water was deleted from the file and the protein with (*S*)-fluoxetine as ligand was saved. This complex was subjected to the protein preparation wizard in Maestro and a grid file for docking was created using (*S*)-fluoxetine as the reference ligand for the positioning of the grid box and the size. Compound **115** was saved in .sdf format from ChemDraw and subjected to the ligprep protocol in Maestro. Up to 30 conformations were generated using the default settings. Then the compound was docked with Glide in standard precision mode. The resulting poses were exported and investigated with MOE.

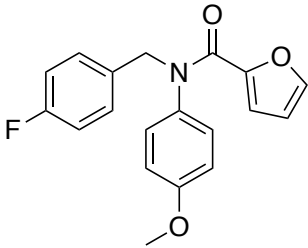
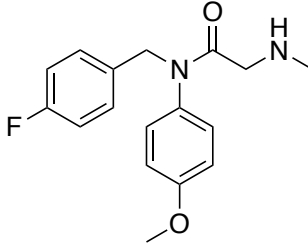
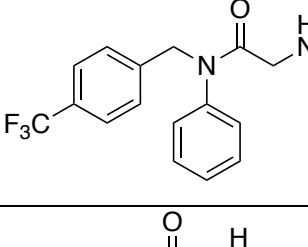
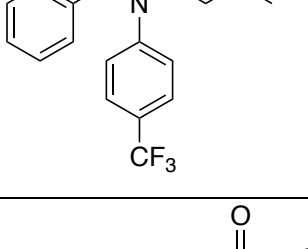
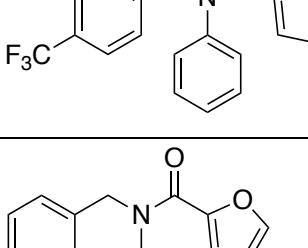
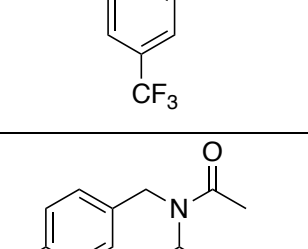
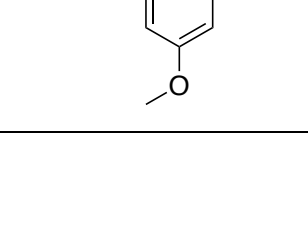


**Figure 83: 115 (turquoise) docked into site A after MD2 frame 625
(S)-fluoxetine as reference in purple**

In Figure 83 it can be observed that the trifluoromethyl moiety and the fluoro group of **115** occupy the same space in the pocket. Furthermore, the oxygen of fluoxetine and the nitrogen of the amide group of **115** nearly overlap. The carbonyl group of **115** makes a hydrogen bond with a nitrogen of the guanidine group of the arginine 307. The aromatic ring with the methoxy group resides well centred in the hydrophobic region of the pocket opening with the methoxy group pointing towards a more hydrophilic region. The furan ring occupies a small pocket that is not accessible to fluoxetine.

The structure of **115** fulfils the requirement of an achiral backbone and the substituents resemble those found on fluoxetine. Furthermore, it fits perfectly into the identified site A. Therefore, a classic medicinal chemistry approach to structure activity relationship studies was chosen. Features of fluoxetine and **115** were combined and typical replacements especially for the furan ring were made. Roberto Manganaro and his Erasmus Moira Lorenzo Lopez synthesised most of the compounds. Lisa Bauer carried out the cell-based virological evaluation. In the following table the structures of the analogues of **115** and the **fluoxetine-115** mix compounds.

Table 80: First round of compound 115 analogues synthesised and tested

Nr	Structure	EC ₅₀ (μM)	CC ₅₀ (μM)	SI
115		1.12 ± 0.28	>50	>44.6
155		NA	>50	-
156		NA	>20	-
157		NA	>50	-
158		0.84 ± 0.03	>50	>50
159		2.42 ± 0.42	>50	>20
160		NA	>50	-

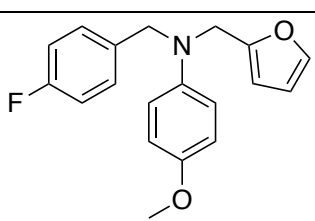
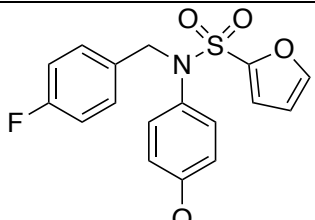
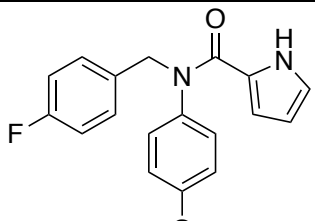
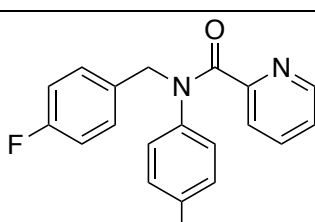
161		NA	>50	-
162		NA	>50	-
163		1.04 ± 0.16	>50	>16
164		6.18 ± 0.12	>50	>7.8

Table 81: Evaluation of compound 115 analogues on a panel of entero- and rhinoviruses (unit μM)

	EVA71 BrCr	CVB3 Nancy	Polio 1 Sabin	EVD68 Fermon	HRV-2	HRV-14	CC ₅₀
146	-	0.50 ± 0.14	-	0.67 ± 0.22	8.99 ± 0.24	6.33 ± 1.02	>21
152	-	1.76 ± 1.06	-	0.34 ± 0.05	-	2.11 ± 0.34	>50
153	-	1.22 ± 0.1	-	0.55 ± 0.04	-	-	>50
115	-	1.01 ± 0.2	-	0.31 ± 0.06		10.85 ± 1.86	>50
158	0.39 ± 0.05	0.836 ± 0.03	21 ± 0.85	1.395 ± 0.2	15.15 ± 1.13	4.76 ± 2.8	>50
159	-	2.42 ± 0.42		3.16 ± 0.12	-	-	>50
163	-	1.04 ± 0.16	-	1.39 ± 0.42	-	-	>50
164	-	6.18 ± 0.12	-	11.82 ± 0.56	-	-	>50

Among the compounds that were synthesised and tested based on compound 2 from Zuo et al. 2016 we identified several combinations of features between fluoxetine and **115** that

improved the activity towards CVB3 and other enteroviruses. Therefore, this scaffold can be used to design pan-enterovirus inhibitors with improved activity. A more extensive structure-activity-relationship study is still conducted but some of the compounds were not yet synthesised and tested. Therefore, it is too early to conclude which features are contributing in which way to the activity of the compounds. So far it can only be concluded that the five-membered ring cannot be substituted by an amino-chain as it was found in fluoxetine. The substitutions on the two aromatic rings seem to favour a fluoro or trifluoromethyl group in the benzene ring. The other ring should either remain unsubstituted or contain a methoxy-group in *para* position. Other substituents were not tested so far.

This small SAR study will be the basis of a more extended synthetic exploration of the backbone of **115**. With the new insights on the co-crystallised (S)-fluoxetine and its pocket, structure-based investigations and improvements of the **115** analogues that possibly also target the newly identified site might help design new 2C inhibitors with an even better activity against CVB3. With the use of homology models the pan-inhibitor activity might be improved as well.

3.9 Conclusions

From the beginning of this study, that was initiated with the release of the crystal structure of EV-A71 2C protein, to this point, several aspects of the protein were explored and the collaborative investigations between Lisa Bauer, Roberto Manganaro and Bruno Coutard led to several advances in the understanding of the 2C protein and its inhibitors.

First, the supposed binding site of the C-terminal helix was targeted with a HTVS approach but the efforts did not yield any active compounds. This might be due to the fact that the interface between the two oligomers of 2C does not rely only on the C-terminal helix to bind into its pocket but that several other factors are crucial for oligomerisation so that further studies might be needed. Perhaps, the interface, although well characterised by Guan et al. does not possess the expected biological relevance.

Then structures of higher order were investigated to understand the interactions between the monomers if the 2C protein indeed forms hexameric ring-shaped structures as it is commonly believed. A hexameric model was obtained with SymmDock and subjected to molecular dynamics simulations. One of the snapshots during the simulation showed that the ATP binding site incorporated a magnesium ion close to D176 and D177 that are expected to bind divalent cations in physiological conditions. Furthermore, it could be demonstrated that under hexameric conditions the R240 and R241 are in close proximity to ATP, sometimes even engaging in hydrogen bonds between one another. These results were used by Roberto

Manganaro to target the ATP binding site using an ensemble docking approach on different frames obtained from the molecular dynamics simulations. One active hit compound was obtained from his efforts and might be developed further in the future.

Using the information available for previously identified 2C inhibitors, it was aimed to identify potential pockets on the structure that could be the target sites for these known inhibitors. In depth investigations using computational modelling (docking and molecular dynamics simulations) led to the identification of two potential target sites.

The attention was then shifted to the characterisation of the mode-of-action of fluoxetine that is marketed as racemic mixture and for which the antiviral activity of the two enantiomers was not yet evaluated separately in such depth. It was identified through virological characterisation, binding assays on the purified 2C protein and mutations studies that only (S)-fluoxetine is responsible for the antiviral effect of the compound. Furthermore, Bruno Coutard was able to obtain crystals of 2C protein with (S)-fluoxetine crystallised within a new and previously uncharacterised pocket.

