Reaction-Driven Fragment-Based Design of Viral Protease Inhibitors

Inaugural-Dissertation
to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat)
submitted to the Department of Biology, Chemistry and Pharmacy
of Freie Universität Berlin

by
Robert Schulz
from Berlin, Germany

2017

The presented thesis was prepared from April 2013 till April 2017 under the supervision of Prof. Dr. Gerhard Wolber at the Institute of Pharmacy of the Freie Universität Berlin.

1. Gutachter: Prof. Dr. Gerhard Wolber
2. Gutachter: Prof. Dr. Jörg Rademann

Disputation am: 12.12.2017

Acknowledgments

Firstly, I would like to express my sincere gratitude to my supervisor Prof. Gerhard Wolber for his continuous support, the opportunities and freedom he granted me during the thesis.

Secondly, I would like to thank Prof. Jörg Rademann for the great collaboration that enabled the thesis through the use of material and equipment of his lab.

Furthermore, I want to thank past and present members of the computer-aided drug design group for the nice and conducive working atmosphere, especially first floor inhabitants Alexandra Naß, Anne Frank and David Schaller.

I would also express my gratitude to Dr. Susanne Dupré for her contribution to the work group and especially for the thesis.

Further thanks go to members of the group of Prof. Rademann for their help with experimental methods and procedures. I want to especially mention Daniel Becker, Franziska Gottschalk and Caroline Tauber for the fruitful discussions and collaborative work.

Gratefully acknowledged are the synthesis efforts of Amira Atef Suliman and Je-Ryun Lee.

I would also like to thank all people involved in the EUROPIN PhD program for the critical review of my research and the insightful opportunities offered by participation.

Furthermore, I want to thank past and present members of the assistant team for the drug analytic teaching for making this duty so enjoyable.

Finally, I would like to thank my family and friends for their continuous support during this time, and most of all Jule.

Science is not everything,
but science is very beautiful.

J. Robert Oppenheimer

Abstract

Highly pathogenic enteroviruses such as CVB3, EV D68 or EV71 have emerged as global threat for the human health with no treatment available to date. Polyprotein processing by the viral 3C protease ($3C^{pro}$) represents an essential step within the replication cycle of all enteroviruses, which spurred the development of anti-enterovirus agents by targeting the $3C^{pro}$. Peptidomimetic analogues of the natural substrate are the most potent inhibitors developed so far. Yet, their inherent poor pharmacokinetic properties due to their peptidic nature limit clinical applicability.

The primary aim of this thesis is the design of non-peptidic, covalent 3C protease inhibitors to cope with the impending threat of enterovirus infections. This was pursued through a computer-driven fragment-based approach in close collaboration with the group of Prof. Rademann.

At first, the structural characteristics essential for covalent CVB3 $3C^{pro}$ binding were derived from co-crystal structures and transferred into a 3D pharmacophore model, which served as basis for the subsequent fragment screening of an in-house small molecule collection. A virtual screening enriched selection of fragments was evaluated in a FRET-based enzyme kinetic assay, which led to the identification of a fragment hit with favorable $3C^{pro}$ binding characteristics. Due to its hydrolytic instability, scaffold hopping within commercial fragment space was pursued, which led the identification of additional $3C^{pro}$ binding fragments. Integration of protein mass spectrometry data into the established structural model enabled the rationalization of the fragment $3C^{pro}$ binding through a novel type of covalently binding substructure. For the fragment optimization, a rational *de novo* design workflow was developed, in which commercially available building blocks and synthesis oriented reaction rules are applied to grow the fragment within the active site of the $3C^{pro}$ *in silico*. Proposed fragment analogues were synthesized by the collaborating group according to the virtual design scheme, which led to the identification of $3C^{pro}$ inhibitors active in the micromolar range.

In the second part of the thesis, the analysis of a noncompetitive $3C^{pro}$ inhibiting fragment, identified through the scaffold hopping, is described. Structure activity relationships derived from literature and through the biological evaluation of analogues were used for subsequent binding hypothesis generation. An allosteric pocket was identified through computational hot spot analysis and molecular dynamics simulations within the RNA binding site of the $3C^{pro}$, which essen-

tial role in enterovirus genome replication provides biological relevance. A model for allosteric $3C^{pro}$ binding was developed through integration of the structure activity relationship data and statistical information of protein-ligand interaction patterns derived from molecular dynamics simulations.

The reported findings provide a novel way to target the pivotal $3C^{pro}$ as our identified and rationally designed inhibitors represent a new structural basis for the development of non-peptidic anti-enteroviral drugs. In addition, the identified allosteric binding site enables a novel, complementing way to address this essential target and thus to combat enterovirus infections.

Zusammenfassung

Das Auftreten hochpathogener Enteroviren wie CVB3, EV D68 oder EV 71 stellt eine globale Gefahr für die menschliche Gesundheit dar. Die Spaltung des viralen Polyproteins durch die 3C Protease ist ein essentieller Schritt im viralen Replikationszyklus aller Enteroviren, weshalb sich die Entwicklung von Wirkstoffen auf die 3C Protease konzentriert hat. Die potentesten daraus hervorgegangenen Inhibitoren der 3C Protease sind Analoga des natürlichen Substrates, deren peptidische Struktur zu pharmakokinetischen Nachteilen führt und den klinischen Nutzen einschränkt.

Das Ziel dieser Arbeit besteht in dem Design neuartiger, nicht-peptidischer Inhibitoren für die 3C Protease als Grundlage für Entwicklung anti-enteroviraler Wirkstoffe. Dies wurde über einen computergestützten, fragment-basierten Ansatz in Kooperation mit der Arbeitsgruppe von Prof. Rademann verfolgt.

Zunächst wurden die für die kovalente CVB3 3C Protease Bindung essentiellen Strukturmerkmale aus Kristallstrukturen abgeleitet und in ein 3D-Pharmakophormodell überführt, das als Grundlage für ein sich anschließendes Fragment Screening einer in-house Molekülbibliothek diente. Die biologische Testung von durch das Bindungsmodell vorselektierten Fragmenten in einem FRET-basierten enzymkinetischen Assay lieferte einen Hit mit günstigen 3C Protease Bindungseigenschaften. Auf Grund dessen hydrolytischer Instabilität wurde eine Suche nach Strukturanaloga in Bibliotheken von kommerziell erhältlichen Fragmenten durchgeführt, die zur Identifizierung weiterer 3C Protease bindender Fragmente führte. Mit Hilfe massenspektrometrischer Daten konnte die Protein-Fragment Bindung über eine neuartige kovalent-bindende Substruktur aufgeklärt und eine rationale Bindungshypothese erstellt werden. Für die Optimierung der Fragmente wurde ein *de novo*-Design basierter Workflow entwickelt, in dem Moleküle in der aktiven Tasche der 3C Protease virtuell durch das Anknüpfen chemischer Bausteine optimiert werden. Vorgeschlagene Verbindungen wurden in der Kooperationsgruppe dem virtuellen Schema folgend synthetisiert und führten zur Identifizierung micro molarer 3C Protease Inhibitoren.

Im zweiten Teil der Arbeit wird die Analyse eines nicht-kompetitiv bindenden Fragments beschrieben, welches ebenfalls im Rahmen der Suche nach Strukturanaloga identifiziert wurde. Strukturaktivitätsbeziehungen wurden durch die biologische Testung weiterer Analoga gewonnen oder der Literatur entnommen und für die anschließende Generierung von Bindungshypothesen verwendet. Eine

allosterische Bindungsstelle wurde durch computer-basierte Hotspot Analysen und Molekulardynamik-Simulationen innerhalb der RNA-Bindungsdomäne der 3C Protease identifiziert, die aufgrund ihrer essentiellen Rolle für die Enterovirus Genomreplikation biologische Relevanz aufweist. Die Integration der erhaltenen Strukturaktivitätsbeziehungen und von Molekulardynamik-Simulationen abgeleiteten Protein-Ligand Interaktionsmustern ermöglichte die Generierung eines dynamischen Modells für die allosterische Bindung von Inhibitoren an die 3C Protease.

Die im Rahmen dieser Arbeit identifizierten und rational entworfenen Inhibitoren ermöglichen einen neuen Ansatz zur Hemmung der 3C Protease für die Entwicklung nicht-peptidischer Wirkstoffe zur Therapie von Enterovirus Infektionen. Darüber hinaus ermöglicht die identifizierte allosterische Bindungsstelle eine ergänzende Möglichkeit, die Aktivität der 3C Protease zu modulieren.

Contents

Chapter 1

Introduction

1.1 Viruses as human pathogens

During the preparation of this thesis the public has been faced with the occurrence of newly emerging viruses. In 2013, coronavirus infections in the Middle East reawakened the memories of the SARS epidemic only a decade ago [1]. Western Africa has been facing the most severe outbreak of hemorrhagic fever caused by Ebola virus infections in 2015 with over ten thousand fatal cases so far [2]. Increasing incidence of microcephaly in new-borns in Brazil were associated with ZIKA virus infection in early 2016 [3]. Less noticed by the public, health systems are facing the spread of mosquito-borne flavivirus infections and increasing frequencies of severe enterovirus outbreaks [4–6].

For none of these viral infections a specific treatment has been and is still not available. This exemplifies that viral infections are a serious threat to the human health and the development of antiviral drugs is of utmost importance. The following section will provide an introduction to the viral replication, means for antiviral interferences and the enteroviruses as emerging pathogens will be described in detail.

1.1.1 (Breaking) the infectious cycle

Origin and classification of viruses with respect to other biological entities is still a matter of extensive debate [7]. Viruses have distinctive features, which differentiate them from all other biological entities. First of all, viruses are particle consisting of genetic material within a protein shell, the capsid. Some viruses, termed enveloped viruses, additionally have a lipid membrane around the capsid. The most important difference is that viruses cannot replicate by themselves but are dependent on cellular organisms for reproduction. This is achieved by infecting host cells, hijacking of the cellular machinery and inducing the generation of new viral particles, which can in turn infect new cells, resulting in the so called infectious cycle.

The complete dependence on host cells for reproduction explains why viruses are pathogens. In order to understand in which way drugs can be used to combat viral infections, it is necessary to have a closer look at the infectious cycle (Figure 1.1). The virus-host interaction is very specific so that a virus can only enter specific, *susceptible* cells and in turn is only replicated in specific, *permissive* cells.

The first step of viral replication is the attachment of a virus particle to a susceptible cell characterized by the presence of a specific protein at the cellular mem-

brane, which functions as specific receptor for the virus independent from its inherent function. Binding of the virus to its receptor enables fusion with the cellular membrane in case of enveloped viruses or triggers endocytosis for non-enveloped viruses causing the virus particle to enter the host cell.

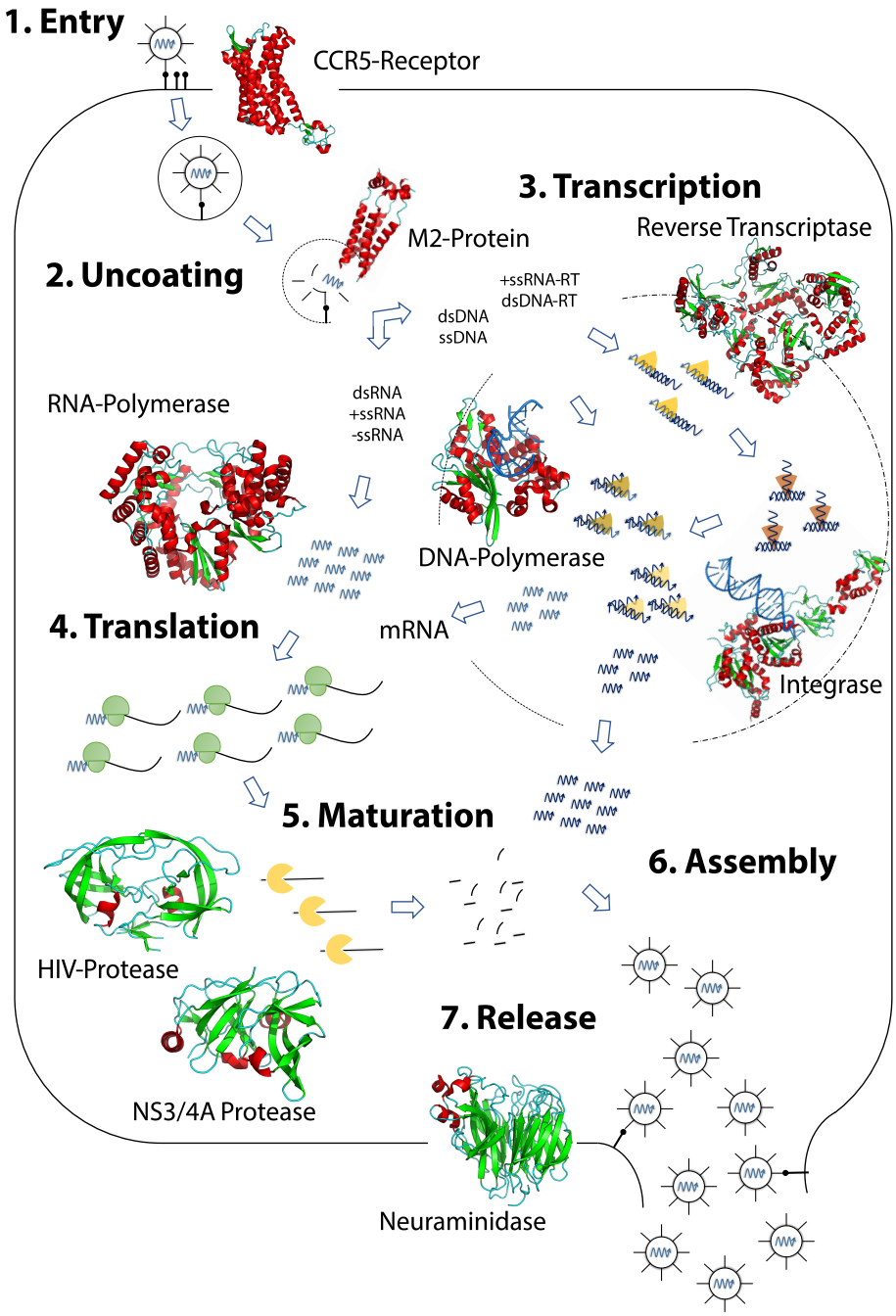

Figure 1.1: Viral replication scheme including key steps and involved drug targets.

Subsequently, the viral genome is released from the capsid facilitated by structural changes of capsid proteins due to acidification along the endocytic pathway.

The released genome and accompanying viral proteins take over the cellular machinery by blocking protein synthesis and immune defense mechanisms. For the production of new virus particles, +mRNA needs to be formed, from which the necessary viral proteins can be expressed. The Baltimore scheme, as depicted in Figure 1.2, represents a classification system of viruses based on their genetic material and their path toward +mRNA [8]. Two main groups can be differentiated: Those that either contain or form DNA and those that stay on the RNA level. Viruses of the groups III, IV and V belong to the latter. They dependent on the RNA-dependent RNA polymerase (RdRp), which does not exists in humans, for +mRNA synthesis upon release from the capsid either directly from the negative strand present in the groups III and V or via the synthesis of a negative strand intermediate as template (group IV). DNA viruses enter the nucleus of the host cell where the dsDNA is either transcribed into mRNA directly (group I), after synthesis of the complementary strand (group II) or after filling of the gaps in the group VII virus genomes. Group VI, to which the pathogenic human immunodeficiency virus (HIV) belongs, follows an unique mechanism. These viruses contain a specific enzyme that can produce DNA from RNA, which is termed reverse transcriptase. It generates dsDNA from the +ssRNA genome of the retroviruses. This dsDNA is subsequently incorporated into the host cell genome by the integrase, another specific enzyme of retroviruses.