In parallel new compounds with improved antiviral activity against CVB3 and even broad spectrum activity towards other enteroviruses were designed and synthesised based on the results of the fluoxetine studies and the known 2C inhibitor **115** from Zuo et al. These studies have not reached the point of completion yet, and will be further explored during the next months.

In conclusion, the 2C protein presents a very interesting target for drug discovery against enteroviruses and rhinoviruses. With this studies we contributed to the elucidation of the mode-of-action of known 2C inhibitors and we identified several new compounds with antiviral activity against 2C. This work also initiated and facilitated studies towards the resolution of a new crystal structure of the 2C protein of CVB3 together with the 2C inhibitor (S)-fluoxetine.

3.10 Experimental

3.10.1 Computational studies

A PC equipped with a 1.80 GHz Intel Xeon CPU (8 cores), 32GB of RAM and a GPU Zotac GeForce GTX 1080-Ti Mini with 11GB GDDR5X of dedicated RAM, running Ubuntu 14.04 LTS was used for molecular modelling studies. The methods used during this work are described within the corresponding sections.

3.10.2 Virological assays

3.10.2.1 Cell Culture

Buffalo Green Monkey cells (BGM) and HeLa R19 cells were cultured in Dulbecco's modified Eagle's medium (DMEM, Lonza) supplemented with 10 % fetal bovine serum (FBS). Huh7-Lunet 7/T7, a stable cell pool expressing T7 RNA polymerase and blasticidin S-deaminase (Backes et al. 2010), were cultured in DMEM supplemented with 10% FBS and 10 µg/ml blasticidin. All cell lines were grown at 37°C in 5 % CO₂.

3.10.2.2 Viruses

CVB3 and CVB3 mutant viruses were obtained by transfecting BGM cells with RNA transcripts derived from the full-length infectious clones p53CB3/T7 ((Wessels et al. 2006) and 2C [C179Y], 2C [A224V], 2C [I227V], 2C [A229V], 2C [A224V/I227V], 2C [A224V/A229V], 2C [I227V/A229V], 2C [A224V/I227V/A229V] (De Palma et al. 2008)). Rluc-CVB3, which contains a Renilla luciferase gene upstream of the capsid coding region, was previously described (Lanke et al. 2009). Rluc-CVB3, was obtained by transfecting Huh7-Lunet/T7 (Backes et al. 2010) cells with MluI-linearized pRluc-53CB3/T7 plasmid.

3.10.2.3 Single-cycle virus Infection

Virus infections were performed by incubating subconfluent HeLa R19 cells with virus corresponding multiplicity of infection (MOI) at 37°C for 30 min. Next, the medium was removed and fresh (compound-containing) medium was added to the cells. After the indicated time points, in case of Rluc-CVB3, infection, the medium was discarded and cells were lysed to determine the Renilla luciferase activity using the Renilla luciferase Assay System (Promega) according to the manufacturer's protocol. Where indicated, cell viability was determined in parallel using the AQueous One Solution Cell Proliferation Assay (Promega) according to the manufacturer's protocol. The optical density at 490 nm was determined using a microplate reader. For measurements of infectious particles, virus was released from the cells by three freeze-thawing cycles. Virus titers were determined by end-point dilution assay and calculated by the method of Reed and Muench (Reed & Muench 1938).

3.10.2.4 Multicycle CPE reduction Assay

Subconfluent layers of HeLa R19 cells were seeded in 96-wells and treated with serial dilutions of the corresponding compounds. Cells were infected with CVB3 at the lowest possible MOI (MOI 0,001) resulting in full CPE within 2-3 day. Subsequently the cells were incubated at 37°C for 2-3 days until full CPE was observed in the infected untreated cell controls. Cell viability was determined in parallel using the AQueous One Solution Cell Proliferation Assay (Promega) according to the manufacturer's protocol. The optical density at 490 nm was determined using

a microplate reader. Raw OD values were converted to percentage of untreated and uninfected cell control after subtraction of the background.

3.10.2.5 Calculations

The concentration of compound that inhibits virus-induced cell death by 50% (50% effective concentration [EC₅₀]) was calculated by nonlinear regression analysis. Cytotoxicity of the compounds was assessed in a similar set-up, and 50% cytotoxic concentration (CC₅₀) values were derived from cell viability values determined with an MTS assay. Each experiment was performed at least in triplicate. The nonlinear regression and the graphs were done with GraphPad Prism Version 6.

4 General Conclusions

In the general aims of this work the constant need for antiviral research and drug discovery was outlined and the contributions to this process that can be made with computational and medicinal chemistry approaches were discussed.

Several molecular modelling techniques were applied to the two main parts of this thesis: the search for novel anti-chikungunya virus compounds and the search for novel anti-enterovirus compounds. The results for the individual chapters were already presented in the appropriate section. Here only a brief general overview will be given.

In order to target chikungunya virus a structure-based pharmacophore and docking approach was applied to identify hit compounds that showed activity in cell-based antiviral assays. One compound seemed particularly promising, thus a series of analogues was synthesised to understand the contributions of the different substituents on the molecule. While these novel compounds were evaluated for their antiviral activity, in depth studies to clarify the mechanism of action of the hit compounds were undertaken. The compound was designed to target CHIKV macro domain, but biochemical assays investigating binding to the purified protein could not confirm a direct interaction. Therefore, other possible modes of action were considered. So far it seems that the compound plays a role as entry inhibitor but it cannot be excluded that it also inhibits viral replication. Unfortunately, neither the physicochemical properties nor the activity of any of the analogues could be significantly improved, therefore the process in raising resistance mutations, which sometimes leads to the identification of a target, was not successful. In a second screen against the distal ribose pocket of CHIKV macro domain one active compound could be identified but due to time constraints its mode of action could not be further analysed.

Homology modelling was applied to contribute to the structural understanding of the alphavirus unique domain of CHIKV. This also led to the successful linkage of the crystallised component of the P23 precursor of CHIKV from the protease domain to the alphavirus unique domain. This P23 precursor was then targeted using a HTVS approach, but the purchased compounds resulted to be inactive.

In the enterovirus part of the project that was initiated as a collaboration between Lisa Bauer, Roberto Manganaro and myself we aimed to target first the recently crystallised enterovirus 2C protein with computational methods in order to find new antiviral against enteroviruses. Enteroviruses are a large group of related viruses and the 2C protein is a crucial non-structural protein in the lifecycle of these pathogens. Due to sequence similarity homology models were created that could be investigated for the discovery of a potential pan-enterovirus inhibitor. Furthermore, several compounds were previously identified in cell-based screenings to target

the 2C protein of enteroviruses. The mode of action of some of these compounds was investigated and a mode of action of Fluoxetine could be hypothesised. Based on docking studies and virological data new antiviral compounds were synthesised with improved activity and a broader spectrum against different enteroviruses. The discovery of the stereoselective effect of fluoxetine as enterovirus inhibitor has also contributed to the resolution of a crystal structure with the compound co-crystallised.

In conclusion, this thesis has employed computational methods to contribute to a better understanding of CHIKV and enterovirus non-structural proteins and has identified several novel antiviral compounds that might be further developed in the future.

5 References

- Abdel-Magid, A.F. et al., 1996. Reductive Amination of Aldehydes and Ketones with Sodium Triacetoxyborohydride. Studies on Direct and Indirect Reductive Amination Procedures ¹. *The Journal of Organic Chemistry*, 61(11), pp.3849–3862.
- Abdelnabi, R. et al., 2017. Protein kinases C as potential host targets for the inhibition of chikungunya virus replication. *Antiviral Research*, 139, pp.79–87.
- Abdelnabi, R., Neyts, J. & Delang, L., 2015. Towards antivirals against chikungunya virus. *Antiviral Research*, 121, pp.59–68.
- Aggarwal, M. et al., 2015. Kinetic characterization of trans-proteolytic activity of Chikungunya virus capsid protease and development of a FRET-based HTS assay. *Scientific Reports*, 5(14753), pp.1–12.
- Albulescu, L. et al., 2015. Broad-range inhibition of enterovirus replication by OSW-1, a natural compound targeting OSBP. *Antiviral Research*, 117, pp.110–114.
- Albulescu, L. et al., 2017. Uncovering oxysterol-binding protein (OSBP) as a target of the anti-enteroviral compound TTP-8307. *Antiviral Research*, 140, pp.37–44.
- Allen, M.D. et al., 2003. The crystal structure of AF1521 a protein from *Archaeoglobus fulgidus* with homology to the non-histone domain of macroH2A. *Journal of Molecular Biology*, 330(3), pp.503–511.
- Amombo, G.M.O. et al., 2012. Modification of a promiscuous inhibitor shifts the inhibition from γ -secretase to FLT-3. *Bioorganic and Medicinal Chemistry Letters*, 22(24), pp.7634–7640.
- Andries, K. et al., 1992. In vitro activity of pirodavir (R 77975), a substituted phenoxy-pyridazinamine with broad-spectrum antipicornaviral activity. *Antimicrobial Agents and Chemotherapy*, 36(1), pp.100–107.
- Ao, D., Sun, S.Q. & Guo, H.C., 2014. Topology and biological function of enterovirus non-structural protein 2B as a member of the viroporin family. *Veterinary Research*, 45(1), pp.1–9.
- Arita, M. et al., 2013. Oxysterol-Binding Protein Family I Is the Target of Minor Enviroxime-Like Compounds. *Journal of Virology*, 87(8), pp.4252–4260.
- Asare, E. et al., 2016. A Single Amino Acid Substitution in Poliovirus Nonstructural Protein 2C ATPase Causes Conditional Defects in Encapsidation and Uncoating. *Journal of Virology*, 90(14), pp.6174–6186.
- Ashburn, T.T. & Thor, K.B., 2004. Drug repositioning: identifying and developing new uses for existing drugs. *Nature Reviews Drug Discovery*, 3(8), pp.673–683.
- Bachmair, A., Finley, D. & Varshavsky, A., 1986. In vivo half-life of a protein is a function of its amino-terminal residue. *Science*, 234(4773), pp.179–186.
- Backes, P. et al., 2010. Role of Annexin A2 in the Production of Infectious Hepatitis C Virus

- Particles. *Journal of Virology*, 84(11), pp.5775–5789.
- Badorff, C. et al., 1999. Enteroviral protease 2A cleaves dystrophin: Evidence of cytoskeletal disruption in an acquired cardiomyopathy. *Nature Medicine*, 5(3), pp.320–326.
- Baggen, J. et al., 2018. The life cycle of non-polio enteroviruses and how to target it. *Nature Reviews Microbiology*, 16(June), pp.368–381.
- Baltimore, D., 1971. Expression of Animal Virus Genomes. *Bacteriological reviews*, 35(3), pp.235–241.
- Barkauskaite, E., Jankevicius, G. & Ahel, I., 2015. Structures and Mechanisms of Enzymes Employed in the Synthesis and Degradation of PARP-Dependent Protein ADP-Ribosylation. *Molecular Cell*, 58(6), pp.935–946.
- Baron, S., Fons, M. & Albrecht, T., 1996. *Viral Pathogenesis*, University of Texas Medical Branch at Galveston.
- Barton, D.J. & Flanagan, J.B., 1997. Synchronous replication of poliovirus RNA: initiation of negative-strand RNA synthesis requires the guanidine-inhibited activity of protein 2C. *Journal of virology*, 71(11), pp.8482–9.
- Bassetto, M. et al., 2013. Computer-aided identification, design and synthesis of a novel series of compounds with selective antiviral activity against chikungunya virus. *Antiviral Research*, 98(1), pp.12–18.
- Bauer, L. et al., 2017. Direct-acting antivirals and host-targeting strategies to combat enterovirus infections. *Current Opinion in Virology*, 24, pp.1–8.
- Berman, H.M. et al., 2000. The Protein Data Bank Helen. *Nucleic acids research*, 28(1), pp.235–242.
- Bernstein, F. et al., 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *European Journal of Biochemistry*, 80(1), pp.319–324.
- Bettadapura, J. et al., 2013. Approaches to the treatment of disease induced by chikungunya virus. *The Indian journal of medical research*, 138(5), pp.762–765.
- BioSolveIT GmbH, 2006. LeadIT version 2.1.8.
- Blaising, J., Polyak, S.J. & Pécheur, E.I., 2014. Arbidol as a broad-spectrum antiviral: An update. *Antiviral Research*, 107(1), pp.84–94.
- Bonderoff, J.M., LaRey, J.L. & Lloyd, R.E., 2008. Cleavage of Poly(A)-Binding Protein by Poliovirus 3C Proteinase Inhibits Viral Internal Ribosome Entry Site-Mediated Translation. *Journal of Virology*, 82(19), pp.9389–9399.
- Bos, L., 1999. Beijerinck's work on tobacco mosaic virus: Historical context and legacy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 354(1383), pp.675–685.
- Bourjot, M. et al., 2012. Prostratin and 12-O-Tetradecanoylphorbol 13-Acetate Are Potent and

- Selective Inhibitors of Chikungunya Virus Replication. *Journal of Natural Products*, 75(12), pp.2183–2187.
- Bowers, K.J. et al., 2006. Scalable algorithms for molecular dynamics simulations on commodity clusters. In *Proceedings of the 2006 ACM/IEEE conference on Supercomputing - SC '06*. p. 84.
- Case, D. et al., 2012. *AMBER 12, University of California, San Francisco.*,
- Cavasotto, C.N. & Phatak, S.S., 2009. Homology modeling in drug discovery : current trends and applications. *Drug Discovery Today*, 14(July), pp.676–683.
- CDC, C. of D.C., 2016a. Clinical Evaluation & Disease | Chikungunya virus | CDC. Available at: <http://www.cdc.gov/chikungunya/hc/clinicalevaluation.html> [Accessed May 13, 2016].
- CDC, C. of D.C., 2019. Non-polio enterovirus. <https://www.cdc.gov/non-polio-enterovirus/about/transmission.html>.
- CDC, C. of D.C., 2016b. Symptoms, Diagnosis, & Treatment | Chikungunya virus | CDC. Available at: <http://www.cdc.gov/chikungunya/symptoms/index.html> [Accessed May 13, 2016].
- Charlotte, E. et al., 2017. Escaping Host Factor PI4KB Inhibition : Enterovirus. , 21(3).
- Chemical Computing Group Inc., 2016. Molecular Operating Environment (MOE).
- Chemical Computing Group Inc., 2014. Site Finder. *MOE Tutorial*.
- Chen, D. et al., 2011. Identification of macrodomain proteins as novel O-acetyl-ADP-ribose deacetylases. *Journal of Biological Chemistry*, 286(15), pp.13261–13271.
- Cho, M.W. et al., 1994. Membrane rearrangement and vesicle induction by recombinant poliovirus 2C and 2BC in human cells. *Virology*, 202(1), pp.129–145.
- De Clercq, E. & Li, G., 2016. Approved Antiviral Drugs over the Past 50 Years. *Clinical Microbiology Reviews*, 29(3), pp.695–747.
- ClinicalTrials.gov ID:NCT00391313, 2006. CuraChik : A Trial of the Efficacy and Safety of Chloroquine as Therapeutic Treatment of Chikungunya Disease. <https://clinicaltrials.gov/ct2/show/NCT00391313>. Available at: <https://clinicaltrials.gov/ct2/show/NCT00391313> [Accessed June 5, 2019].
- ClinicalTrials.gov ID:NCT01489358, 2016. VRC 311: A Phase 1 Open Label, Dose-Escalation Clinical Trial to Evaluate the Safety and Immunogenicity of a Virus-Like Particle (VLP) Chikungunya Vaccine, VRC-CHKVLP059-00-VP, in Healthy Adults. *NCT01489358*. Available at: <https://clinicaltrials.gov/ct2/show/study/NCT01489358> [Accessed May 8, 2019].
- ClinicalTrials.gov ID:NCT02562482, 2015. Phase 2 Randomized, Placebo-Controlled Trial to Evaluate the Safety and Immunogenicity of a Chikungunya Virus-Like Particle Vaccine, VRC-CHKVLP059-00-VP, in Healthy Adults. *NCT02562482*. Available at: <https://clinicaltrials.gov/ct2/show/NCT02562482> [Accessed May 8, 2019].

- ClinicalTrials.gov ID:NCT03483961, 2018. A Phase 2 Parallel-Group, Randomized, Double-Blind Study to Assess the Safety and Immunogenicity of PXVX0317 (Chikungunya Virus Virus-Like Particle Vaccine [CHIKV-VLP], Unadjuvanted or Alum-adjuvanted). *NCT03483961*. Available at: <https://clinicaltrials.gov/ct2/show/NCT03483961> [Accessed May 8, 2019].
- Collett, M.S. et al., 2017. Antiviral activity of pocapavir in a randomized, blinded, placebo-controlled human oral poliovirus vaccine challenge model. *Journal of Infectious Diseases*, 215(3), pp.335–343.
- Coulson, C.A., 1956. *Valence*, Clarendon Press.
- Coutard, B. et al., 2014. Assessment of Dengue virus helicase and methyltransferase as targets for fragment-based drug discovery. *Antiviral Research*, 106(1), pp.61–70.
- Cruz, D.J.M. et al., 2013. Identification of Novel Compounds Inhibiting Chikungunya Virus-Induced Cell Death by High Throughput Screening of a Kinase Inhibitor Library. *PLoS Neglected Tropical Diseases*, 7(10).
- Cui, M. et al., 2011. Molecular Docking: A powerful approach for structure-based drug discovery. *Current Computer-Aided Drug Design*, 7(2), pp.146–157.
- Das, P.K. et al., 2016. Design and Validation of Novel Chikungunya Virus Protease Inhibitors. *Antimicrobial agents and chemotherapy*, 60(12), pp.7382–7395.
- Delang, L. et al., 2014. Mutations in the chikungunya virus non-structural proteins cause resistance to favipiravir (T-705), a broad-spectrum antiviral. *The Journal of antimicrobial chemotherapy*, 69(10), pp.2770–84.
- Delogu, I. et al., 2011. In vitro antiviral activity of arbidol against Chikungunya virus and characteristics of a selected resistant mutant. *Antiviral Research*, 90(3), pp.99–107.
- Diallo, M. et al., 1999. Vectors of Chikungunya virus in Senegal: current data and transmission cycles. *The American journal of tropical medicine and hygiene*, 60(2), pp.281–6.
- Diamond, M.S., Zachariah, M. & Harris, E., 2002. Mycophenolic acid inhibits dengue virus infection by preventing replication of viral RNA. *Virology*, 304(2), pp.211–221.
- DrugBank, 2019. DrugBank - Chloroquine. Available at: <https://www.drugbank.ca/drugs/DB00608> [Accessed May 9, 2019].
- van Duijl-Richter, M.K.S. et al., 2015. Early events in chikungunya virus infection—from virus cell binding to membrane fusion. *Viruses*, 7(7), pp.3647–3674.
- Eckej, L. et al., 2017. The conserved macrodomains of the non-structural proteins of Chikungunya virus and other pathogenic positive strand RNA viruses function as mono-ADP-ribosylhydrolases. *Scientific Reports*, 7(February), p.41746.
- Edelman, R. et al., 2000. Phase II safety and immunogenicity study of live chikungunya virus vaccine TSI-GSD-218. *The American Journal of Tropical Medicine and Hygiene*, 62(6), pp.681–685.