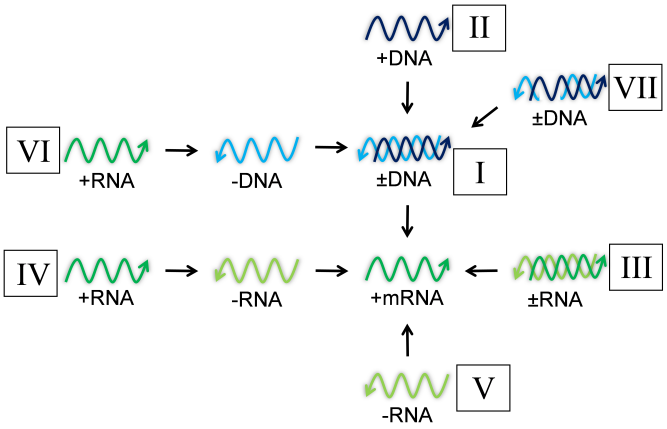

Figure 1.2: Baltimore scheme for the classification of viruses.

After formation of viral mRNA, it is translated by host cell ribosomes. Viruses have evolved strategies to abrogate the translation of host mRNA and simultaneously to enable viral protein synthesis from mRNA missing specific elements essential for the interaction with the ribosome and translation factors such as the cap at the 5′ end. In contrast to eukaryotic mRNA, which generally encodes a

single protein from one open reading frame (ORF), viral mRNA from the picornaviridae, containing only a single ORF, is expressed as a polyprotein comprising all viral proteins. Through subsequent proteolytic cleavage by viral proteases the viral proteins are released, which is termed maturation. Different cleavage sites and degrees of proteolysis result in viral proteins and premature adducts contributing to viral replication.

Enveloped viruses assemble at the cellular membrane, where they induce the formation of buds in the membrane, from which the new, enveloped particles are formed. Therefore, this is termed budding. Non-enveloped viruses, which do not bud, use different pathways for release. Some take advantage of cellular vesicle formation and trafficking system to be released via exocytosis. Other viruses exploit the disintegration of the cellular membrane resulting from apoptosis, lack of integral membrane proteins or pore forming viral proteins.

Infected organisms have evolved mechanisms to detect and fight virus infections. Characteristic structural patterns of viruses are sensed by specific receptors of the immune system such as the toll-like receptor 3, which recognizes dsRNA [9]. This initiates the production of cytokines, which induce the formation of antiviral proteins, attract cells of the innate immune system and stimulate the adaptive immune response. The selective pressures of the host immune system have let viruses to evolve strategies to evade defense mechanisms. This is facilitated by their replication mechanism in which a single viral particle is replicated ever so often leading to a vast number of slightly different copies (quasispecies) [10]. The differences result from mutations introduced during DNA replication and transcription, variable protein expression due to frame shifting, leaky scanning and altered initiation or termination during translation and reassortment and rearrangement of the viral genome during assembly. Altogether, this leads to a huge diversity in the viral progeny, from which those that are most adapt to the current environmental pressures, like host immune system or drugs, are selected and can infect new cells.

The emergence and spread of new, potentially hazardous viruses is not only driven by the evolution of the virus. High density human population and frequent travel allow rapid spread of viral infections. Whereas the invasion of ecosystem and environmental changes foster the transmission of viruses to new host species or previously inaccessible ones [11]. These encounters are often severe as neither the host nor the virus have been adapted to one another.

Despite the threat of existing and emerging viruses, the number of available

Table 1.1: Major targets of antiviral drugs. The viruses are abbreviated as follows: HIV, human immunodeficiency virus; HSV, herpes simplex virus; HCV, hepatitis C virus.

Step	Virus	Target	Drug
Attachment	HIV	CCR5	Maraviroc
Uncoating	Influenza virus	M2-Protein	Amantadine
Replication	HSV	DNA-Polymerase	Aciclovir
	HCV	RNA-Polymerase	Sofosbuvir
	HIV	Reverse transcriptase	Emtricitabine
	HIV	Integrase	Raltegravir
Maturation	HIV	HIV-Protease	Darunavir
	HCV	NS3/4A Protease	Boceprevir
Budding	Influenza virus	Neuraminidase	Oseltamivir
Unspecific	HCV	NS5A	Ledipasvir

antiviral agents is limited [12]. Table 1.1 shows that these cover most of the steps within the viral replication cycle. The attachment of HIV through the interaction with its receptor proteins is blocked by the drug maraviroc, which binds to CCR5 [13]. Uncoating of the influenza virus is prevented by amatandine, which blocks the M2-protein [14], a proton transporter located at the viral envelope, which forwards the acidification during endocytosis to the capsid and thus induces uncoating. Nucleoside analogues are used to inhibit viral polymerases like acivlovir in the treatment of herpes simplex infections [15]. Inhibitors of the reverse transcriptase, such as emtricitabine, are the backbone of the highly active antiretroviral therapy (HAART) for the treatment of HIV infections [16]. Two nucleoside reverse transcriptase inhibitors (NRTI) are thereby combined with either a non-nucleoside reverse transcriptase inhibitor (NNRTI), like efavirenz [17], an integrase inhibitor such as raltegravir [18] or an inhibitor of the HIV-protease like darunavir [19]. The therapy of hepatitis C virus infections has experienced a most recent breakthrough with the approval of the viral RNA-polymerase (NS5B) inhibitor sofosbuvir [20], especially in combination with inhibitors of the viral NS5A protein such as ledipasvir [21], which results in high cure rates for this otherwise chronic infection [22]. Budding of influenza virus particles can be prevented by neuraminidase inhibitors like oseltamivir [23].

From this scarce overview of antiviral drugs, it becomes apparent that only a handful of virus infections can specifically be treated by small molecule drugs. This indicates the challenges for the development of antivirals but also the lack of inhibitors with broad spectrum activity. An exception to this is ribavirin, a nu-

cleoside analogue, with antiviral activity against several RNA and DNA-viruses through a multitude of biological effects including the inhibition of the viral enzyme RdRp and inosine-5′-monophosphate dehydrogenase [24].

1.1.2 Enteroviruses, an emerging threat

Within the picornaviridae, enteroviruses form a genus comprising several viruses infecting humans most frequently. This includes the enterovirus species A-D (E-J infect animals) and the rhinovirus species A-C. These are small non-enveloped viruses of the group IV in the Baltimore classification scheme. The +ssRNA contains a single ORF framed by untranslated regions (UTR) regions at the 5′- and 3′-end. These contain secondary RNA structures essential for replication (cloverleaf) and cap-independent translation (Figure 1.3) [25]. The expressed polyprotein is subsequently cleaved by the viral proteases 2A and 3C to yield the structural and functional viral proteins listed in Table 1.2.

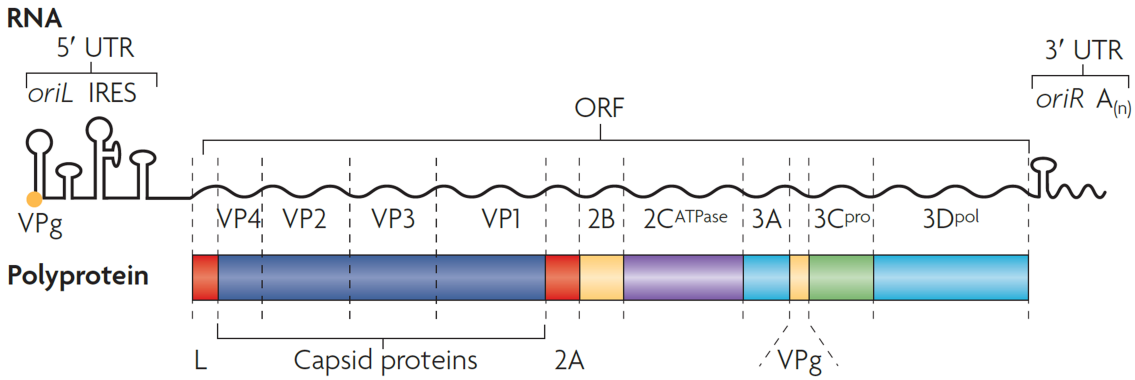

Figure 1.3: Schematic representation of the picornaviridae genome adapted from Ago et al. [26].

Enteroviruses cause mild flu-like symptoms after infections of the upper respiratory- or the gastrointestinal tract. But several members have emerged as cause for severe diseases especially in children [27]. Infections with coxsackieviruses typically lead to the hand, foot and mouth disease characterized by rashes at the respective body parts [28]. Most recently the incidence of coxsackievirus infections associated with the formation aseptic meningitis increased significantly [29]. Furthermore, coxsackievirus infections are the most common cause for viral myocarditis, which can lead to sudden cardiac death in acute or dilated cardiomyopathy in chronic infections [30]. The pathophysiological mechanism has been pinpointed to the ability of the 2A protease to cleave dystrophin [31, 32]. In addition, infections of the pancreas cannot only cause pancreatitis but may also play a role in

Table 1.2: Proteins encoded by the picornavirus polyprotein and their functions.

Protein(s)	Function(s)
VP1, VP2, VP3, VP4	Capsid formation
L (not conserved)	Polyprotein & host protein processing
2A	Polyprotein & host protein processing
2B	Modulation of secretory pathways
2C	Vesicle formation, NTPase activity
3A	Shut down of intracellular transport
3B (VPg)	RNA replication priming
3C	Polyprotein & host protein processing
3D	RNA-dependent RNA polymerase

the etiology of type-I diabetes, which is still controversially debated [33–35]. A recent publication by Mantri *et al.* proposed a link between coxsackievirus infections and dementia in an immunodeficient patient [36]. This would not only provide a new perspective on the pathogenesis and possible therapy of Alzheimer's disease but also for the pathogenic threat originating from coxsackievirus infections.

Another emerging pathogen is the enterovirus EV D68, which was once rare, but occurred in several serotypes over the past decade as the cause for severe respiratory illnesses and acute flaccid myelitis [37, 38]. The latter is better known as poliomyelitis, which occurs in 1-2 % of patients infected with the poliovirus and leads to lifelong deformity of the limbs or even death if muscles of respiration are affected [39]. This threatening disease has triggered extensive research on enteroviruses enabling the development of a vaccine against the poliovirus, which has led to an almost complete eradication of poliovirus infections [40]. Apart from the enterovirus EV 71, no vaccine is available for any other enterovirus, which is due to the huge number and variety of serotypes for the different enterovirus genera. Especially the human rhinovirus comprises over 100 serotypes, which circulate within the human population and are the most frequent cause of the common cold [41]. Even though, this infection of the upper respiratory tract generally progresses mildly it is of major significance due to high frequency of infection and the resulting economic impact due to the loss of work days [42].

The lack of vaccines for the majority of enteroviruses is accompanied by the absence of an approved drug, which poses a threatening situation in connection with the looming of emerging enteroviruses.

1.2 Targeting peptide cleaving enzymes

Enzymes are essential for all biological systems as they enable reactions to take place within the physical conditions and the time frame that these can tolerate [43]. Due to this relevance, enzymes play a key role in the pathogenesis and treatment of diseases [44]. The viral genomes primarily encode enzymes due to its limited size, which are thus drug targets for the development of antivirals [45].

The following chapter gives a detailed understanding of molecular mechanisms and kinetic evaluation of enzyme catalysis and inhibition. Furthermore, the enzyme class of proteases will be introduced in general and the viral 3C protease as protein target of this thesis in detail.

1.2.1 Enzymes as biological catalysts

Enzymes catalyze the conversion of one or more substrate into one or more products. Four digits EC numbers are used to classify enzymes based on the type catalyzed reaction [46]. The protein-based nature of enzymes enables the formation of a catalytically active environment, the active site, due to the variable composition of amino acids. The active site in general makes up only a small portion of the total protein structure. The non-active site or allosteric parts of an enzyme are also of relevance as they can mediate interactions with proteins, membranes or small molecules, which modulate the catalytic activity through the induction of conformational changes. Allosteric modulation can occur within metabolic reaction cascades as part of a negative feedback loop or as inhibition mechanism of drugs.

A catalytic reaction can be described as a sequence of subsequent steps by the following scheme:

$$E + S \rightleftharpoons ES \rightleftharpoons ES^{\ddagger} \rightleftharpoons EP \rightleftharpoons E + P$$

in which E denotes the free enzyme, S the substrate, ES the initial encounter complex, $\mathrm{ES}^{\ddagger}$ the transition state, P the product and EP the enzyme-product complex after catalysis. The first step of the enzymatic reaction is the formation of the initial encounter complex ES. The substrate is bound through hydrophobic contacts and the formation of vectorized electrostatic interactions, which are essential for a proper spatial arrangement in the complementarily shaped active site. Emil Fischer proposed the "lock and key" principle to illustrate this specific binding event [47]. This rigid model is in contrast to many observed characteristics and

was extended by Koshland, who suggested that enzymes exist as an ensemble of conformations [48]. The substrate induces formation (induced-fit) or binds to the catalytically active conformer (conformational selection).

The reversible formation of the initial encounter complex ES is characterized by an association rate constant k_{on} and a dissociation rate constant k_{off}:

$$E + S \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} ES$$

The formed equilibrium can be quantitatively described by an equilibrium dissociation constant K_d, which is termed K_s for the substrate, according to the law of mass action:

$$K_s = \frac{[E] \cdot [S]}{[ES]} = \frac{k_{off}}{k_{on}} \tag{1.1}$$

Equation 1.1 shows that the formation of ES has to be energetically favored. The value of the free binding energy ΔG for the formation of ES can be calculated by applying the Gibbs free energy equation to K_s as in equation 1.2, in which R is the ideal gas constant and T the absolute temperature in kelvin. Through this equation, the free energy of binding for the formation of any enzyme small molecule complex can be calculated, which is valuable for the quantitative and also qualitative evaluation of enzyme inhibitors.

$$\Delta G_{ES} = -RT \cdot ln \left[\frac{1}{K_s} \right] \tag{1.2}$$

During catalysis, the substrate undergoes a multitude of small chemical transformations and conformational rearrangements. These are described by the rate constant k_{cat} according to the following scheme:

$$E + S \rightleftharpoons ES \xrightarrow{k_{cat}} E + P$$

As k_{cat} refers to all occurring enzyme substrate complexes formed during catalysis, it is a composite rate constant. In practice, the major energetic barrier that needs to be overcome along the reaction path is the transition state ($ES^{\ddagger}$) according to the transition state theory [49]. Therefore, k_{cat} can be approximated by the first-order rate constant for the formation of $ES^{\ddagger}$ from ES:

$$E + S \rightleftharpoons ES \xrightarrow{k_{cat}} ES^{\ddagger}$$

The ability of enzymes to accelerate a chemical reaction and thus to have high values of k_{cat} is based on their ability to reduce ΔG by stabilizing the transition state $ES^{\ddagger}$ [50]. This is accomplished by applying four strategies: (i) Approximation, (ii) covalent catalysis, (iii) acid/base catalysis and (iv) conformational distortion. The architecture of the enzyme active site leads to the proper spatial positioning of the substrate during the formation of the ES, which facilitates subsequent chemical transformations through interactions with a second substrate, a cofactor or enzyme residues. This so-called approximation effect significantly increases the probability of a successful encounter in comparison to diffusion mediated molecule collision. Some enzymes engage in a covalent complex with their substrate, which mimics the transition state of the catalyzed reaction. Thereby, the necessary energy for the formation of $ES^{\ddagger}$ is compensated by the covalent bond. Enzyme active sites contain protic amino acid side chains, which can engage in proton transfers and act as acid or base. This induces polarization of atomic bonds in the substrate and stabilizes charged intermediates formed along the catalytic reaction. The last strategy is based on the structural dynamics of enzymes. The substrate is bound in an energetically less favored conformation. Upon binding, the enzyme strives to an energetically more favored conformation complimentary to the transition state. This induces strain and distortion in the substrate, which facilitates conversion of the substrate toward the transition state.