- Eggers, H.J., 2004. Successful treatment of enterovirus-infected mice by 2-(alpha-hydroxybenzyl)-benzimidazole and guanidine. *Journal of Experimental Medicine*, 143(6), pp.1367–1381.
- Eggers, H.J. & Tamm, I., 1962. On the mechanism of selective inhibition of enterovirus multiplication by 2-(α -hydroxybenzyl)-benzimidazole. *Virology*, 18(3), pp.426–438.
- Eggers, H.J. & Tamm, I., 1961. SPECTRUM AND CHARACTERISTICS OF THE VIRUS INHIBITORY ACTION OF 2-(alpha-HYDROXYBENZYL)-BENZIMIDAZOLE. *Journal of Experimental Medicine*, 113(4), pp.657–682.
- Egloff, M.-P. et al., 2006. Structural and functional basis for ADP-ribose and poly(ADP-ribose) binding by viral macro domains. *Journal of virology*, 80(17), pp.8493–502.
- Ehrlich, P., 1909. Über den jetzigen Stand der Chemotherapie. *Berichte der deutschen chemischen Gesellschaft*, 42(1), pp.17–47.
- Enemark, E.J. & Joshua-Tor, L., 2006. Mechanism of DNA translocation in a replicative hexameric helicase. *Nature*, 442(7100), pp.270–275.
- Eriksson, K.K. et al., 2008. Mouse Hepatitis Virus Liver Pathology Is Dependent on ADP-Ribose-1''-Phosphatase, a Viral Function Conserved in the Alpha-Like Supergroup. *Journal of Virology*, 82(24), pp.12325–12334.
- FDA U.S. Food & Drug Administration, 2016. Drug Innovation - Novel Drug Approvals for 2016. Available at: <https://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugInnovation/ucm483775.htm> [Accessed November 22, 2018].
- FDA U.S. Food & Drug Administration, 2017. Drug Innovation - Novel Drug Approvals for 2017. Available at: <https://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugInnovation/ucm537040.htm> [Accessed November 22, 2018].
- FDA U.S. Food & Drug Administration, 2018. Drug Innovation - Novel Drug Approvals for 2018. Available at: <https://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugInnovation/ucm592464.htm> [Accessed November 22, 2018].
- Fehr, A.R. et al., 2016. The Conserved Coronavirus Macrodomain Promotes Virulence and Suppresses the Innate Immune Response during Severe Acute Respiratory Syndrome Coronavirus Infection. *mBio*, 7(6), pp.1–12.
- Fehr, A.R. et al., 2014. The nsp3 Macrodomain Promotes Virulence in Mice with Coronavirus-Induced Encephalitis. *Journal of Virology*, 89(3), pp.1523–1536.
- Fehr, A.R. et al., 2018. Viral Macrodomains: Unique Mediators of Viral Replication and Pathogenesis. *Trends in Microbiology*, 26(7), pp.598–610.

- Feibelman, K.M. et al., 2018. Identification of small molecule inhibitors of the Chikungunya virus nsP1 RNA capping enzyme. *Antiviral research*, 154, pp.124–131.
- Feijs, K.L.H. et al., 2013. Macrodomein-containing proteins: regulating new intracellular functions of mono(ADP-ribosyl)ation. *Nature reviews. Molecular cell biology*, 14(7), pp.443–51.
- Ferrer-Orta, C., Ferrero, D. & Verdaguer, N., 2015. RNA-dependent RNA polymerases of picornaviruses: From the structure to regulatory mechanisms. *Viruses*, 7(8), pp.4438–4460.
- Fields, W. & Kielian, M., 2015. Interactions involved in pH protection of the alphavirus fusion protein. *Virology*, 486, pp.173–179.
- Fleming, F.F. et al., 2011. Nitrile-Containing Pharmaceuticals: Efficacious Roles of the Nitrile Pharmacophore. *Journal of Medicinal Chemistry*, 53(22), pp.7902–7917.
- Flint, S. et al., 2009. *Principles of Virology: Molecular Biology*,
- Forst, A.H. et al., 2013. Recognition of mono-ADP-ribosylated ARTD10 substrates by ARTD8 macrodomains. *Structure*, 21(3), pp.462–475.
- Franco, E.J. et al., 2018. The effectiveness of antiviral agents with broad-spectrum activity against chikungunya virus varies between host cell lines. *Antiviral Chemistry and Chemotherapy*, 26, pp.1–7.
- Friesner, R.A. et al., 2004. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of Medicinal Chemistry*, 47(7), pp.1739–49.
- Fros, J.J. et al., 2013. The C-terminal domain of chikungunya virus nsP2 independently governs viral RNA replication, cytopathicity, and inhibition of interferon signaling. *Journal of virology*, 87(18), pp.10394–400.
- Furuta, Y. et al., 2013. Favipiravir (T-705), a novel viral RNA polymerase inhibitor. *Antiviral research*, 100(2), pp.446–54.
- Gazina, E. V. et al., 2005. Ion transport blockers inhibit human rhinovirus 2 release. *Antiviral Research*, 67(2), pp.98–106.
- Gerber, P.R. & Müller, K., 1995. MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry. *Journal of Computer-Aided Molecular Design*, 9(3), pp.251–268.
- Gofshateyn, J., Cárdenas, A.M. & Bearden, D., 2016. Treatment of Chronic Enterovirus Encephalitis With Fluoxetine in a Patient With X-Linked Agammaglobulinemia. *Pediatric Neurology*, 64, pp.94–98.
- Gorbalenya, A.E., Koonin, E. V & Wolf, Y.I., 1990. NTP -binding domains encoded by genomes of small DNA and RNA viruses. *FEBS Letters*, 262(1), pp.145–148.

- Gorchakov, R. et al., 2012. Attenuation of Chikungunya Virus Vaccine Strain 181/Clone 25 Is Determined by Two Amino Acid Substitutions in the E2 Envelope Glycoprotein. *Journal of Virology*, 86(11), pp.6084–6096.
- Gould, E.A. et al., 2010. Understanding the alphaviruses: Recent research on important emerging pathogens and progress towards their control. *Antiviral Research*, 87(2), pp.111–124.
- Gould, E.A. & Higgs, S., 2009. Impact of climate change and other factors on emerging arbovirus diseases. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 103(2), pp.109–21.
- Goulding, J., 2016. Virus Replication. <https://www.immunology.org/public-information/bitesized-immunology/patógenos-y-enfermedades/virus-replication>.
- Goyal, M. et al., 2018. Recent development in the strategies projected for chikungunya vaccine in humans. *Drug Design, Development and Therapy*, Volume 12, pp.4195–4206.
- GPEI, G.P.E.I., 2017. *Annual report 2017*,
- GPEI, G.P.E.I., 2018. Polio-EB144/9 in English. , EB144/9(November 2018), pp.1–7.
- de Groot, R.J. et al., 1990. Cleavage-site preferences of Sindbis virus polyproteins containing the non-structural proteinase. Evidence for temporal regulation of polyprotein processing in vivo. *The EMBO journal*, 9(8), pp.2631–8.
- Guan, H. et al., 2017. Crystal structure of 2C helicase from enterovirus 71. *Science Advances*, 3(4), pp.1–10.
- Guan, H. et al., 2018. Crystal structure of a soluble fragment of poliovirus 2CATPase. *PLoS Pathogens*, 14(9), pp.1–24.
- Gutierrez, C.D., Bavetsias, V. & McDonald, E., 2005. TiCl(OiPr)₃ and NaBH(OAc)₃: An efficient reagent combination for the reductive amination of aldehydes by electron-deficient amines. *Tetrahedron Letters*, 46(20), pp.3595–3597.
- Haguenau, F. et al., 2003. Key Events in the History of Electron Microscopy. *Microscopy and Microanalysis*, 9, pp.96–138.
- Harder, E. et al., 2016. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *Journal of Chemical Theory and Computation*, 12(1), pp.281–296.
- Harrison, V.R. et al., 1971. Production and evaluation of a formalin-killed Chikungunya vaccine. *Journal of immunology (Baltimore, Md. : 1950)*, 107(3), pp.643–7.
- Hayden, F.G. et al., 2003. Phase II, Randomized, Double-Blind, Placebo-Controlled Studies of Rupintrivir Nasal Spray 2-Percent Suspension for Prevention and Treatment of Experimentally Induced Rhinovirus Colds in Healthy Volunteers. *Antimicrobial Agents and Chemotherapy*, 47(12), pp.3907–3916.
- Her, Z. et al., 2009. Chikungunya: a bending reality. *Microbes and Infection*, 11(14–15),

pp.1165–1176.