These mechanisms can be exploited to inhibit enzymes through binding of small molecules, which mimic the transition state or are complementarily shaped to the active site. The increased reactivity of catalytic residues of enzymes pursuing covalent catalysis, such as proteases, can be used to covalently inactive an enzyme, which provides advantageous target binding characteristics and will be described in more detail in section 1.3.

1.2.2 Peptide cleaving enzymes

Proteases or peptidases catalyze the hydrolytic cleavage of peptide bonds within proteins and peptides. Since these are involved in essentially all kind of biological process, proteases, which regulate their activity, are associated with or a multitude of pathological conditions [51–54]. Therefore, proteases are important targets for approved drugs as listed in Table 1.3 and novel therapeutics [55, 56].

The size of protease substrates can vary significantly from small peptides to huge polyproteins expressed during the viral replication from a single ORF (subsection 1.1.1). For most proteases, the substrate recognition is independent of the

Table 1.3: Overview of drugs targeting proteases based on the clinical indication.

Indication	Drug	Target	Protease type
Anticoagulation	Rivaroxaban	Factor Xa	Serine protease
Bacterial infections	Cilastatin	Dehydropeptidase	Serine protease
Cancer	Bortezomib	Proteasome	Threonine protease
Type-I diabetes	Vildagliptin	Dipeptidyl peptidase-4	Serine protease
Diarrhea	Racecadotril	Enkephalinase	Metallo protease
Heart failure	Sacubitril	Neprilysin	Metallo protease
Hypertension	Ramipril	ACE	Metallo protease
Viral infections	Darunavir	HIV-Protease	Aspartic protease

size of the substrate but depends on the recognition of specific amino acid patterns within the peptide chain [57]. The active site of a protease comprises several pockets, which host the amino acid side chains upon substrate binding. This interaction can be described in a conventional scheme proposed by Berger and Schechter, which allows a uniformed denotation of the active site pockets and substrate residues binding to them (Figure 1.4) [58]. Starting from the scissile peptide bond, the first residue at the N-terminal site is named P1 and the pocket, which binds the respective side chain, S1. The following residue is termed P2 and binds to the S2 pocket consequently. For the C-terminal end site it is done accordingly but in opposite direction and with the addition of the ″ ′ ″ sign. The second residue involved in the scissile bond is therefore named P1′ and binds the S1′ pocket. The peptide backbone binds in an extensive conformation through the formation of hydrogen bonds in a β-sheet like interaction pattern [59]. The structure and arrangement of these pockets within the active site are key determents for the specificity of a protease. The S1 pocket is hereby of exceptional importance since proper binding of the P1 residue is essential for the necessary spatial positioning of the scissile bond in respect to the catalytic residues [57].

Proteases can be classified according to the position of the cleft peptide bond within the substrate peptide chain [60]. Three main groups are differentiated: (i) Exopeptidases, (ii) endopeptidases and (iii) omega peptidases. Exopeptidases process peptides and proteins at the N-(aminopeptidases) or the C-terminus (carboxypeptidases). Endopeptidases, which make up for most of the known proteases, cleave their substrate within the peptide chain marked by a specific recognition sequence. Omega peptidases cleave peptide bonds that have not been formed between the α-amino and an α-carboxyl group, which are termed isopeptide bonds.

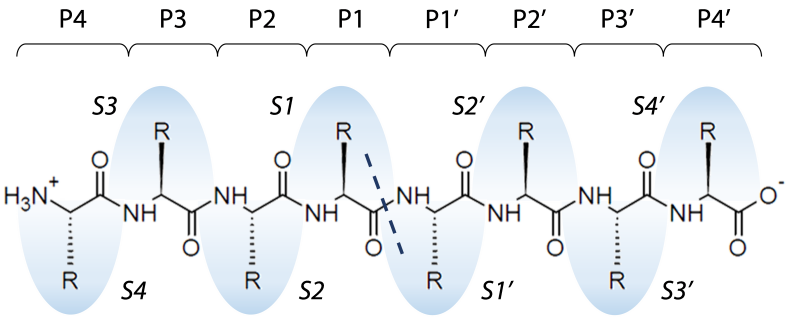

Figure 1.4: Protease active site and substrate nomenclature according to Schechter. [57]

An alternative, but more common way to classify proteases is based on the mechanism of catalyzes and the involved residues. Thereby, two main groups are differentiated. The first group of proteases, comprising serine, threonine and cysteine proteases, forms a covalent acyl-enzyme intermediate during catalysis, whereas the second group, which includes metallo-, aspartic and glutamic peptidases, does not. The active sites of the proteases proceeding via an acyl-enzyme intermediate share similar structural features: (i) a catalytic triad and (ii) an oxyanion hole. The catalytic triad refers to a specific arrangement of three amino acids within the active site, which determines the catalytic activity of the proteases. It generally comprises the residues serine (or cysteine), histidine and aspartate within a linear arrangement, which enables the formation of hydrogen bonds [61]. In serine proteases, the hydroxyl group containing side chain forms a hydrogen bond to the imine nitrogen of the imidazole in the histidine side chain, which in turn engages in a second hydrogen bond to the carboxyl moiety of the aspartate (Figure 1.5 I) [62]. The ability of the imidazole to act as hydrogen bond donor and acceptor as well as its $\mathrm{pK_a}$ around the neutral pH range are essential for the catalytic mechanism [63]. In cysteine proteases, the catalytic residues cysteine and histidine often exist as a preformed ion-pair due to the higher acidity of the cysteine thiol moiety (Figure 1.6 I) [64].

The substrate binds to the active site mediated by specific sub-pocket interactions with its scissile peptide bond positioned in close proximity to the catalytic serine (figure 1.5 II) or cysteine, respectively (Figure 1.6 II). Subsequently, the hydroxyl moiety of serine or the cysteine thiolate engages in a nucleophilic attack at substrate peptide bond mediated by the histidine, which acts as a base either during the covalent bond formation or before leading to the formation of a tetrahedral, negatively charged transition state intermediate (Figure 1.5 III and 1.6 III).

The oxyanion hole is a small pocket adjacent to the catalytic residues, in which

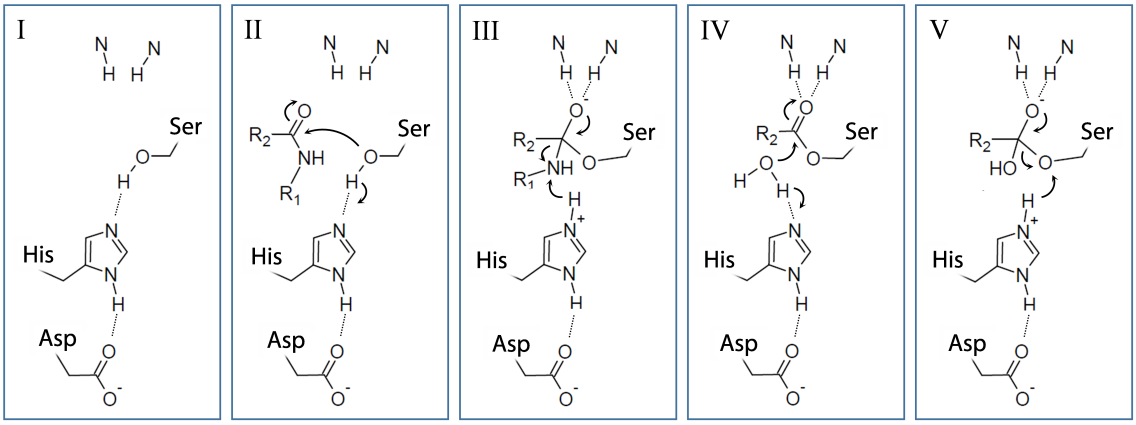

Figure 1.5: Schematic depiction of the essential steps within the catalytic reaction of serine proteases.

the formed oxyanion is bound and stabilized through the formation of hydrogen bonds with backbone nitrogen of the pocket forming residues [63]. In addition, the oxyanion hole induces conformational strain on the planar peptide bond, which forces it toward the better stabilized tetrahedral conformation. In the next

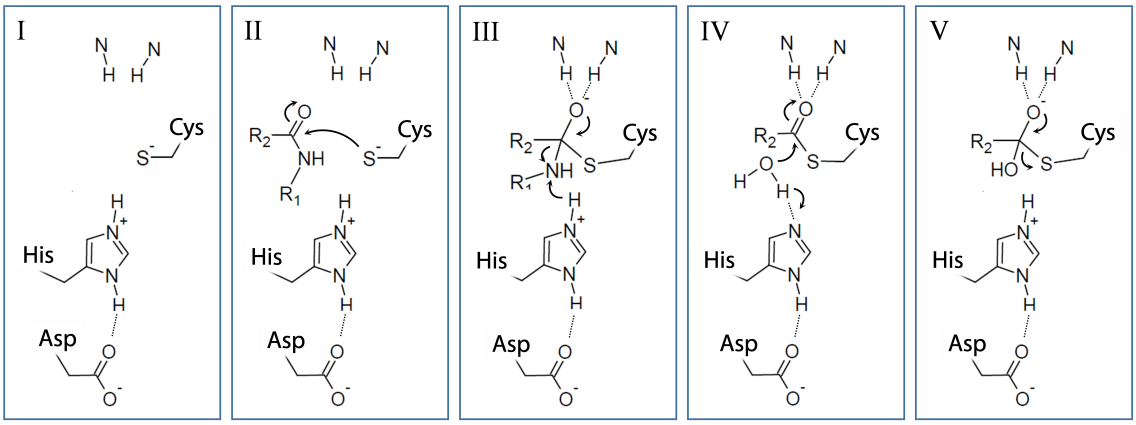

Figure 1.6: Schematic depiction of the essential steps within the catalytic reaction of cysteine proteases.

step, the histidine imidazolium ion acts as an acid to promote the release of the C-terminal amine of the substrate leading to the formation of an acyl-enzyme intermediate, an ester for serine proteases (Figure 1.5 IV) or a thioester in cysteine proteases (Figure 1.6 IV), respectively.

For the cleavage of the acyl-enzyme intermediate the histidine imidazole acts again as hydrogen bond acceptor for a water molecule. The formed tetrahedral intermediate is stabilized as described above and the release of the N-terminal part of the substrate is mediated trough proton donation from the imidazolium ion (Figure 1.5 V, 1.6 V). The third amino acid in the catalytic triad, the aspartic acid,

ensures the correct orientation of the histidine side chain tautomer and stabilizes the imidazolium ion [65]. Exchange of either one of the amino acids within the catalytic triad for alanine leads to a significant reduction in catalytic activity [60]. But it is not abolished, which illustrates the relevance of the overall architecture of the active site and the oxyanion hole for transitions state stabilization [60].

The detailed understanding of the catalytic mechanism of proteases is key for the successful design of inhibitors by exploiting essential features of the catalysis like the formation of covalent intermediates or transition-state stabilization by the oxyanion hole.

1.2.3 Viral 3C protease

Picornain 3C (EC: "3.4.22.28"), or 3C protease ($3C^{pro}$), is a cysteine-dependent endopeptidase, which is found in the different genera within the family of picornaviridae [60]. Its main function is the processing of the viral polyprotein (see section 1.1.1). The overall structure, composed of two topologically equivalent six stranded β-barrels (figure 1.7A), is similar to those of the trypsin-like serine proteases, which led to the suggestion of the existence of a common evolutionary precursor [66]. In analogy, the active site, accommodating the catalytic residues C147, H40 and E71, is positioned in a cleft between the two β-barrels (Figure 1.7B), which are preceded by a 14 amino acid long α-helix at the N-terminus and a mostly unstructured stretch of 20 amino acids connecting both β-barrels opposite to the active site (Figure 1.7A) [67]. Substrates are cleaved after glutamine, especially between Gln-Gly pairs but also Gln-Ser and Gln-Ala. The high specificity for glutamine at the P1 position is a distinctive feature of the $3C^{pro}$ and results from the ability of glutamines terminal amide moiety to form three hydrogen bonds with H161 and T142 within the S1 pocket (Figure 1.7B). Apart from this, the preferences for the respective residues can vary for different 3C proteases. In general, at the P2 and P3 position residues with lipophilic side chains are preferred. Whereas the P4 position is quite specific for alanine [68]. The primed site is less structured and thus of little relevance for substrate binding.

During maturation, $3C^{pro}$ cleaves the viral polyprotein at specific sites, marked by glutamine containing recognition sequences at unstructured regions within the enteroviral polyprotein as depicted in Figure 1.8. In order to facilitate binding of the huge viral polyprotein, the $3C^{pro}$ active site is wide and hosts only few pockets, which accommodate the key substrate residue side chains. The cleavage leads to the subsequent release of functional proteins but also of intermediate viral protein

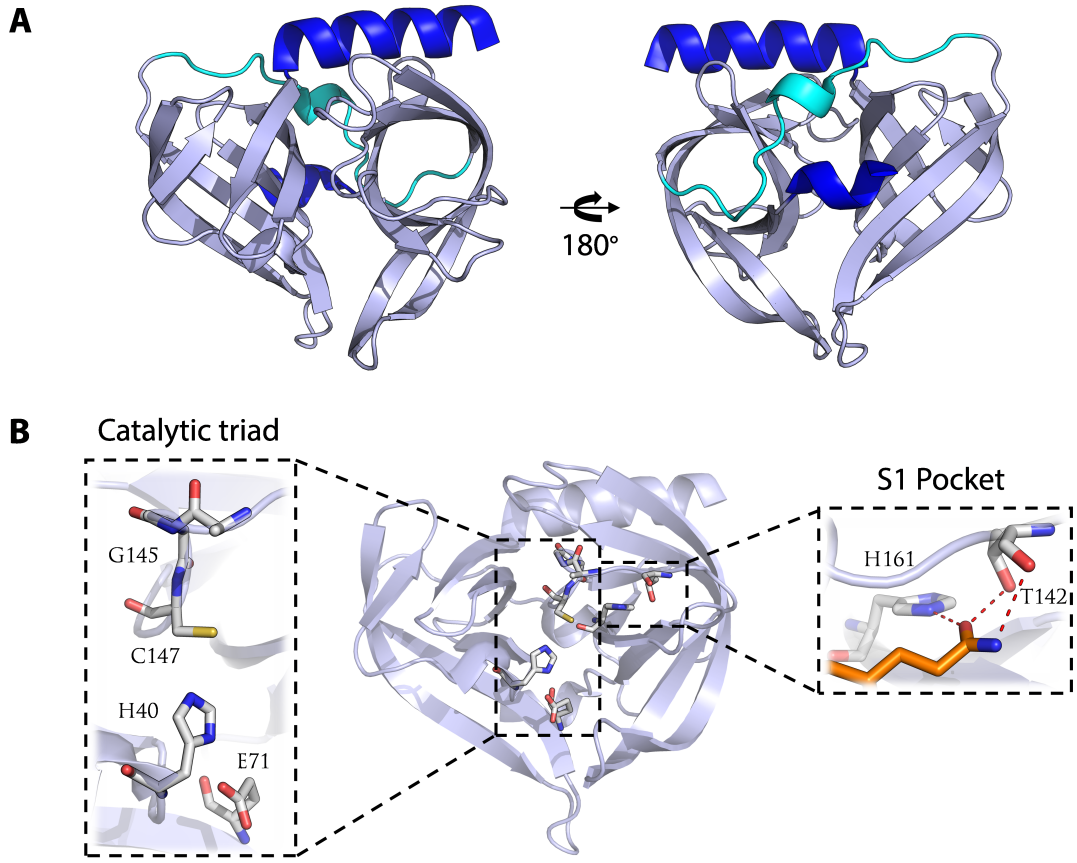

Figure 1.7: 3C Protease structural characteristics. (A) Cartoon representation of the $3C^{pro}$ with β-sheets colored in light blue, helical regions in dark blue and unstructured regions in cyan. (B) $3C^{pro}$ active site with a close up on the catalytic triad (left) and the S1 pocket (right).

conjugates, like the 3CD protein, which is composed of the $3C^{pro}$ and the RdRp 3D (section 1.1.2). This $3C^{pro}$ precursor not only participates in the polyprotein processing but also plays a major role in the viral RNA replication. The proteolytic activity of the $3C^{pro}$ enables the cleavage of several host cell proteins, which promotes hijacking of the cellular machinery and abrogates immune responses [69]. The cleavage of transcription factors like TBP or CREB within the nucleus by 3CD hampers host cell RNA synthesis [25]. Cap-mediated translation is shut down through the cleavage of eukaryotic translation initiation factor 4G (eIF4G) [70]. 3C proteases from different picornaviral genera take part in evasion from host immune response by cleavage of TRIF, RIG-1, MAVS or NEMO [71, 72].

The $3C^{pro}$ precursor 3CD plays also an essential role in the RNA replication of enteroviruses in which it binds to secondary structural elements in the 5'-UTR of the RNA [73]. On the one hand, it interacts with the *cis*-acting replication ele-

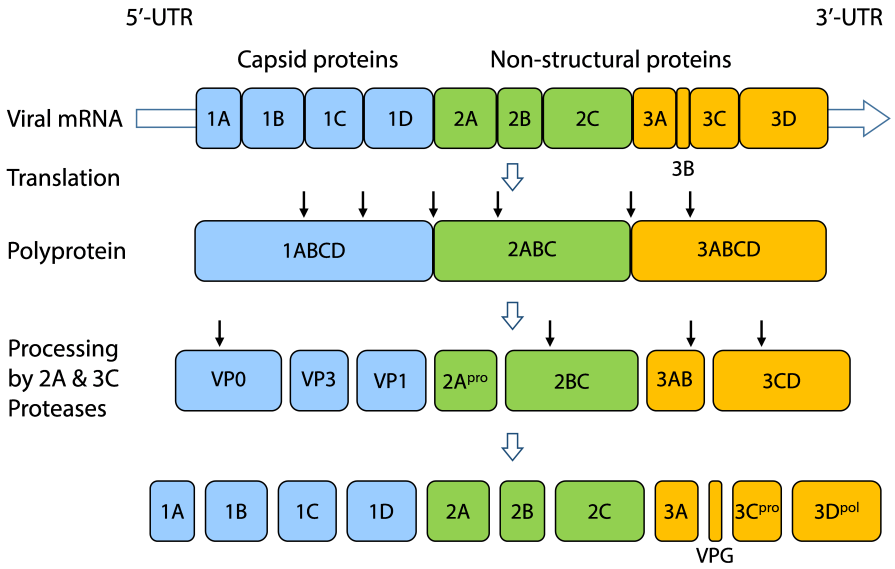

Figure 1.8: Polyprotein processing by viral proteases. Processing of the enterovirus polyprotein by proteases 2A, 3C and 3CD. Black arrows indicate cleaving sites. Sequential cleavage leads to the formation of functional viral precursor proteins.

ment (Cre) around which a replication complex is formed that generates the protein primer for replication through uridylation of the VPg protein [74]. On the other hand, 3CD binds to the stem loop D (SLD) of the cloverleaf structure (Figure 1.9A,B), which enables cyclization of the RNA through the interaction of the 3CD with the poly(A)-binding protein (PAbP) bound to the poly(A) tail at the 3'-end [75, 76]. The interaction of the 3CD with the SLD takes place opposite to the catalytic site of the $3C^{pro}$ part and involves a series of highly conserved residues, the KFRDI motif (Figure 1.9C) [77].

Due to the significance of the 3C protease for the replication and pathogeneses of the picornaviridae, it was indicated as a potential target for the therapy of enterovirus infections [78]. Inhibitors of the human rhinovirus (HRV) $3C^{pro}$ have first been developed by Agouron Pharmaceuticals, now Pfizer, to treat the common cold [79]. Peptide aldehyde inhibitors were designed based on the structure of the substrate. These show intramolecular cyclization to a hemiaminal due to the reaction of the P1 glutamine side chain amide moiety with the aldehyde, which was circumvented by converting the amide into a γ-lactam (Figure 1.10A) [80]. Further optimization led to a series of tetrapeptides, which mimic the natural substrate recognition sequence and in addition contain a Michael acceptor moiety for the covalent interaction with the catalytic cysteine of the $3C^{pro}$ (Figure 1.10B) [81–86]. The compound AG7088, termed rupintrivir, showed promising *in vitro*

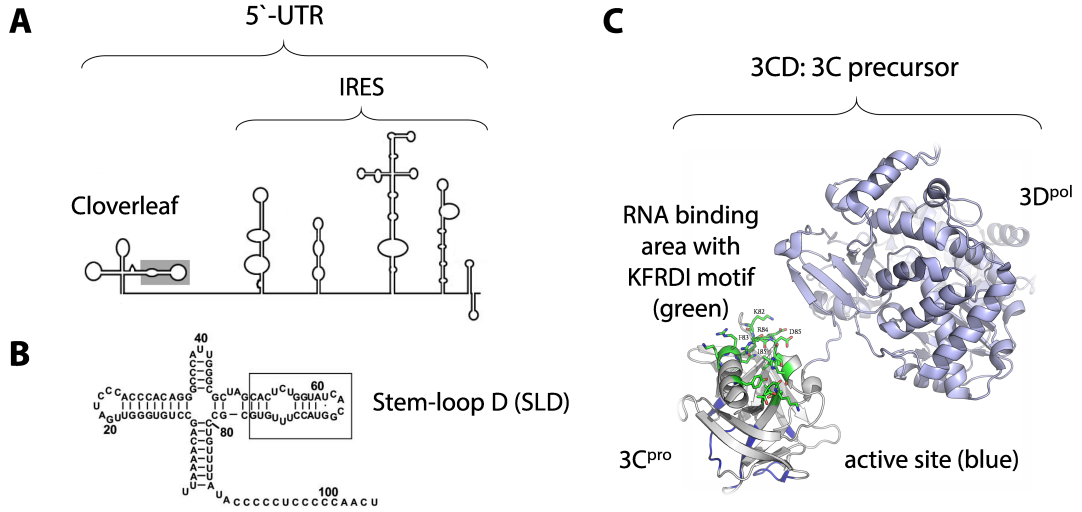

Figure 1.9: 3CD RNA binding characteristics. (A) Schematic representation of the secondary structures within the 5'-UTR of the enterovirus RNA. (B) Close up on the cloverleaf structure with the framed SLD adapted from Ohlenschlager *et al.* [75]. (C) 3CD protein structure with key residues involved in RNA binding colored in green and active site residue backbone colored in blue.

data and was therefore evaluated in clinical studies up to phase IIb [87]. However, development was haltered as primary endpoints were not met [88]. These compounds show unfavorable pharmacokinetic properties due to their peptidic nature [89]. In addition, the development of anti-rhinovirus agents is challenging as the success in treatment depends of the precise timing of drug administration during infection because of its acute progress [41]. The inevitable occurrence of resistances is further raised by the high incidence rates of HRV infections.

Despite this setback, $3C^{pro}$ is still considered a valuable target for the development of pan-enterovirus inhibitors. This is supported by the ability of rupintrivir to inhibit multiple enteroviruses *in vitro* and to prevent organ pathogenesis in animal models [90–93]. Furthermore, emerging enteroviruses are devoid of the challenging rhinovirus infection characteristics and are thus more susceptible to antiviral treatment. Therefore, a large variety of peptidic $3C^{pro}$ inhibitors has been developed to address these emerging enteroviruses based on the structure of rupintrivir and its analogues (Figure 1.10B). The reported variations include alteration of the Michael acceptor through the incorporation of different alcohols [94] or exchange to α-ketoamide warheads for inhibition of the the EV 71 $3C^{pro}$ [95]. Kawatkar *et al.* reported the incorporation of substituted prolines at the P2 position to optimize HRV $3C^{pro}$ inhibition [96]. Variations in the P3, 4 positions have been reported by Tan *et al.* as inhibitors for the EV D68 [97].

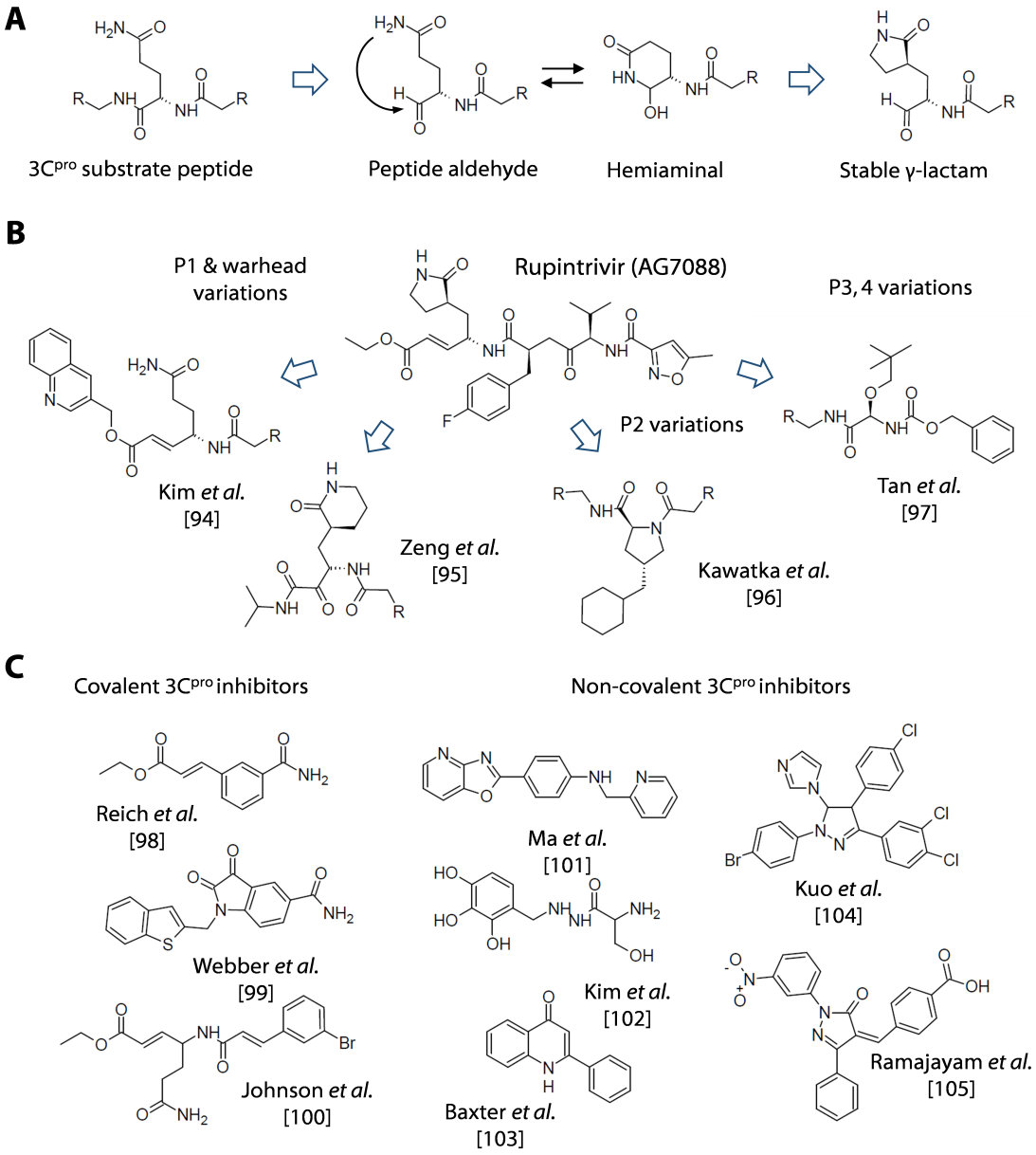

Figure 1.10: 3C^{pro} inhibitors reported in literature. (A) Starting point for the initial design of peptidic HRV 3C^{pro} inhibitors. (B) Chemical structure of rupintrivir and reported representative variations. (C) Chemical structures of reported non-peptidic 3C^{pro} inhibitors.

The identification of non-peptidic inhibitors has been a primary objective, due to the pharmacokinetic issues of the peptidic 3C^{pro} inhibitors. Several non-peptidic inhibitors have been reported in the last decade, which includes covalent as well as non-covalent 3C^{pro} inhibitors (Figure 1.10C) [98–105]. An issue of non-peptidic 3C^{pro} inhibitors is the often unknown binding mode and the questionable specificity, which is especially the case for the natural-product based compounds excluded in this overview.

1.2.4 Enzyme kinetic evaluations

For the design of enzyme inhibitors, a qualitative and quantitative evaluation of the enzyme-inhibitor interactions is indispensable. The enzymatic reaction can be described by the conversion of a substrate (S) into a product (P). A quantification is based on the increase of product over time or the disappearance of the substrate. The rate of conversion or velocity (v) is described by the differential equation 1.3. The plot of this linear relationship yields a straight line with the slope equal to the velocity (Figure 1.11A).

$$v = \frac{-d[S]}{dt} = \frac{d[P]}{dt} = k_{\text{cat}} \cdot [ES] \tag{1.3}$$

This linearity is only given as long as the concentration of ES ([ES]) is constant. The product, formed during the progress of the enzymatic reaction, promotes the reverse reaction and the consumption of the substrate diminishes the formation of ES, which leads to a decrease of the velocity (Figure 1.11B). The condition in which ES is constant is termed steady-state and it is desired for enzyme kinetic analysis. The steady state velocity (v_i) is indicated by using the index 'i' as it is the initial velocity. This is achieved through the use of an excess of substrate. Thereby, it was observed that the velocity and the substrate concentration follow a saturable kinetic (Figure 1.12A). This can be explained by the rapid equilibrium model [106].