- Hinchliffe, A., 2003. *Molecular Modelling for Beginners*, John Wiley & Sons, Ltd.
- Hoffmann, R., 1963. An Extended Hückel Theory. I. Hydrocarbons. *The journal of chemical physics*, 39(6), pp.1397–1412.
- Hoke, C.H. et al., 2012. US Military contributions to the global response to pandemic Chikungunya. *Vaccine*, 30(47), pp.6713–6720.
- Hoorweg, T.E. et al., 2016. Dynamics of chikungunya virus cell entry unraveled by single virus tracking in living cells. *Journal of Virology*, 90(9), pp.4745–4756.
- Hospital, A. et al., 2015. Molecular Dynamics Simulations: Advances and Applications. *Advances and Applications in Bioinformatics and Chemistry*, 8, pp.37–47.
- Hottiger, M.O. et al., 2010. Toward a unified nomenclature for mammalian ADP-ribosyltransferases. *Trends in Biochemical Sciences*, 35(4), pp.208–219.
- Jacobs, S.E. et al., 2013. Human rhinoviruses. *Clinical Microbiology Reviews*, 26(1), pp.135–162.
- Jankevicius, G. et al., 2013. A family of macrodomain proteins reverses cellular mono-ADP-ribosylation. *Nature structural & molecular biology*, 20(4), pp.508–14.
- Jin, J. et al., 2015. Neutralizing Monoclonal Antibodies Block Chikungunya Virus Entry and Release by Targeting an Epitope Critical to Viral Pathogenesis. *Cell Reports*, 13(11), pp.2553–2564.
- Jorgensen, W.L. et al., 1983. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2), pp.926–935.
- Jorgensen, W.L., Maxwell, D.S. & Tirado-Rives, J., 1996. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society*, 118(45), pp.11225–11236.
- Jose, J., Snyder, J.E. & Kuhn, R.J., 2009. A structural and functional perspective of alphavirus replication and assembly. *Future microbiology*, 4(7), pp.837–56.
- Karlberg, T. et al., 2013. Structural biology of the writers, readers, and erasers in mono- and poly(ADP-ribose) mediated signaling. *Molecular Aspects of Medicine*, 34(6), pp.1088–1108.
- Karpe, Y. a, Aher, P.P. & Lole, K.S., 2011. NTPase and 5'-RNA triphosphatase activities of Chikungunya virus nsP2 protein. *PloS one*, 6(7), p.e22336.
- Kaufman, H.E., 1962. Clinical cure of herpes simplex keratitis by 5-iodo-2-deoxyuridine. *Proceedings of the Society for Experimental Biology and Medicine*, 109, pp.251–2.
- Kaur, P. & Chu, J.J.H., 2013. Chikungunya virus: an update on antiviral development and challenges. *Drug discovery today*, 18(19–20), pp.969–83.
- Keen, E.C., 2015. A century of phage research: Bacteriophages and the shaping of modern biology. *Bioessays*, 37(1), pp.6–9.

- Klebe, G., 2006. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discovery Today*, 11(13–14), pp.580–594.
- Klebe, G., 2009. *Wirkstoffdesign: Entwurf und Wirkung von Arzneistoffen* 2.Auflage., Spektrum Akademischer Verlag.
- Klein, M. et al., 2000. Picornavirus replication inhibitors HBB and guanidine in the echovirus-9 system: The significance of viral protein 2C. *Journal of General Virology*, 81(4), pp.895–901.
- Kondekar, S. & Gogtay, N., 2006. Why Chikungunya is called Chikungunya. *Journal of Postgraduate Medicine*, 52(4), p.307.
- Koonin, E. V et al., 1992. Computer-assisted assignment of functional domains in the nonstructural polyprotein of hepatitis E virus: delineation of an additional group of positive-strand RNA plant and animal viruses. *Proceedings of the National Academy of Sciences USA*, 89(September), pp.8259–8263.
- Korb, O., Stutzle, T. & Exner, T., 2009. Empirical scoring functions for advanced protein– ligand docking with PLANTS. *Journal of chemical information and Modeling*, 49(1), pp.84–96.
- Krishna, S.S., Majumdar, I. & Grishin, N. V., 2003. Structural classification of zinc fingers. *Nucleic Acids Research*, 31(2), pp.532–550.
- Krissinel, E. & Henrick, K., 2007. Interference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology*, 372, pp.774–797.
- Kumaran, D., Eswaramoorthy, S. & Studier, F.W., 2005. Structure and mechanism of ADP-ribose-1''-monophosphate (Appr-1''-pase) a ubiquitous cellular processing enzyme. *Protein Science*, pp.719–726.
- Kuri, T. et al., 2011. The ADP-ribose-1''-monophosphatase domains of severe acute respiratory syndrome coronavirus and human coronavirus 229E mediate resistance to antiviral interferon responses. *Journal of General Virology*, 92(8), pp.1899–1905.
- Laakkonen, P. et al., 1998. Alphavirus Replicase Protein NSP1 Induces Filopodia and Rearrangement of Actin Filaments. *Journal of Virology*, 72(12), pp.10265–10269.
- Labute, P., 2008a. Protonate3D: Assignment of Ionization States and Hydrogen Coordinates to Macromolecular Structures. *Proteins*, 75, pp.187–205.
- Labute, P., 2008b. The generalized Born/volume integral implicit solvent model: Estimation of the free energy of hydration using London dispersion instead of atomic surface area. *Journal of Computational Chemistry*, 29(10), pp.1693–1698.
- de Lamballerie, X., Ninove, L. & Charrel, R., 2012. Antiviral Treatment of Chikungunya Virus Infection. *Infectious Disorders - Drug Targets*, 9(2), pp.101–104.
- Lanke, K.H.W. et al., 2009. GBF1, a Guanine Nucleotide Exchange Factor for Arf, Is Crucial for Coxsackievirus B3 RNA Replication. *Journal of Virology*, 83(22), pp.11940–11949.

- Law, Y.-S. et al., 2019. Structural insights into RNA recognition by the Chikungunya virus nsP2 helicase. *Proceedings of the National Academy of Sciences*, p.201900656.
- Leach, A.R., 2001. *Molecular Modelling - Principles and Applications* 2nd Editio., Pearson Education Limited.
- Lechevalier, H., 1972. Dmitri Iosifovich Ivanovski (1864-1920). *Bacteriological reviews*, 36(2), pp.135–45.
- Lefkowitz, E.J. et al., 2018. Virus taxonomy: The database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Research*, 46(D1), pp.D708–D717.
- Lei, X. et al., 2013. Cleavage of Interferon Regulatory Factor 7 by Enterovirus 71 3C Suppresses Cellular Responses. *Journal of Virology*, 87(3), pp.1690–1698.
- Lei, X., Xiao, X. & Wang, J., 2016. Innate immunity evasion by enteroviruses: Insights into virus-host interaction. *Viruses*, 8(1), pp.1–13.
- Lemm, J.A. et al., 1994. Polypeptide requirements for assembly of functional Sindbis virus replication complexes: a model for the temporal regulation of minus- and plus-strand RNA synthesis. *The EMBO journal*, 13(12), pp.2925–34.
- Leung, A.K.L., McPherson, R.L. & Griffin, D.E., 2018. Macrodomein ADP-ribosylhydrolase and the pathogenesis of infectious diseases. *PLOS Pathogens*, 14(3), p.e1006864.
- Levitt, N.H. et al., 1986. Development of an attenuated strain of chikungunya virus for use in vaccine production. *Vaccine*, 4(3), pp.157–162.
- Li, C. et al., 2015. mRNA Capping by Venezuelan Equine Encephalitis Virus nsP1: Functional Characterization and Implications for Antiviral Research. *Journal of Virology*, 89(16), pp.8292–303.
- Li, C. et al., 2016. Viral Macro Domains Reverse Protein ADP-ribosylation. *Journal of Virology*, 90(July), p.JVI.00705-16.
- Li, J.-P. & Baltimore, D., 1990. An Intragenic Revertant of a Poliovirus 2C Mutant Has an Uncoating Defect. *Journal of Virology*, 64(3), pp.1102–1107.
- van der Linden, L., Wolthers, K.C. & van Kuppeveld, F.J.M., 2015. Replication and inhibitors of enteroviruses and parechoviruses. *Viruses*, 7(8), pp.4529–4562.
- Lipinski, C.A. et al., 2012. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, 64, pp.4–17.
- Liu, Y. et al., 2010. Direct interaction between two viral proteins, the nonstructural protein 2CATPase and the capsid protein VP3, is required for enterovirus morphogenesis. *PLoS Pathogens*, 6(8), pp.75–76.
- Loschmidt, J., 1861. *Chemische Studien, Constitutions-Formeln der organischen Chemie in graphischer Darstellung: Das Mariott'sche Gesetz* - Google Play, Wien: Carl Gerold's

Sohn.