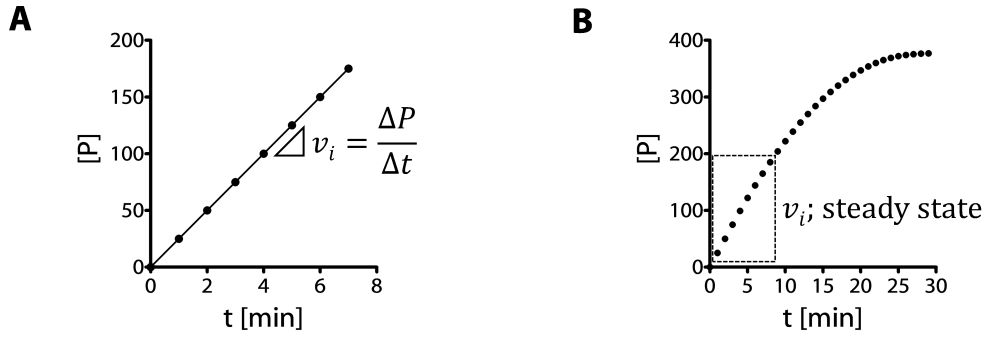

Figure 1.11: Progress curve of an enzymatic reaction. The formation of product as a function of time for short (A) and long assay duration (B).

The formation of ES is assumed to be faster than the subsequent catalysis. Thus, at higher substrate concentrations all enzyme molecules are occupied and no increase in velocity can be obtained. The maximum velocity at an infinite substrate concentration, termed V_{max}, only depends on the enzyme concentration ([E]) and

the substrate turnover given by the rate constant k_{cat} (see subsection 1.2.1).

$$V_{\text{max}} = [E] \cdot k_{\text{cat}} \tag{1.4}$$

For the mathematical description Michaelis and Menten proposed equation 1.5, which was subsequently optimized by Haldane and Briggs, through the introduction of another constant, $\mathrm{K_M}$ [50].

$$v = \frac{V_{\text{max}} \cdot [S]}{K_M + [S]} \tag{1.5}$$

$\mathrm{K_M}$, generally referred to as Michaelis constant, is not an equilibrium binding constant like $\mathrm{K_s}$, although it allows a similar estimation of the stability of the ES complex and has units of molarity. For substrate concentrations ([S]) equal to $\mathrm{K_M}$, equation 1.5 can be transformed to equation 1.6 and subsequently reduced to equation 1.7, which shows that $\mathrm{K_M}$ is the substrate concentration at half maximum velocity and thus of half maximal enzyme saturation, which can be directly linked to the stability of ES.

$$v = \frac{V_{\text{max}} \cdot [S]}{[S] + [S]} \tag{1.6}$$

$$v = \frac{V_{\text{max}}}{2} \tag{1.7}$$

The determination of $\mathrm{K_M}$ can be performed through non-linear regression of equation 1.5 or graphically through the use of its double reciprocal form. The obtained equation 1.8 is linear for the inverse velocity as a function of the inverse substrate concentration with the slope of $\mathrm{K_M/V_{max}}$ and the y-intercept being $1/\mathrm{V_{max}}$. This kind of plot is referred to as Lineweaver-Burk plot (Figure 1.12B) [107].

$$\frac{1}{v} = \left(\frac{K_M}{V_{\text{max}}}\frac{1}{[S]}\right) + \frac{1}{V_{\text{max}}} \tag{1.8}$$

$\mathrm{K_M}$ and $\mathrm{k_{cat}}$ are the key kinetics values for the description of the catalytic reaction. The ratio of $\mathrm{k_{cat}}$ to $\mathrm{K_M}$ describes how efficient an enzyme converts a certain substrate and is termed catalytic efficiency. As $\mathrm{k_{cat}}$ and $\mathrm{K_M}$ are related to specific steps in the catalytic pathway, changes in their values due to alterations of the experimental set up (pH, ion strength, small molecules) can be linked to either affecting the formation of ES or its conversion to $\mathrm{ES^{\ddagger}}$. These values are of great relevance for determining the mode of action of an enzyme inhibitor. The activity

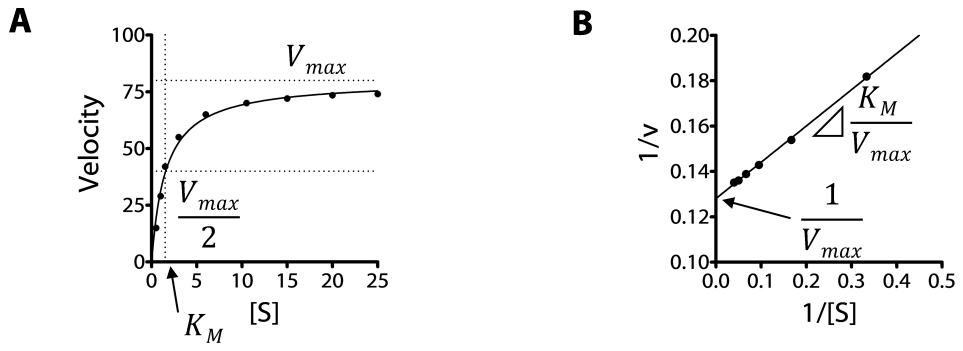

Figure 1.12: Enzyme substrate reaction kinetics. (A) Plot of the enzyme velocity as a function of substrate concentration. (B) Double reciprocal plot of the enzyme velocity as a function of substrate concentration.

of an enzyme can be modulated through the binding of a small molecule to the active site or an allosteric site but the kinetic description depends on the step of the catalysis, which is thereby affected. The possibilities are schematically depicted in Figure 1.13A. Most inhibitors bind exclusively to the free enzyme leading to the

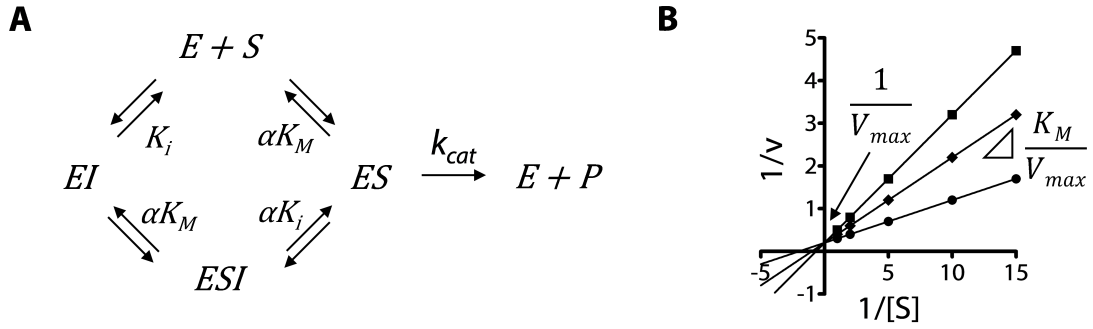

Figure 1.13: Enzyme inhibition modalities. (A) Schematic overview of the different possibilities of an inhibitor to interact in the catalytic pathway. (B) Lineweaver-Burk plot of competitive inhibitors.

formation of an enzyme inhibitor complex EI. This can be quantitatively described by an equilibrium dissociation constant, termed K_i. The binding in the active site of the free enzyme is in competition to the formation of ES, and these inhibitors are therefore termed competitive inhibitors. The opposite type of inhibitors bind only to ES leading to the formation of an enzyme substrate inhibitor complex ESI. These are termed uncompetitive inhibitors. The third type of inhibitors bind to the free enzyme as well as to the enzyme substrate complex and are referred to as noncompetitive inhibitors [50]. Competitive inhibitors in general block the formation of ES through the occupation of the substrate binding area. This affects the value of K_M and depends on the extent of the active site occupation by the in-

hibitor, which correlates with the inhibitor concentration [I]. Equation 1.9, which is an updated version of equation 1.5, can be utilized to determine the enzyme velocity in the presence of a competitive inhibitor.

$$v = \frac{V_{\text{max}} \cdot [S]}{[S] + K_M \left(1 + \frac{[I]}{K_i}\right)} \tag{1.9}$$

A double reciprocal (Lineweaver-Burk) plot of equation 1.9 is depicted in Figure 1.13B. It shows that the lines, referring to different concentrations of a competitive inhibitor, have variable slopes but intersect at the y-axis. This illustrates the theoretical considerations of $\mathrm{K_M}$ alteration by a competitive inhibitor but not of $\mathrm{V_{max}}$. The binding of noncompetitive inhibitors is characterized by two equilibrium dissociation constants: $\mathrm{K_i}$ for the formation of EI and $\alpha\mathrm{K_i}$ for the formation of ESI. α describes the preference of the inhibitor for either E or ES. A value of unity for α implies equal affinity for E and ES, which is considered as true noncompetitive inhibition and $\alpha \neq 0$ is termed mixed-type inhibition. $\mathrm{K_M}$ as well as $\mathrm{V_{max}}$ are altered due to the binding to E and ES, which can be expressed through equation 1.10.

$$v = \frac{\left(\frac{V_{\text{max}}}{\left(1+\frac{[I]}{\alpha K_i}\right)}\right)[S]}{[S] + K_M \left[\frac{\left(1+\frac{[I]}{\alpha K_i}\right)}{\left(1+\frac{[I]}{K_i}\right)}\right]} \tag{1.10}$$

This can be simplified by multiplication with the term $(1 + [I]/\alpha K_i)$ leading to equation 1.11.

$$v = \frac{V_{\text{max}} \cdot [S]}{[S]\left(1 + \frac{[I]}{\alpha K_i}\right) + K_M \left(1 + \frac{[I]}{K_i}\right)} \tag{1.11}$$

For noncompetitive inhbitors with $\alpha = 1$, equation 1.11 reduces to

$$v = \frac{V_{\text{max}} \cdot [S]}{([S] + K_M)\left(1 + \frac{[I]}{K_i}\right)}. \tag{1.12}$$

The lines of different concentrations of a noncompetitive inhibitor in the Lineweaver-Burk plot intersect at the x-axis and intercept with the y-axis at different values, which reflects the alteration of $\mathrm{V_{max}}$ and $\mathrm{K_M}$ (Figure 1.14A). The binding of uncompetitive inhibitors is solely quantified by the equilibrium dissociation constant $\alpha\mathrm{K_i}$ for the binding to ES. This affects $\mathrm{k_{cat}}$ and thus $\mathrm{V_{max}}$. But also $\mathrm{K_M}$ as it not only refers to the stability of ES but also of ESI. This is reflected by equation 1.13 for

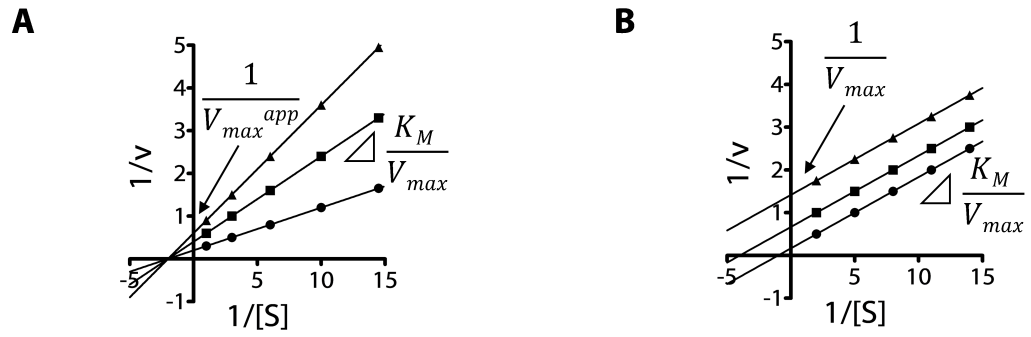

Figure 1.14: Lineweaver-Burk plots of noncompetitive (A) and uncompetitive inhibitors (B).

the calculation of the enzyme velocity in the presence an uncompetitive inhibitor,

$$v = \frac{\frac{V_{\max}}{\left(1+\frac{[I]}{\alpha K_i}\right)}[S]}{[S] + \frac{K_M}{\left(1+\frac{[I]}{\alpha K_i}\right)}}, \tag{1.13}$$

which can be simplified by multiplication with the term $(1 + [I]/\alpha K_i)$.

$$v = \frac{V_{\max} \cdot [S]}{[S] \cdot \left(1 + \frac{[I]}{\alpha K_i}\right) + K_M} \tag{1.14}$$

A Lineweaver-Burk plot shows parallel lines for different concentrations due to the mutual effect on $\mathrm{V_{max}}$ and $\mathrm{K_M}$ (Figure 1.14).

More reliable graphical tools for determining the mode of action are Dixon- and Cornish-Bowden-plots [108, 109]. In the Dixon plot v^{-1} is plotted as a function of [I], whereas in the Cornish-Bowden plot the ratio of [S] to [v] is plotted as a function of [I]. The combination of both yields a pair of plots, which is characteristic for each type of enzyme inhibition due to the specific perturbation of the kinetic parameters. Thus, inhibition modalities can be determined as well as $\mathrm{K_i}$ values, which are equal to the x-value at which the lines intersect. The Dixon- and Cornish-Bowden plots for competitive, for noncompetitive and for uncompetitive inhibitors are depicted in figure 1.15A-C.

The so far described kinetic analysis primarily provide a qualitative assessment of the interaction between enzyme and inhibitor. These analyses are only done for inhibitors which provide sufficient activity justifying these more extensive investigations. Independent from the binding modality, enzyme inhibition by an inhibitor leads to a reduced substrate turnover, which results in a diminished enzyme velocity $\mathrm{v_i}$ according to equation 1.3. This can be set in relation

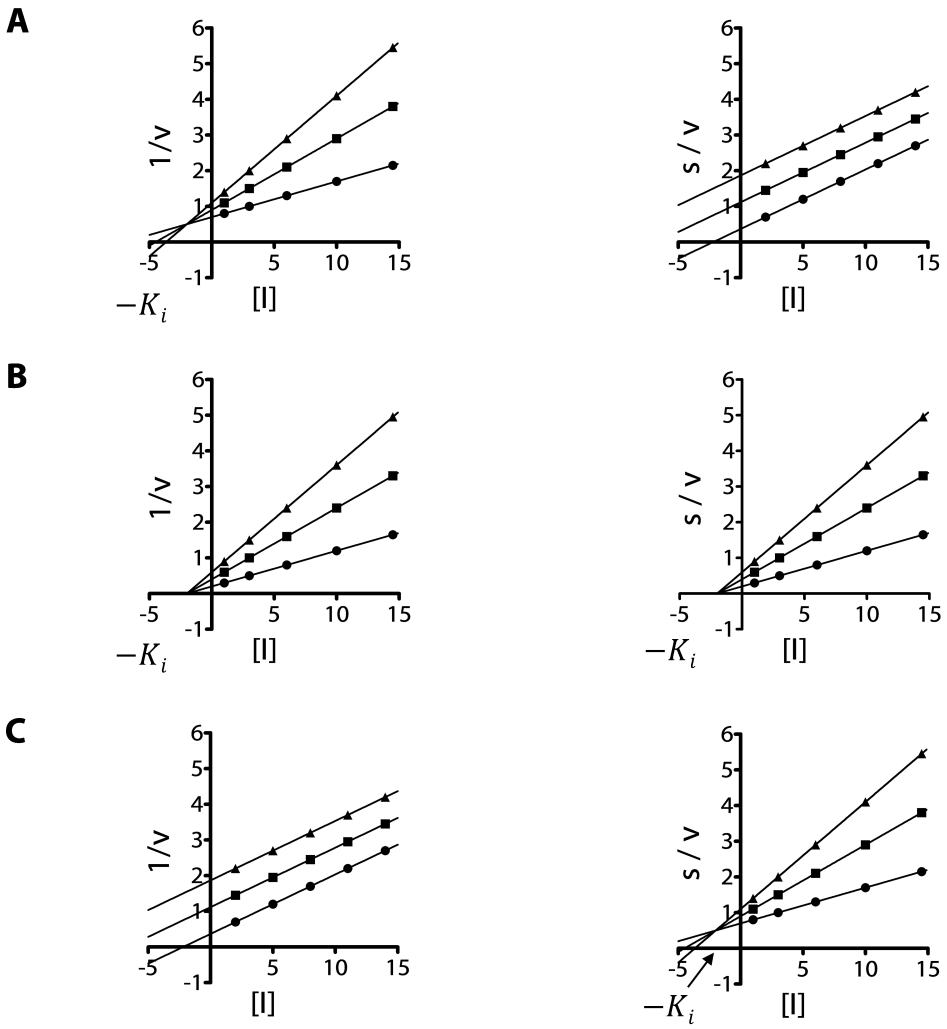

Figure 1.15: Dixon and Cornish-Bowden plots of enzyme inhibitors. Dixon (left) and Cornish-Bowden (right) plots for competitive (A), noncompetitive (B) and uncompetitive (C) enzyme inhibition.

to the uninhibited enzyme velocity v_0, which yields the fractional or remaining enzyme activity at the respective inhibitor concentration. This is performed for a wide range of inhibitor concentrations in concentration- or dose-response experiments. Representative plots for the enzyme activity in the presence of an inhibitor at different concentrations is given in Figure 1.16 in a linear (A) and a logarithmic concentration scale (B). The semi-log plot shows a sigmoidal curvature with plateaus at high and low inhibitor concentrations, which is typical for a well-behaved enzyme-inhibitor interaction and exemplifies the equilibrium-based nature of binding. The midpoint of the sigmoidal curve represents the inhibitor concentration, which leads to an enzyme activity of 50 % or likewise 50 % enyzme inhibition and is referred to as IC_{50}. It can be determined through non-linear fit

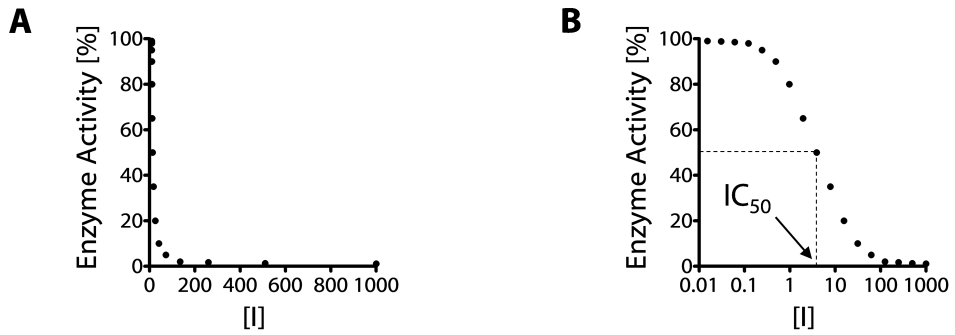

Figure 1.16: Enzyme inhibition plots. Plot of the enzyme activity as a function of the inhibitor concentration in linear (A) and logarithmic concentration scale (B).

to equation 1.15.

$$\text{Enzyme Activity} = \frac{v_i}{v_0} = \frac{1}{\left(1 + \frac{[I]}{IC_{50}}\right)} \tag{1.15}$$

The depicted sigmoidal dose-response curve illustrates an ideal enzyme-inhibitor interaction referring to a binding stoichiometry of $1:1$. However, there are situations, in which this is not the case and thus equation 1.15 needs to be adjusted to equation 1.16 to account for these cases.

$$\text{Enyzme Activity} = \frac{v_i}{v_0} = \frac{1}{\left(1 + \frac{[I]}{IC_{50}}\right)^h} \tag{1.16}$$

This is done through the introduction of the term h, which is referred to as Hill coefficient or Hill slope as it describes the slope of the sigmoidal dose-response curve. Non-stoichiometric binding is known for some cases but greater deviation of the Hill coefficient from unity suggest an unspecific binding event. The Hill slope is therefore of great valuable for the evaluation of enzyme inhibitor interactions.

The IC_{50} is the most common value to report the potency of an enzyme inhibitor. But it only refers to the half-maximum enzyme inhibition under the conditions it was determined. Thus the IC_{50} can vary significantly upon perturbation of the experimental conditions such as the substrate concentration. Depending on the modality of action, an increase in substrate concentration can promote (uncompetitive inhibitors), hamper enzyme binding (competitive inhibitors) or go either way for uncompetitive inhibitors depending on the value of α. To enable comparison of the potency of inhibitors, derived under different conditions, it is necessary to use the $\mathrm{K_i}$. The equations of Cheng and Prusoff listed below provide

a convenient way to transform IC_{50} into K_i values based on the experimentally utilized substrate concentration and the K_M [110]. The ratio of K_M to [S] refers to the above stated influence of the inhibition modality. In the case of noncompetitive inhibitors with $\alpha = 1$ the K_i equals the IC_{50}.

Competitive Inhibition:

$$IC_{50} = \frac{K_i}{\left(1 + \frac{[S]}{K_M}\right)} \tag{1.17}$$

Noncompetitive Inhibition:

$$IC_{50} = \frac{[S] + K_M}{\left(\frac{K_M}{K_i} + \frac{[S]}{\alpha K_i}\right)} \tag{1.18}$$

Uncompetitive Inhibition:

$$IC_{50} = \alpha K_i \left(1 + \frac{K_M}{[S]}\right). \tag{1.19}$$

The so far described enzyme inhibition is based on the assumption that the equilibrium between the free enzyme and the enzyme inhibitor complex is established rapidly and thus is constant over the time course of the kinetic analysis. In some cases, the equilibrium is slowly established over a longer period of time. The progress curve of the enzymatic reaction in the presence of a so called slow-binding inhibitor is linear during the initial phase, which is described by v_i. Over the course of the reaction, the enzyme velocity diminishes until an equilibrium has been established and the system is in a steady-state with a velocity v_s (Figure 1.17A). To determine the potency of a slow-binding inhibitor, the comparison of v_i and v_o will not provide a proper estimate as it varies overtime until the steady-state is reached, which can outlast the time course of the experimental measurement. The progress curve can be described by equation 1.20, which contains the pseudo-first order rate constant k_{obs}. It describes the curvature of the progress curve and thus the conversion of v_i to v_s.

$$[P] = v_s \cdot t + \left(\frac{v_i - v_s}{k_{\text{obs}}}\right)\left(1 - e^{(-k_{\text{obs}} \cdot t)}\right) \tag{1.20}$$

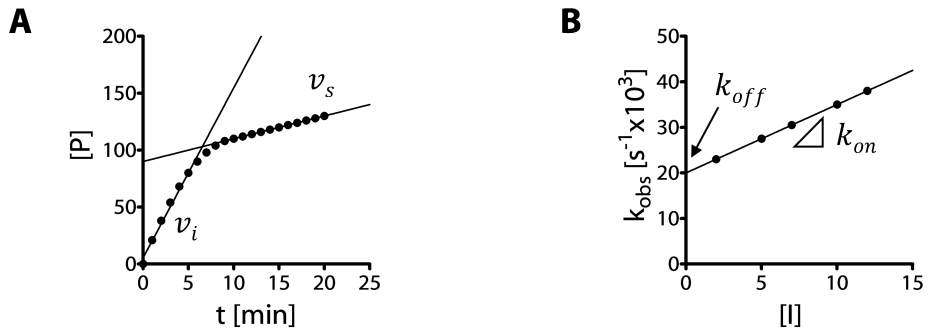

Figure 1.17: Enzyme inhibition by slow binding inhibitors. (A) Plot of [P] as a function of time for an enzymatic reaction in the presence of a slow-binding inhibitor. (B) Plot of k_{obs} as a function of [I] for a slow-binding inhibitor.

One of the most common reason for slow binding is the formation of a covalent bond, which can be part of the catalytic reaction of enzymes (subsection 1.2.2) or the binding mechanism of an enzyme inhibitor (section 1.3). This is either reversible or irreversible. The reversible covalent binding is described by the rapid formation of the non-covalent enzyme inhibitor complex EI, which then converts slowly upon covalent bond formation to the covalent enzyme inhibitor complex E-I (Figure 1.18A). As the latter is in general the rate limiting step, the kinetic description can be reduced to the formation of E-I characterized by the rate constants k_{on} and k_{off} (Figure 1.18B). Thus a plot of k_{obs} as a function of [I] yields a linear relationship (Figure 1.17) quantified by equation 1.21. As the formed equilibrium

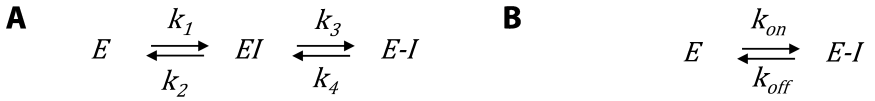

Figure 1.18: Reversible covalent enzyme inhibitor interaction scheme in general (A) and reduced (B).

only depends on k_{on} and k_{off}, the K_i of an reversible covalent enzyme inhibitor can be calculated from the ratio of k_{off} to k_{on} according to equation 1.22. This is an apparent value as k_{on} depends on the substrate concentration and the inhibition modality, which have to be considered (equation 1.23).

$$k_{obs} = k_{on} \cdot [I] + k_{off} \tag{1.21}$$

$$K_i = \frac{k_{off}}{k_{on}} \tag{1.22}$$

$$k_{\text{obs}} = \frac{k_{\text{on}} \cdot [I]}{\left(1 + \frac{[S]}{K_M}\right)} + k_{\text{off}} \tag{1.23}$$

Irreversible inhibition differs significantly from the so far described kinetics as no equilibrium is formed throughout the progress of the enzymatic reaction and the enzyme velocity after an infinite time is zero. A plot of the enzyme velocity as a function of time shows a hyperbola due to its continuous decline (Figure 1.20A). Equation 1.20 can thus be simplified to equation 1.24 for the description of the progress curve of an irreversible enzyme inhibitor.

$$[P] = \frac{v_i}{k_{\text{obs}}} \left(1 - e^{(-k_{\text{obs}} \cdot t)}\right) \tag{1.24}$$

The irreversible enzyme inhibition either takes place through a single- or a two-step mechanism as shown by the schemes in Figure 1.19. A plot of k_{obs} as a func-

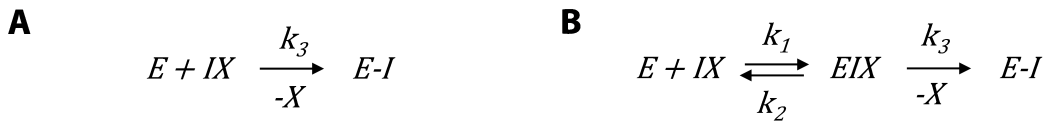

Figure 1.19: Irreversible covalent enzyme inhibitor interaction scheme for a single-step (A) and a two-step mechanism (B).

tion of [I] is linear for an inhibitor with to a single-step mechanism as its binding is depends only on k_3 (Figure 1.19). k_{off} is zero as there is no dissociation from the E-I and the line passes through the origin (Figure 1.20C). For an inhibitor with a two-step kinetic, the rate of covalent bond formation (k_3) is affected by the formation of the initial complex EI, described by the ratio of k_2 to k_1, similar to the enzyme substrate interaction in the Michaelis-Menten kinetics and described by equation 1.25. Thus a plot of k_{obs} as a function of [I] is linear at small inhibitor concentrations but shows saturable kinetics with a hyperbolic curvature due the rate limiting formation of EI (Figure 1.20B).

$$k_{\text{obs}} = \frac{k_{\text{inact}} \cdot [I]}{K_I + [I]} \tag{1.25}$$

k_{inact} refers to the maximum rate of enzyme inactivation at an infinite inhibitor concentration. K_I is in analogy to K_M the inhibitor concentration displaying half-maximum enzyme inactivation. Equation 1.25 can be transformed by division of

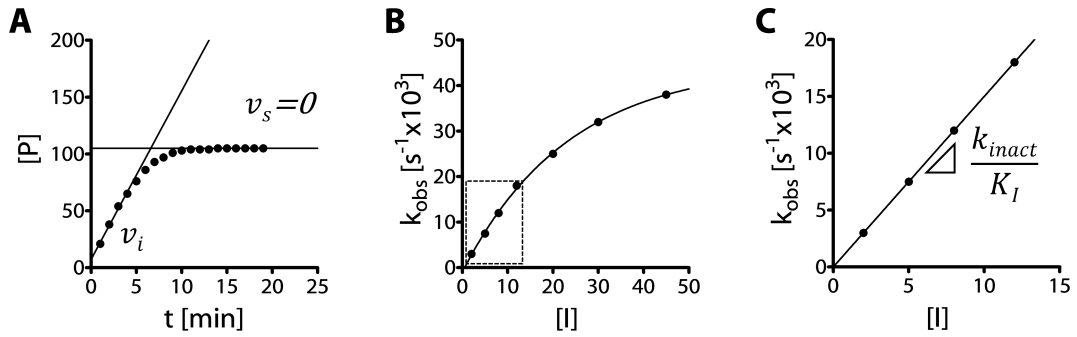

Figure 1.20: Irreversible enzyme inhibition. (A) Plot of [P] as a function of time for an enzymatic reaction in the presence of an irreversible inhibitor. Plot of k_{obs} as a function of k_I for an irreversible inhibitor proceeding via a two-step (B) or a single-step mechanism (C).

both sides by [I] leading to equation 1.26.

$$\frac{k_{\text{obs}}}{[I]} = \frac{k_{\text{inact}}}{K_I + [I]} \tag{1.26}$$

At very low inhibitor concentrations, in which $[I] << K_I$, equation 1.26 can be simplified to equation 1.27. These conditions represent the linear relationship between k_{obs} and [I] similar to the single-step irreversible inhibition. The slope of the linear function is therefore described by k_{inact}/K_I. This second-order rate constant describes best the enzyme inactivation rate and is most commonly used to report the potency of an irreversible enzyme inhibitor.

$$\frac{k_{\text{obs}}}{[I]} = \frac{k_{\text{inact}}}{K_I} \tag{1.27}$$

The here described mathematical and graphical methods are the foundation for the enzyme kinetic analysis performed in this thesis.

1.3 Covalent modifiers in drug design

The formation of a covalent bond provides unique features for the protein ligand interaction, which can be utilized throughout the drug design process to address challenging targets and to identify promising bioactive ligands. This chapter will provide an overview of the key characteristics of covalently binding ligands and the chemical moieties involved in covalent bond formation

1.3.1 Covalent protein-ligand interaction

Binding of a small molecule to a protein, like the substrate of an enzyme or a drug, is the result of shape and electrostatic complementarity, which favors the binding through the formation of specific electrostatic interactions and the displacement of water from the binding site. In addition, this can involve a covalent interaction. Also a great number of drugs covering a variety of indications pursue this mechanism as shown by Table 1.4. This includes many important drugs, like acetylsalicylic acid, which is unique among the nonsteroidal anti-inflammatory drugs (NSAID) due to its ability to block thrombocyte aggregation as a result of the covalent mode of action [111]. Furthermore, the β-lactam antibiotics [112], which are still the most important group of antibiotics, the proton-pump inhibitors like omeprazole or clopidogrel, a thienopyridine-class antiplatelet agent, which is the second best-selling drug [113], act through a covalent mechanism.