- Lucas-Hourani, M. et al., 2013. A phenotypic assay to identify chikungunya virus inhibitors targeting the nonstructural protein nsP2. *Journal of Biomolecular Screening*, 18(2), pp.172–179.
- Lumsden, W.H., 1955. An epidemic of virus disease in Southern Province, Tanganyika Territory, in 1952-53. II. General description and epidemiology. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 49(1), pp.33–57.
- Lyoo, H. et al., 2019. ACBD3 Is an Essential Pan-enterovirus Host Factor That Mediates the Interaction between Viral 3A Protein and Cellular Protein PI4KB. *mBio*, 10(1), pp.1–15.
- Malet, H. et al., 2009. The crystal structures of Chikungunya and Venezuelan equine encephalitis virus nsP3 macro domains define a conserved adenosine binding pocket. *Journal of virology*, 83(13), pp.6534–45.
- Marrapu, V.K. et al., 2011. Novel aryloxy azolyl chalcones with potent activity against Mycobacterium tuberculosis H37Rv. *European Journal of Medicinal Chemistry*, 46(9), pp.4302–4310.
- Martzen, M.R. et al., 1999. A Biochemical Genomics Approach for Identifying Genes by the Activity of Their Products. *Science*, 286(5442), pp.1153–1155.
- McLaughlin, M., Palucki, M. & Davies, I.W., 2006. Efficient access to cyclic ureas via Pd-catalyzed cyclization. *Organic Letters*, 8(15), pp.3311–3314.
- McPherson, R.L. et al., 2017. ADP-ribosylhydrolase activity of Chikungunya virus macrodomain is critical for virus replication and virulence. *Proceedings of the National Academy of Sciences*, 114(7), pp.1666–1671.
- Meinel, C., 1992. *Blick in die Wissenschaft : Forschungsmagazin der Universität Regensburg*, Univ.-Verl.
- Melton, J. V et al., 2002. Alphavirus 6K Proteins Form Ion Channels. *Journal of Biological Chemistry*, 277(49), pp.46923–46931.
- Messacar, K. et al., 2019. Safety, tolerability, and efficacy of fluoxetine as an antiviral for acute flaccid myelitis. *Neurology*, 92(18), pp.e2118–e2126.
- Di Mola, A. et al., 2014. Structure-activity relationship study of arbidol derivatives as inhibitors of chikungunya virus replication. *Bioorganic & medicinal chemistry*, 22(21), pp.6014–25.
- Molecular Biology Institute at the University of California, L.A., 2016. SAVES - Structure Analysis and Verification Sever. *version 4*.
- Morin, B. et al., 2014. *Crystal Structure of Chikungunya virus nsP3 Macro Domain in Complex with a 2'-5' oligoadenylate trimer*,
- NIH US National Library of Medicine & Public Health Service Historian, 2002. Smallpox: Variolation. Available at:

https://www.nlm.nih.gov/exhibition/smallpox/sp_variolation.html [Accessed June 8, 2019].

- Novoa, I. & Carrasco, L., 2015. Cleavage of Eukaryotic Translation Initiation Factor 4G by Exogenously Added Hybrid Proteins Containing Poliovirus 2A pro in HeLa Cells: Effects on Gene Expression. *Molecular and Cellular Biology*, 19(4), pp.2445–2454.
- Oestereich, L. et al., 2014. Successful treatment of advanced Ebola virus infection with T-705 (favipiravir) in a small animal model. *Antiviral Research*, 105(1), pp.17–21.
- Ooi, M.H. et al., 2010. Clinical features, diagnosis, and management of enterovirus 71. *The Lancet Neurology*, 9(11), pp.1097–1105.
- Ozden, S. et al., 2008. Inhibition of Chikungunya Virus Infection in Cultured Human Muscle Cells by Furin Inhibitors. , 283(32), pp.21899–21908.
- Palazzo, L., Mikoč, A. & Ahel, I., 2017. ADP-ribosylation: new facets of an ancient modification. *FEBS Journal*, 284(18), pp.2932–2946.
- De Palma, A.M. et al., 2007. Anti-enterovirus activity and structure-activity relationship of a series of 2,6-dihalophenyl-substituted 1H,3H-thiazolo[3,4-a]benzimidazoles. *Biochemical and Biophysical Research Communications*, 353(3), pp.628–632.
- De Palma, A.M. et al., 2008. The Thiazolobenzimidazole TBZE-029 Inhibits Enterovirus Replication by Targeting a Short Region Immediately Downstream from Motif C in the Nonstructural Protein 2C. *Journal of Virology*, 82(10), pp.4720–4730.
- Papageorgiou, N. et al., 2010. The 2C putative helicase of echovirus 30 adopts a hexameric ring-shaped structure. *Acta Crystallographica Section D: Biological Crystallography*, 66(10), pp.1116–1120.
- Park, E. & Griffin, D.E., 2009a. Interaction of Sindbis virus non-structural protein 3 with poly(ADP-ribose) polymerase 1 in neuronal cells. *The Journal of general virology*, 90(Pt 9), pp.2073–80.
- Park, E. & Griffin, D.E., 2009b. The nsP3 macro domain is important for Sindbis virus replication in neurons and neurovirulence in mice. *Virology*, 388(2), pp.305–14.
- Parvez, M.K., 2015. The hepatitis E virus ORF1 “X-domain” residues form a putative macrodomain protein/Appr-1”-pase catalytic-site, critical for viral RNA replication. *Gene*, 566(1), pp.47–53.
- Patrick, G.L., 2009. An Introduction to Medicinal Chemistry. *UCL lecture notes*, 40, p.752.
- Pauling, L., Corey, R.B. & Branson, H.R., 1951. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences of the United States of America*, 37(4), pp.205–11.
- Pevear, D.C. et al., 1999. Activity of pleconaril against enteroviruses. *Antimicrobial Agents and Chemotherapy*, 43(9), pp.2109–2115.

- Pfister, T. & Wimmer, E., 1999. Characterization of the nucleoside triphosphate activity of poliovirus protein 2C reveals a mechanism by which guanidine. *J Biol Chem*, 274(11), pp.6992–7001.
- Pincus, S.E. et al., 1986. Guanidine-selected mutants of poliovirus: mapping of point mutations to polypeptide 2C. *Journal of virology*, 57(2), pp.638–46.
- Plotkin, S., 2014. History of vaccination. *Proceedings of the National Academy of Sciences of the United States of America*, 111(34), pp.12283–7.
- Pohjala, L. et al., 2011. Inhibitors of alphavirus entry and replication identified with a stable Chikungunya replicon cell line and virus-based assays. *PLoS ONE*, 6(12).
- Pons-Salort, M., Parker, E.P.K. & Grassly, N.C., 2015. The epidemiology of non-polio enteroviruses: Recent advances and outstanding questions. *Current Opinion in Infectious Diseases*, 28(5), pp.479–487.
- Powers, A.M., 2018. Vaccine and therapeutic options to control chikungunya virus. *Clinical Microbiology Reviews*, 31(1), pp.1–29.
- Promega Corporation, 2006. CellTiter 96 AQueous Non-Radioactive Cell Proliferation Assay. *Technical Bulletin*, 313(5783), pp.45–45.
- Prusoff, W.H., 1959. Synthesis and biological activities of iododeoxyuridine, an analog of thymidine. *Biochimica et Biophysica Acta*, 32(1), pp.295–296.
- Pulmanausahakul, R. et al., 2011. Chikungunya in Southeast Asia: understanding the emergence and finding solutions. *International journal of infectious diseases*, 15(10), pp.e671-6.
- Putics, A. et al., 2005. ADP-Ribose-1"-Monophosphatase: a Conserved Coronavirus Enzyme That Is Dispensable for Viral Replication in Tissue Culture. *Journal of Virology*, 79(20), pp.12721–12731.
- Rada, B. & Dragúň, M., 1977. ANTIVIRAL ACTION AND SELECTIVITY OF 6-AZAURIDINE. *Annals of the New York Academy of Sciences*, 284(1 Third Confere), pp.410–417.
- Rana, J. et al., 2014. Network mapping among the functional domains of Chikungunya virus nonstructural proteins. *Proteins: Structure Function and Bioinformatics*, 82(10), pp.2403–11.
- Rarey, M. et al., 1996. A fast flexible docking method using an incremental construction algorithm. *Journal of molecular biology*, 261(3), pp.470–489.
- Rashad, A.A., Mahalingam, S. & Keller, P.A., 2014. Chikungunya Virus : Emerging Targets and New Opportunities for Medicinal Chemistry. *Journal of medicinal chemistry*, 57(4), pp.1147–1166.
- Rathore, A.P.S. et al., 2014. Chikungunya virus nsP3 & nsP4 interacts with HSP-90 to promote virus replication: HSP-90 inhibitors reduce CHIKV infection and inflammation in vivo.

Antiviral research, 103, pp.7–16.