The significance and their distinctive binding characteristics most often result from the covalent bond formation, yet the majority of covalent drugs were discovered through phenotypic screenings and the true mode of action was elucidated hindsight. Furthermore, covalent ligands were disregarded by the pharmaceutical industry due to the toxicological risk of covalently binding to off-target proteins [114].

During the last decade, there has been a change of minds in the perception of covalently binding ligands, which is appropriately described by the title of a review by Singh *et al.* entitled "The resurgence of covalent drugs" [115]. By now the covalent bond is not only associated with toxicological risks but also with advantageous target binding characteristics [116]. These include: (i) Prolonged target interaction, (ii) increased binding affinity, (iii) pharmacokinetic and pharmacodynamic decoupling, (iv) improved selectivity, (v) stalled resistance formation. The formation of a covalent bind can either be reversible or irreversible (see subsection 1.2.4). In either case the target residence time is prolonged [117], even to

Table 1.4: Overview of covalent drugs, their indication and the respective target.

Indication	Drug	Target
Inflammation	Acetylsalicylic acid	Cyclooxygenase
Infection	Ampicillin	D-Ala-D-Ala transpeptidase
GERD	Omeprazol	H^+/K^+-ATPase
Cancer	Afatinib	Bruton's tyrosine kinase
Obesity	Orlistat	Triacylglycerol lipase
Anticoagulation	Clopidogrel	$P2Y_{12}$-Receptor
Epilepsy	Vigabatrin	GABA transaminase
BPH	Dutasteride	5α-Reductase
Parkinson's disease	Rasageline	Monoamine oxidase
Type-I diabetes	Vildagliptinee	Dipeptidyl peptidase-4

infinity for irreversible inhibitors. This leads to an increased duration of the biological effect, which can outlast the presence of the drug in the body. This leads to a decoupling of the pharmacokinetics from the pharmacodynamics, which makes covalent drugs less sensitive to elimination, protein binding or metabolism and enables longer dosing intervals. This is key as the potential for idiosyncratic drug reactions is depending on the total drug exposure of the patient [118]. Furthermore, the formation of a covalent bond provides additional binding energy, which exceeds typical non-covalent binding energy [119]. This enables addressing of challenging target proteins comprising interaction sites devoid of small pockets that are else necessary for ligand binding [120]. The non-equilibrium binding of irreversible inhibitors not only inactivates the targeted proteins until it has been resynthesized but is also beneficial to overcome high concentration of an endogenous ligand with which a non-covalent inhibitor would has to compete with [121]. Especially in the field of kinase inhibitors, covalent binding is utilized to address poorly conserved residues, typically cysteines, within the active site and thereby to improve selectivity over closely related kinases [122]. Furthermore, irreversible target inactivation delays the emergence of drug-resistance as no selection is occurring [123].

Table 1.4 shows that most targets of covalent drugs are enzymes. On the one hand, this due to the fact that the covalent bond formation is often facilitated by the inherent activity of the catalytic residues. On the other hand, the major goal of addressing an enzyme is the complete inactivation, for which covalent inhibitors are well suited. However, for other target classes, like ion-channels or G protein-coupled receptors, for which the shot-time activity modulation is essential, cova-

lent inactivation can be disadvantages [124].

Despite the above listed benefits of covalent binding, the inherent reactivity poses a major risk and challenge for the development of covalent drugs. Nonetheless, for anti-infective agents, covalent binding is an attractive strategy as enzyme are typically addressed with the aim of complete inactivation during a short time frame of administration, which limits the drug exposure and thus the possibilities for adverse drug reactions [114].

1.3.2 Warheads

The chemical substructure that engages in covalent bond formation with the respective target is referred to as warhead. A description and classification of the different warheads is done based on (i) reversibility of covalent bond formation, (ii) the proceeding chemical reaction or (iii) the mechanism of target inactivation. The latter is applied for enzyme inhibitors and will therefore be used for description. Following the suggestions of Kranz *et al.*, enzyme inhibitors are group into mechanism-based inhibitors, which are depending on the catalytic machinery of the enzyme, and affinity labels, which act independently [125]. Mechanism-based inhibitors are further subdivided: (i) dead-end inhibitors, (ii) transition state inhibitors and (iii) enzyme-activated suicide inhibitors. Dead-end inhibitors react with the enzyme according to the catalytic reaction but form a covalent adduct, which cannot react any further. The transition state inhibitors are a sub-group of the dead-end inhibitor class, which form a covalent transition state mimicking complex. The enzyme-activated suicide inhibitors are unreactive prior to the interaction with the enzyme, which catalytically transforms the inhibitor by introducing a reactive moiety.

Affinity labels interact with an enzyme independent from its catalytic mechanism. Albeit differences are made based on the characteristic to react specifically with the catalytic residues. These are termed quiescent affinity labels and show in general no affinity toward non-catalytic residues. In contrast, reactivity-based affinity labels interact with any kind of nucleophile.

For the rational development of covalent drugs, warheads are explicitly designed to be part of the molecule. The selection of suitable warheads not only depends on the mechanism of inhibition, which consequently excludes overly reactive affinity labels and enzyme activated inhibitors due to the complex and often hard to predict mode of action, but also on the addressed residue of the respective target [126]. The majority of covalent protein modifications take place via

the cysteine thiol moiety as also non-catalytic cysteines show sufficient reactivity. Serine modifications are also common due to the prevalence of serine in the catalytic triad of hydrolytic enzymes. To a lesser extent threonine, histidine, lysine and tyrosine residues are addressed [127]. The reaction of hydroxyl moieties with carbonyl compounds leads to the acylation of the hydroxyl group in the case of warheads, which contain a leaving group like esters, carbamates, urea or activated amides. This product is similar to the acyl enzyme intermediate found in the catalytic reaction described for serine and cysteine proteases. For carbonyl containing warheads without a leaving group, like ketones, the reaction leads to the formation of a hemiketal or hemicacetal in the case of aldehydes. These mimic the negatively charged tetrahedral catalytic reaction intermediate (Figure 1.5 in subsection 1.2.2). The activity of these warheads depends on the stabilization of the ketal form, which is fostered by the electron withdrawing effect of neighboring groups such as a trifluormethyl group or carbonyl moieties in α-position. Boronic acids are very specific warheads for catalytic serine or threonine residues, as the oxygen in the hydroxyl moiety can fill the empty boron p-orbital leading to the formation of a tetrahedral transition state analogue [128].

The chemical nature of the cysteine thiol moiety as a soft nucleophile promotes the formation of thioethers with warheads like epoxides, aziridines, acetylenes, Michael acceptors or substituted α-methyl ketones, which are therefore more specific for cysteine modification. Moreover, the thiol moiety can interact in analogy to the hydroxyl group with carbonyl compound to form thioesters or thiohemiketals, respectively [129].

Despite these common principles, the proceeding reaction is also heavily influenced by the active site and thus can vary for different enzymes of the same catalytic type. The structural core of a covalent inhibitor influences the reactivity of the warhead. First, through alterations of the electronic distribution. Second, the binding of the inhibitor is initially facilitated through non-covalent interactions leading to the initial complex EI as described in subsection 1.2.4 and depicted in the following scheme:

$$E \underset{k_2}{\overset{k_1}{\rightleftharpoons}} EI \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} E\text{-}I$$

The covalent bond is formed subsequently, which therefore depends on the spatial arrangement in the active site determined by the initial complex. Hence, for the

design of covalent inhibitors balancing of the contribution of the warhead and the core is crucial but also challenging.

1.4 Fragment-based drug design

During the last twenty years fragment-based approaches have risen to be mainstream for the development of novel drugs. Its success is exemplified by the large number compounds in clinical trials originating from fragment-based approaches of which two have been approved [130–132]. This section will introduce the origin and conceptual foundations, the practice as well as the application of computational methodologies in fragment-based approaches.

1.4.1 The origin of FBDD

Fragment-based design refers to the utilization of small molecules for the identification and subsequent development of bioactive ligands through the use of structural information on protein binding. The conceptual framework originates from the 80s and 90s. A key contribution came from Jencks, who illustrated that the free energy of binding is additive upon combination of two ligands [133]. The observations of solvent molecules bound to protein active sites in X-ray structures [134] fostered the idea to identify protein hot spots similar to computational sampling of the protein surface by probes [135]. A practical application was first conducted by Hajduk *et al.* with the famous "SAR by NMR" approach, in which weakly active ligands were linked guided by structural information derived from NMR to obtain bioactive inhibitors [136]. The focus on small molecules is encouraged by the complexity model proposed by Hann *et al.*, which suggests an inverse relationship for the probability of finding a hit and the complexity of molecules [137]. Analysis of the maximal binding affinity by Kuntz *et al.* showed that the affinity only increases linear with molecular size up to a heavy atom count of 15 and then levels off [138]. The screening at a small molecular size enables more efficient sampling of the chemical space in comparison to the conventional high throughput screening (HTS), in which compound collections comprising millions of drug-sized molecules are screened. Estimations of chemical space suggest that more than 10^{60} drug-like molecules could exist, of which HTS collections cover only a tiny fraction [139]. The chemical space of fragment-sized molecules is significantly smaller so that it can be better represented by libraries of a few thousand fragments. Integral to fragment-based approaches is the reliance on structural information about the binding of fragment hits due to the limited affinity of the fragment but also for the rational optimization. This includes X-ray crystallography and nuclear magnetic resonance (NMR)-based methodologies [140].

In order to evaluate the attractiveness of merely active fragment hits, several metrics have been introduced. The most widely used one is the ligand efficiency (LE), which normalizes the free energy of binding (equation 1.2 in subsection 1.2.2) to the number of heavy atoms (equation 1.28). This can be simplified to equation 1.29 and subsequently to equation 1.30 [141]. In general LE values of 0.3 or above are considered attractive as this is the typical increase in binding affinity during optimization [142] and is in accordance with the evaluation on the maximal binding affinity by Kuntz *et al.* [138].

$$LE = -\frac{\Delta G}{HA} \tag{1.28}$$

$$LE = -\frac{2.3 \cdot RT \cdot pK_i}{HA} \tag{1.29}$$

$$LE = -\frac{1.4 \cdot pK_i}{HA} \tag{1.30}$$

The hallmark characteristics of fragment-based drug design are (i) more efficient chemical space sampling, (ii) higher probability of hit finding, (iii) smaller compound libraries, (iv) enthalpy-driven binding with improved pharmacokinetics and (v) rational optimization guided by structural information.

1.4.2 The practice of FBDD

Fragment-based approaches generally consist of the three following parts: (i) Fragment library design, (ii) fragment screening and (iii) hit optimization.

The proper composition of the fragment library is a key determinant for the success of the later steps due to the limited number of contained fragments [143]. These are pooled from larger collections through *in silico* filtering based on their physicochemical properties, like the rule of three [144], tailoring toward the utilized screening methods, the addressed target, potential synthetic follow up and structural diversity. Substructures associated with undesired properties such as assay interference, instability or toxicity are removed [145].

Two aspects make fragment screenings to significantly differ from HTS. The low binding affinities in the high μM to mM range, which require more sensitive detection methods and the fragment concentrations promoting artificial binding. The primary applied screening methods are surface plasmon resonance (SPR),

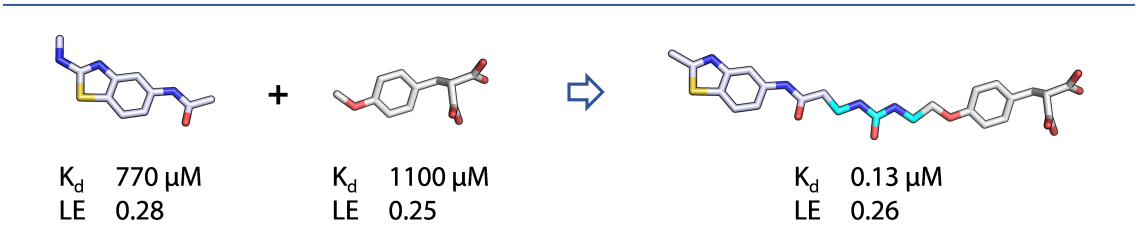

Figure 1.21: Literature-derived example for hit optimization through fragment linking. [157]

X-ray crystallography, ligand- and protein-observed NMR, functional assays and to a lesser extend isothermal titration calorimetry (ITC), mass spectrometry (MS) and virtual screening (VS) [146]. An advantage of VS is the high sampling power, which enables it to cover a great portion of the fragment space [147]. The simplicity of fragments pose a significant challenge especially for docking tools, which is the primarily applied VS technique in fragment screening, as these fail to prioritize fragments, due to the few interactions formed [148]. Therefore, it has been suggested to include interaction-based evaluation metrics such as interaction fingerprints or 3D pharmacophores [149–151].

Hits obtained from primary screenings are evaluated using a second method, an orthogonal assay, to verify activity. Only a portion of the primary hits are confirmed, due to artificial activity but also due to the differences in representation of the biological system by the respective method [152]. The last step in the primary fragment screening involves the determination of structural information, primarily through X-ray crystallography, to guide the subsequent optimization [153]. In the absence of crystal structures, qualitative information about the fragment binding can also be retrieved by NMR or structure-based modeling [154].

The optimization of confirmed fragment hits is pursued through tight integration of chemical transformations with biological and structural evaluation [155]. Three different strategies are pursued: (i) Fragment linking, (ii) fragment merging and (iii) fragment growing. In fragment linking two molecules, which bind adjacently in the active site of a target, are connected by a linker to obtain over additive binding energy as proposed by Jencks [133]. The success of this prototypical fragment optimization strategy depends heavily on the applied linker and its impact on the binding conformation of the initial fragments [156]. A successful fragment linking approach for the design of lactate dehydrogenase inhibitors is given in Figure 1.21 [157]. The structural features of fragments with overlapping binding poses can be combined in one molecule with improved affinity as it shows the interactions of the fragments it originates from. This merging is not restricted to

fragments as structural information of bioactive molecules from other sources can be integrated. Figure 1.22 depicts a fragment merging conducted for the design of lipoprotein-associated phospholipase A_2 inhibitors [158]. In the case of frag-

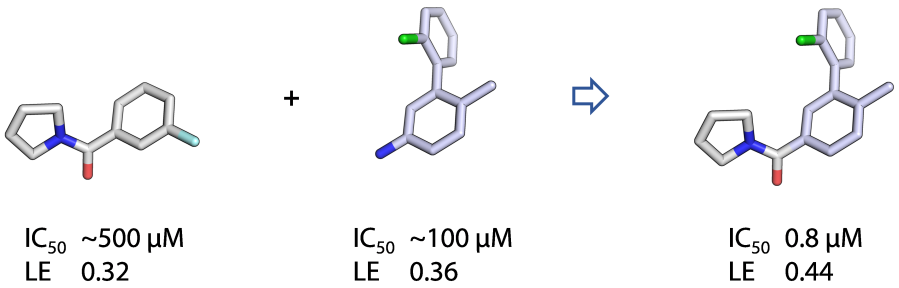

Figure 1.22: Literature-derived example for hit optimization through fragment merging. [158]

ments binding to the same part of the protein, none of the former strategies can be applied. These fragments are synthetically expanded to occupy pockets or to engage in additional interactions in the active site. The growing is depending on structural information of the fragment and computational sampling of putative synthetic analogues to determine promising vectors for growing in the active site. An example for a fragment growing of a fragment hit against Dengue virus RdRp is given in Figure 1.23 [159].