- Reed, L.J. & Muench, H., 1938. A simple method of estimating fifty percent endpoints. *The american journal of hygiene*, 27(3), pp.493–497.
- Reid, C.R., Airo, A.M. & Hobman, T.C., 2015. The Virus-Host Interplay: Biogenesis of +RNA Replication Complexes. *Viruses*, 7(8), pp.4385–4413.
- Rezza, G. et al., 2007. Infection with chikungunya virus in Italy: an outbreak in a temperate region. *Lancet (London, England)*, 370(9602), pp.1840–6.
- Riedel, S., 2005. Edward Jenner and the history of smallpox and vaccination. *Proceedings (Baylor University. Medical Center)*, 18(1), pp.21–5.
- Rightsel, W.A. et al., 1961. Antiviral effect of guanidine. *Science*, 134(3478), pp.558–559.
- Riniker, S., 2018. Fixed-Charge Atomistic Force Fields for Molecular Dynamics Simulations in the Condensed Phase: An Overview. *Journal of Chemical Information and Modeling*, 58(3), pp.565–578.
- Rodriguez, P.L. & Carrasco, L., 1995. Poliovirus Protein 2C Contains Two Regions Involved in RNA Binding Activity. *The Journal of biological chemistry*, 270(17), pp.10105–10112.
- Rosenthal, F. et al., 2013. Macrodomein-containing proteins are new mono-ADP-ribosylhydrolases. *Nature structural & molecular biology*, 20(4), pp.502–7.
- Ross, R.W., 1956. The Newala epidemic. III. The virus: isolation, pathogenic properties and relationship to the epidemic. *The Journal of hygiene*, 54(2), pp.177–191.
- Rothan, H.A. et al., 2016. Mefenamic acid in combination with ribavirin shows significant effects in reducing chikungunya virus infection in vitro and in vivo. *Antiviral Research*, 127, pp.50–56.
- Rupp, J.C. et al., 2015. Alphavirus RNA synthesis and nonstructural protein functions. *The Journal of general virology*, 96(9), pp.2483–2500.
- Russo, A.T., White, M.A. & Watowich, S.J., 2006. The crystal structure of the Venezuelan equine encephalitis alphavirus nsP2 protease. *Structure (London, England : 1993)*, 14(9), pp.1449–58.
- Saarnio, V., 2017. *Antiviral Molecules of Enteroviruses*. University of Jyväskylä.
- Saisawang, C., Saitornuang, S., et al., 2015. Chikungunya nsP2 protease is not a papain-like cysteine protease and the catalytic dyad cysteine is interchangeable with a proximal serine. *Scientific reports*, 5(August), p.17125.
- Saisawang, C., Sillapee, P., et al., 2015. Full length and protease domain activity of chikungunya virus nsP2 differ from other alphavirus nsP2 proteases in recognition of small peptide substrates. *Bioscience reports*, 35(3), pp.1–9.
- Salvador, B. et al., 2009. Characterization of Chikungunya pseudotyped viruses: Identification of refractory cell lines and demonstration of cellular tropism differences mediated by

- mutations in E1 glycoprotein. *Virology*, 393(1), pp.33–41.
- Sastry, G.M. et al., 2013. Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design*, 27(3), pp.221–234.
- Sawicki, D.L. et al., 2006. Role for nsP2 proteins in the cessation of alphavirus minus-strand synthesis by host cells. *Journal of virology*, 80(1), pp.360–71.
- Schilte, C. et al., 2013. Chikungunya Virus-associated Long-term Arthralgia: A 36-month Prospective Longitudinal Study. *PLoS Neglected Tropical Diseases*, 7(3), p.e2137.
- Schlesinger, M., Schlesinger, S. & Kuhn, R., 2011. 13. Togaviruses. In Nicholas H. Acheson, ed. *Fundamentals of Molecular Virology, 2nd Edition*. Wiley-VCH Verlag GmbH, pp. 148–158.
- Schmidt, T., Bergner, A. & Schwede, T., 2014. Modelling three-dimensional protein structures for applications in drug design. *Drug Discovery Today*, 19(7), pp.890–897.
- Schneider, G. & Baringhaus, K.-H., 2008. *Molecular Design: Concepts and Applications*, Wiley-VCH.
- Schneidman-Duhovny, D. et al., 2005a. Geometry-based flexible and symmetric protein docking. *Proteins: Structure, Function and Genetics*, 60(2), pp.224–231.
- Schneidman-Duhovny, D. et al., 2005b. PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Research*, 33(SUPPL. 2), pp.363–367.
- Scholte, F.E.M. et al., 2015. Stress granule components G3BP1 and G3BP2 play a proviral role early in Chikungunya virus replication. *Journal of virology*, 89(8), pp.4457–69.
- Schreiber, V. et al., 2006. Poly(ADP-ribose): novel functions for an old molecule. *Nature reviews. Molecular cell biology*, 7(7), pp.517–28.
- Schrödinger LLC, 2016. Schrödinger Release 2016-1: Maestro, Schrödinger, LLC, New York, NY, 2016.
- Schrödinger LLC, 2018. Schrödinger Release 2018-4: Maestro, Schrödinger, LLC, New York, NY, 2018.
- Schrödinger LLC & DE Shaw Research, 2017. Schrödinger Release 2017-4: Desmond Molecular Dynamics System, DE Shaw Research, New York, NY, 2017.
- Schuffenecker, I. et al., 2006. Genome microevolution of chikungunya viruses causing the Indian Ocean outbreak. *PLoS medicine*, 3(7), p.e263.
- Schwartz, O. & Albert, M.L., 2010. Biology and pathogenesis of chikungunya virus. *Nature reviews. Microbiology*, 8(7), pp.491–500.
- Seyedi, S.S. et al., 2016. Computational Approach Towards Exploring Potential Anti-Chikungunya Activity of Selected Flavonoids. *Scientific Reports*, 6(March), pp.1–8.
- Sharifi, R. et al., 2013. Deficiency of terminal ADP-ribose protein glycohydrolase TARG1/C6orf130 in neurodegenerative disease. *EMBO Journal*, 32(9), pp.1225–1237.

- Shim, A. et al., 2016. Therapeutic and prophylactic activity of itraconazole against human rhinovirus infection in a murine model. *Scientific Reports*, 6(March), pp.1–12.
- Shimizu, H. et al., 2000. Mutations in the 2C Region of Poliovirus Responsible for Altered Sensitivity to Benzimidazole Derivatives. *Journal of Virology*, 74(9), pp.4146–4154.
- Shin, G. et al., 2012. Structural and functional insights into alphavirus polyprotein processing and pathogenesis. *Proceedings of the National Academy of Sciences*, 109(41), pp.16534–16539.
- Shirako, Y. & Strauss, J.H., 1994. Regulation of Sindbis virus RNA replication: uncleaved P123 and nsP4 function in minus-strand RNA synthesis, whereas cleaved products from P123 are required for efficient plus-strand RNA synthesis. *Journal of virology*, 68(3), pp.1874–85.
- da Silva-Júnior, E.F. et al., 2017. The medicinal chemistry of Chikungunya virus. *Bioorganic and Medicinal Chemistry*, 25(16), pp.4219–4244.
- Singh, H. et al., 2018. Chikungunya virus inhibition by peptidomimetic inhibitors targeting virus-specific cysteine protease. *Biochimie*, 149, pp.51–61.
- Singh, K.D. et al., 2012. Homology modeling, molecular dynamics, e-pharmacophore mapping and docking study of Chikungunya virus nsP2 protease. *Journal of molecular modeling*, 18(1), pp.39–51.
- Singh, S.K. & Unni, S.K., 2011. Chikungunya virus : host pathogen interaction. *Reviews in Medical Virology*, 21(2), pp.78–88.
- Smith, K.A., 2011. Edward Jenner and the small pox vaccine. *Frontiers in Immunology*, 2(JUN), pp.1–6.
- Smith, K.A., 2012. Louis Pasteur, the father of immunology? *Frontiers in Immunology*, 3(APR), pp.1–10.
- Soga, S. et al., 2007. Use of Amino Acid Composition to Predict Ligand-Binding Sites. *Journal of Chemical Information and Modeling*, 47(2), pp.400–406.
- Solignat, M. et al., 2009. Replication cycle of chikungunya : A re-emerging arbovirus. *Virology*, 393(2), pp.183–197.
- Sreejith, R. et al., 2012. Mapping interactions of Chikungunya virus nonstructural proteins. *Virus research*, 169(1), pp.231–6.
- Staples, J.E., Breiman, R.F. & Powers, A.M., 2009. Chikungunya fever: an epidemiological review of a re-emerging infectious disease. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 49(6), pp.942–8.
- Strating, J.R.P.M. et al., 2015. Itraconazole inhibits enterovirus replication by targeting the oxysterol-binding protein. *Cell Reports*, 10(4), pp.600–615.
- Strauss, D.M., Glustrom, L.W. & Wuttke, D.S., 2003. Towards an Understanding of the

- Poliovirus Replication Complex: The Solution Structure of the Soluble Domain of the Poliovirus 3A Protein. *Journal of Molecular Biology*, 330(2), pp.225–234.
- Strauss, J.H. & Strauss, E.G., 1994. The alphaviruses: gene expression, replication, and evolution. *Microbiological reviews*, 58(3), pp.491–562.
- Sun, S. et al., 2013. Structural analyses at pseudo atomic resolution of Chikungunya virus and antibodies show mechanisms of neutralization. *eLife*, 2, p.e00435.
- Sweeney, T.R. et al., 2010. Foot-and-mouth disease virus 2C is a hexameric AAA+ protein with a coordinated ATP hydrolysis mechanism. *Journal of Biological Chemistry*, 285(32), pp.24347–24359.
- Tapparel, C. et al., 2013. Picornavirus and enterovirus diversity with associated human diseases. *Infection, Genetics and Evolution*, 14(1), pp.282–293.
- Teterina, N.L. et al., 1997. Poliovirus 2C protein determinants of membrane binding and rearrangements in mammalian cells. *Journal of virology*, 71(12), pp.8962–72.
- Thèves, C., Crubézy, E. & Biagini, P., 2016. History of Smallpox and Its Spread in Human Populations. *Paleomicrobiology of Humans*, (January 2018), pp.161–172.
- Thibaut, H.J., De Palma, A.M. & Neyts, J., 2012. Combating enterovirus replication: State-of-the-art on antiviral research. *Biochemical Pharmacology*, 83(2), pp.185–192.
- Tiefenbrunn, T. et al., 2014. Crystallographic Fragment-Based Drug Discovery: Use of a Brominated Fragment Library Targeting HIV Protease. , pp.141–148.
- Tironi, I.G. et al., 1995. A generalized reaction field method for molecular dynamics simulations. *The Journal of Chemical Physics*, 102(13), pp.5451–5459.
- To, J., Surya, W. & Torres, J., 2016. *Targeting the Channel Activity of Viroporins* 1st ed., Elsevier Inc.
- Tolskaya, E.A. et al., 1994. Genetic studies on the poliovirus 2C protein, an NTPase A plausible mechanism of guanidine effect on the 2C function and evidence for the importance of 2C oligomerization. *Journal of Molecular Biology*, 236(5), pp.1310–1323.
- Tomar, S. et al., 2006. Catalytic core of alphavirus nonstructural protein nsP4 possesses terminal adenylyltransferase activity. *Journal of virology*, 80(20), pp.9962–9969.
- Tracy, S. et al., 2010. Enteroviruses, type 1 diabetes and hygiene: a complex relationship. *Reviews in Medical Virology*, 20, pp.106–116.
- Tracy, S. (Steven), Oberste, M.S. & Drescher, K.M., 2008. *Group B Coxsackieviruses*, Springer.
- Tsetsarkin, K.A. et al., 2009. Epistatic roles of E2 glycoprotein mutations in adaptation of chikungunya virus to *Aedes albopictus* and *Ae. aegypti* mosquitoes. *PLoS one*, 4(8), p.e6835.
- Tsou, Y.L. et al., 2013. Heat Shock protein 90: Role in Enterovirus 71 Entry and Assembly and Potential Target for Therapy. *PLoS ONE*, 8(10), pp.1–13.

- Tuthill, T. et al., 2010. Picornaviruses. *Current topics in microbiology and immunology*, 343, pp.43–89.
- Uchime, O., Fields, W. & Kielian, M., 2013. The role of E3 in pH protection during alphavirus assembly and exit. *Journal of virology*, 87(18), pp.10255–62.
- Ulferts, R. et al., 2016. Screening of a library of FDA-approved drugs identifies several enterovirus replication inhibitors that target viral protein 2C. *Antimicrobial Agents and Chemotherapy*, 60(5), pp.2627–2638.
- Ulferts, R. et al., 2013. Selective serotonin reuptake inhibitor fluoxetine inhibits replication of human enteroviruses B and D by targeting viral protein 2C. *Antimicrobial Agents and Chemotherapy*, 57(4), pp.1952–1956.
- Varjak, M., Zusinaite, E. & Merits, A., 2009. Novel Functions of the Alphavirus Nonstructural Protein nsP3 C-Terminal Region. *Journal of Virology*, 84(5), pp.2352–2364.
- Vasiljeva, L. et al., 2003. Regulation of the sequential processing of Semliki Forest virus replicase polyprotein. *The Journal of biological chemistry*, 278(43), pp.41636–45.
- Vasiljeva, L. et al., 2001. Site-specific protease activity of the carboxyl-terminal domain of Semliki Forest virus replicase protein nsP2. *The Journal of biological chemistry*, 276(33), pp.30786–93.
- Venclovas, Č. & Margelevičius, M., 2005. Comparative modeling in CASP6 using consensus approach to template selection, sequence-structure alignment, and structure assessment. *Proteins: Structure, Function and Genetics*, 61(SUPPL. 7), pp.99–105.
- Vivelo, C.A. & Leung, A.K.L., 2015. Proteomics approaches to identify mono-(ADP-ribosyl)ated and poly(ADP-ribosyl)ated proteins. *Proteomics*, 15(2–3), pp.203–217.
- Voss, J.E. et al., 2010. Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. *Nature*, 468(7324), pp.709–12.
- Walker, E.J. et al., 2013. Rhinovirus 3C Protease Facilitates Specific Nucleoporin Cleavage and Mislocalisation of Nuclear Proteins in Infected Host Cells. *PLoS ONE*, 8(8).
- Wang, C. et al., 2014. A C-terminal, cysteine-rich site in poliovirus 2CATPase is required for morphogenesis. *Journal of General Virology*, 95(PART 6), pp.1255–1265.
- Wang, C. et al., 2012. Alanine Scanning of Poliovirus 2C ATPase Reveals New Genetic Evidence that Capsid Protein/2C ATPase Interactions Are Essential for Morphogenesis. *Journal of Virology*, 86(18), pp.9964–9975.
- Wang, J., Cieplak, P. & Kollman, P.A., 2000. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? Keywords: additive force field; nonadditive force field; restrained electrostatic potential (RESP); torsion. *Journal of Computational Chemistry*, 21(12), pp.1049–1074.

- Wang, R. & Wang, S., 2002. How Does Consensus Scoring Work for Virtual Library Screening? An Idealized Computer Experiment. *Journal of Chemical Information and Computer Sciences*, 41(5), pp.1422–1426.
- Watson, K.G. et al., 2003. An orally bioavailable oxime ether capsid binder with potent activity against human rhinovirus. *Journal of Medicinal Chemistry*, 46(15), pp.3181–3184.
- Weaver, S.C., 2014. Arrival of Chikungunya Virus in the New World: Prospects for Spread and Impact on Public Health. *PLoS Neglected Tropical Diseases*, 8(6), pp.6–9.
- Weaver, S.C. & Barrett, A.D.T., 2004. Transmission cycles, host range, evolution and emergence of arboviral disease. *Nature Reviews Microbiology*, 2(10), pp.789–801.
- Weaver, S.C. & Forrester, N.L., 2015. Chikungunya: Evolutionary history and recent epidemic spread. *Antiviral Research*, 120, pp.32–39.
- Weber, C. et al., 2015. The green tea catechin, epigallocatechin gallate inhibits chikungunya virus infection. *Antiviral Research*, 113, pp.1–3.
- Wermuth, C. et al., 1998. Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). *Pure and Applied Chemistry*, 70(5), pp.1129–1143.
- Wessels, E. et al., 2006. Structure-function analysis of the coxsackievirus protein 3A: Identification of residues important for dimerization, viral RNA replication, and transport inhibition. *Journal of Biological Chemistry*, 281(38), pp.28232–28243.
- Whitehead, J. et al., 2016. Trial design for evaluating novel treatments during an outbreak of an infectious disease. *Clinical Trials*, 13(1), pp.31–38.
- WHO, 2018a. Dengue vaccine: WHO position paper - September 2018. *Weekly epidemiological record*, 93(36), pp.457–476.
- WHO, 2016. Meeting of the Strategic Advisory Group of Experts on immunization, October 2016 - conclusions and recommendations. *Weekly epidemiological record*, 48(91), pp.561–584.
- WHO, 2018b. WHO | Smallpox. *WHO*.
- Wildman, S.A. & Crippen, G.M., 1999. Prediction of Physicochemical Parameters by Atomic Contributions. *Journal of Chemical Information and Modeling*, 39(5), pp.868–873.
- Wilkinson, T.A. et al., 2005. Association of sindbis virus capsid protein with phospholipid membranes and the E2 glycoprotein: implications for alphavirus assembly. *Biochemistry*, 44(8), pp.2800–10.
- Wintachai, P. et al., 2015. Assessment of flavaglines as potential chikungunya virus entry inhibitors. *Microbiology and Immunology*, 59(3), pp.129–141.
- Wong, D.T., Perry, K.W. & Bymaster, F.P., 2005. The Discovery of Fluoxetine Hydrochloride (Prozac). *Nature Reviews Drug Discovery*, 4(9), pp.764–774.
- Xia, H. et al., 2015. Human Enterovirus Nonstructural Protein 2C^{ATPase} Functions as Both an RNA

- Helicase and ATP-Independent RNA Chaperone. *PLoS Pathogens*, 11(7), pp.1–29.
- Yi, E.-J. et al., 2017. Enterovirus 71 infection and vaccines. *Clinical and experimental vaccine research*, 6, pp.4–14.
- Yu, S.F. & Lloyd, R.E., 1991. Identification of essential amino acid residues in the functional activity of poliovirus 2A protease. *Virology*, 182(2), pp.615–625.
- Zell, R. et al., 2017. ICTV virus taxonomy profile: Picornaviridae. *Journal of General Virology*, 98(10), pp.2421–2422.
- Zhang, L. et al., 2015. Discovery of Benzylidene Derivatives as Potent Syk Inhibitors: Synthesis, SAR Analysis, and Biological Evaluation. *Archiv der Pharmazie*, 348(7), pp.463–474.
- Zhen, X.L. et al., 2006. 4-(2,4-Dichlorobenzyloxy)-3-ethoxybenzaldehyde. *Acta Crystallographica Section E: Structure Reports Online*, 62(12), pp.5794–5795.
- Zonsics, B., 2013. *Identification of HIV protease inhibitors by in silico screening*. University of Vienna.
- Zuo, J. et al., 2016. Discovery of structurally diverse small-molecule compounds with broad antiviral activity against enteroviruses. *Antimicrobial Agents and Chemotherapy*, 60(3), pp.1615–1626.
- Zuo, J. et al., 2012. Fluoxetine is a potent inhibitor of coxsackievirus replication. *Antimicrobial Agents and Chemotherapy*, 56(9), pp.4838–4844.