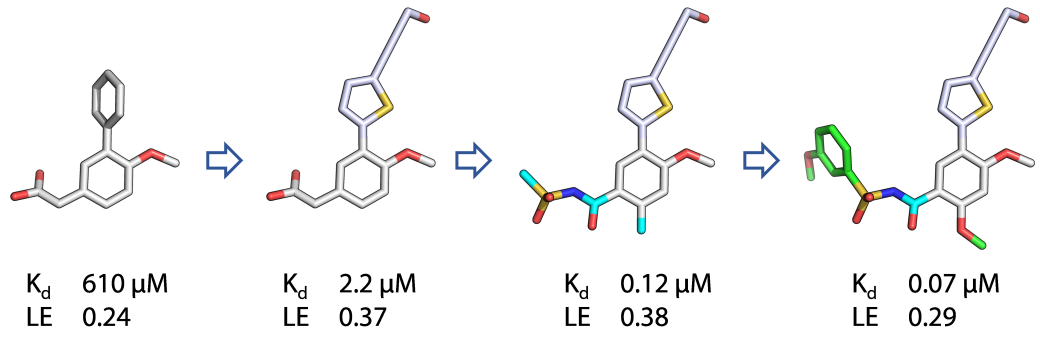

Figure 1.23: Literature-derived example for hit optimization through fragment growing. [159]

The focus on small, weakly active but highly efficient molecules as starting points, primarily identified through biophysical methods , makes fragment-based approaches complementary to conventional HTS. Thereby, FBDD enables the identification of novel chemical structures and allows to address challenging targets for which HTS has failed [160].

Chapter 2

Aim of the Thesis

Aim & Objectives

Emerging enteroviruses, like CVB3, EV D68 or EV71, pose a significant threat for the human health with no treatment available to date. The viral 3C protease plays a pivotal role in enterovirus replication by processing the viral polyprotein as well as host cell proteins, which significantly contributes to viral pathogenesis. Therefore, the 3C protease has been addressed as target for the design of anti-enterovirus agents with limited success as the most potent inhibitors suffer from poor pharmacokinetic properties due to their peptidic nature.

The aim of this thesis is the design of novel, non-peptidic, covalent 3C protease inhibitors through a fragment-based approach. Identification and rational optimization of fragments enable access to uncharted chemical space and thus the identification of novel chemical entities as 3C proteases inhibitors, which lack the pharmacokinetic issues of known peptidic $3C^{pro}$ inhibitors. As starting points, covalently binding fragments were aspired as covalent bond formation is essential for binding in the large and shallow active site of the $3C^{pro}$. On the other hand, covalent binding provides advantages binding kinetics. Furthermore, it enables fragment hit verification through mass spectrometry as well as detailed binding mode analysis in the absence of NMR or X-ray crystallography. The pursue of a fragment-based approach was fostered through a close in-house collaboration with the working group of Prof. Rademann, which provided access to

(i) a tangible library of small molecules as source for fragments,

(ii) a biochemical assay for evaluation of fragment activity against the $3C^{pro}$,

(iii) chemical synthesis capacities for fragment hit optimization.

Computational methods shall be employed to perform specific steps in the fragment-based design approach:

(i) Fragment library design using substructure (2D) and property (1D) filtering,

(ii) virtual screening-based fragment pre-sampling,

(iii) fragment hit exploration through *in silico* analogue and similarity searches,

(iv) *de novo* design-driven fragment optimization.

The close integration of biological and computational data shall drive and enable the rational fragment-based design of $3C^{pro}$ inhibitors. The pursued approach and the objectives to be reached within can be subdivided into three main steps: Fragment screening, fragment hit exploration and fragment hit optimization.

(i) For the initial fragment screening, the structural characteristics of the $3C^{pro}$ active site are elucidated and transferred into a 3D pharmacophore model, which is used to screen a fragment library assembled from the in-house small molecule collection. Virtual hits are subsequently evaluated for biological activity in the enzyme kinetic assay.

(ii) The $3C^{pro}$ binding characteristics of putative fragment hits are biochemically and computationally investigated to generate a rational binding model. This serves as basis for the structural exploration of the fragment hits though searches in commercial fragment libraries to identify suitable starting points for optimization.

(iii) To drive fragment hit optimization, a *de novo* design workflow is developed, which allows exhaustive sampling of putative fragment hit analogues and their rational design based on structural models for $3C^{pro}$ binding. Promising virtually designed compounds are synthesized in collaboration with the working group of Prof. Rademann and biochemically evaluated.

The primary objective of this thesis is the design of novel 3C protease inhibitors, which provide novel possibilities for the development of anti-enterovirus agents and thus fortify available antiviral defenses. The developed structural models and binding mode investigations of the designed inhibitors will provide a more profound understanding of $3C^{pro}$ and thereby guide drug development approaches for this pivotal viral protein.

Chapter 3

Methods

3.1 Computational methods

3.1.1 3D Pharmacophores & virtual screening

The IUPAC defines a pharmacophore as: "*The ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response*" [161]. Therefore, a pharmacophore can be described as an ensemble of chemical features in a specific spatial arrangement, which represents the interactions formed in a protein-ligand complex. This includes electrostatic interactions such as hydrogen bonds, coulomb interactions or π- stacking and entropic contributions resulting from desolvation of hydrophobic areas (Figure 3.1A-I). The shape of the protein active site is described through exclusion volumes spheres placed on the respective protein atoms (Figure 3.1J). The generalized representation allows to identify novel scaffolds as ligands with different chemical structures can comprise the essential functionalities necessary to match the pharmacophore. The abstraction of the complex protein-ligand interaction makes pharmacophore-based screening computationally very efficient [162]. Pharmacophores are either derived from protein-ligand complexes such as crystal structures (structure-based) or from a set of superimposed bioactive ligands (ligand-based) [151]. The most common pharmacophore modeling tools are Catalyst [163], LigandScout [164], MOE [165] and Phase [166]. They differ in terms of the feature definition, feature placement on ligands, the alignment of the ligands and the pharmacophores [167]. In this thesis, LigandScout version 4.09.2 was used for pharmacophore modelling as it provides the extraction of pharmacophore feature from crystal structures [164], an efficient algorithm for pattern-matching [168] and a feature representation for covalent protein-ligand interactions [169].

In virtual screening (VS), computational methods are applied to obtain a set of molecules with favorable characteristics from a large collection. This is referred to as enrichment. A typical example would be the use of a pharmacophore to identify compounds with biological activity to a given target. As for any given model, its performance needs to be verified prior to the application in a VS. This is termed validation of the pharmacophore and is an essential step as a pharmacophore derived from a protein-ligand complex is not a pharmacophore by the definition of IUPAC but a pattern of pharmacophore features representing protein-ligand interactions. In order to identify the essential features biologically active and inactive molecules are used. For the inactives, there are several challenges to be consid-

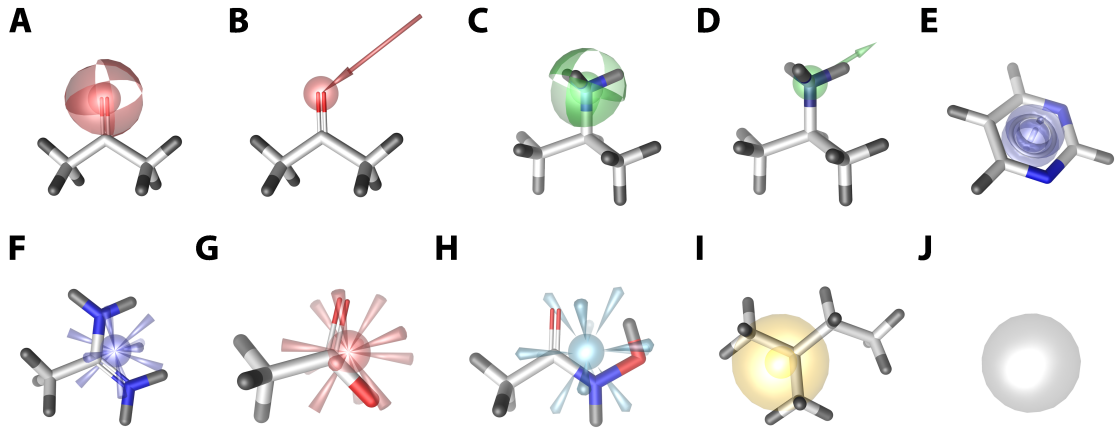

Figure 3.1: Pharmacophore feature representation in LigandScout. (A) Hydrogen bond acceptor as sphere; (B) hydrogen bond acceptor as vector; (C) hydrogen bond donor as sphere; (D) hydrogen bond acceptor as vector; (E) aromatic π-plane; (F) positive ionizable; (G) negative ionizable; (H) metal complexation (I) lipophilic contact; (J) exclusion volume.

ered. At first, data on true inactives is generally sparse. Secondly, the reason for inactivity is sometimes not due to the protein interaction and thus cannot be captured by the pharmacophore model. To overcome these limitations so-called decoy molecules are used [170]. These are compounds that are chemically similar to the actives but are probably not active themselves. As this has not been biologically verified, the use of an excess of decoys in comparison to the number of actives, typically in the ratio of $50:1$, ensures that by chance most of the decoys are true inactives [171]. Secondly, the large number of decoys allows to include a wide range of diverse chemical structures.

Several metrics can be used to evaluate the performance of a pharmacophore model on a given dataset comprising a number of N molecules, which are either biologically active or inactive. Those actives that are correctly identified by the model are termed true positives (TP) and those missed are referred to as false negatives (FN). In analogy, the inactives, identified correctly are termed true negatives (TN), whereas the wrongly identified inactives are referred to as false positives (FP). The sensitivity (Se) describes the ability of the model to identify the actives within the data set based on the ratio of the true actives to all active molecules (equation 3.1).

$$Se = \frac{TP}{TP + FN} \tag{3.1}$$

The specificity (Sp), calculated according to equation 3.2, denotes the ability to reject inactives molecules.

$$Sp = \frac{TN}{TN + FP} \tag{3.2}$$

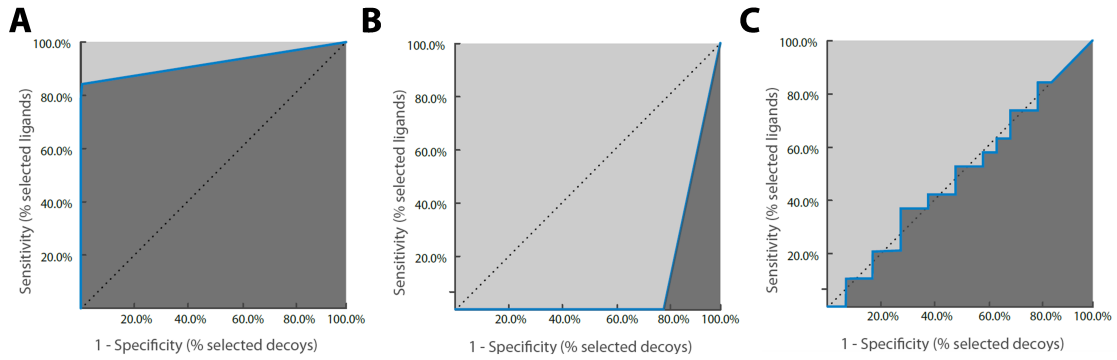

Figure 3.2: Representative ROC curves generated by LigandScout. Performances of a good (A), a poor (B) and a model, which performs equally well as random selection.

The yield of actives (Ya) is the ratio of the true actives to all selected molecules (n) (equation 3.3).

$$Ya = \frac{TP}{n} \tag{3.3}$$

The enrichment factor (EF) describes the enrichment of true actives in the selected set in comparison to a random selection. This is the ratio of Ya to the ratio of all actives (A) to all molecules in the data set (N) (equation 3.4). The enrichment factor is of special use for VS methods like docking tools, which rank compounds according to a scoring scheme, to evaluate if the actives are prioritized.

$$EF = \frac{Ya}{A/N} \tag{3.4}$$

The Receiver Operating Characteristics (ROC) curve is a method for the graphical representation of the discriminative power of a pharmacophore model [172]. It plots the sensitivity as a function of the inverse specificity (1-Sp). A ROC curve is plotted based on a rank-ordered list of hits resulting from a VS, which is passed through from the top scored molecule by calculating the Se and Sp based on the so far encountered molecules. Therefore, the ROC curve for a near-ideal case would rise at the y-axis to a Se close to 100 % before it turns straight to the top right corner (Figure 3.2A). In a poor model, inactives are prioritized over actives and the curve would turn right before it goes up (Figure 3.2B). A more common characteristic of a poor model is its inability to enrich actives, which would results in a random-like selection with the ROC curve proceeding around the median (Figure 3.2C).

The ability of a molecule to match a 3D pharmacophore has at foremost a Boolean characteristic, which depends on the chemical nature of the ligand. In addition, the quality of the matching can be described based on the superposi-

tion of the ligand moieties and the individual features of the pharmacophore for a given conformation. In LigandScout, this is expressed by a so-called pharmacophore fit, which is calculated through equation 3.5 and 3.6. $\mathrm{S_{RMS}}$ is the matched feature pair root mean square (RMS) distance score, $\mathrm{RMS_{FP}}$ is the RMS of the feature pair distances, $\mathrm{S_{FCR}}$ is the feature count/RMS distance score, c is a weighting factor for the number of matched feature pairs with a default value of ten and $\mathrm{N_{MFP}}$ is the number of geometrically matched feature pairs.

$$S_{RMS} = 9 - 3 \cdot min(RMS_{FP}, 3) \tag{3.5}$$

$$S_{FCR} = c \cdot N_{MFP} + S_{RMS} \tag{3.6}$$

The pharmacophore fit is not only a useful descriptor in a 3D pharmacophore-based VS but also for the prioritization of docking poses.

3.1.2 Protein-ligand docking

Protein-ligand docking is one of the primarily applied computational methods in drug design. The main objective is the generation of a ligand binding conformation within the active site of a protein in order to retrospectively rationalize the binding mode for a set of bioactive ligands or to identify ligands with a putative biological activity in prospective VS. The docking tools, of which AutoDock [173], ParaDocks [174], DOCK [175], FlexX [176], GLIDE [177] and GOLD [178] are the most widely used ones, generate binding poses through the iterative search for ligand conformations and evaluation of the generated poses. The conformation search is done either through (i) a systematic, (ii) random or (iii) dynamic approach [179]. Systematic sampling aims for exhaustive search within the conformational space. To cope with the combinatorial explosion of possible conformation, parts of the ligand are incrementally sampled while the rest of the molecule is kept rigid. Random or stochastic methods, which include Monte Carlo searches and genetic algorithms, apply random changes to conformations or population of conformations. Dynamic searches use molecular dynamics to sample the conformational space of the ligand. The evaluation, generally referred to as scoring, is done through scoring functions, which are either (i) force field-based, (ii) empirical or (iii) knowledge-based [180]. Force field-based scoring uses an energy function describing the atomic interactions derived from experimental data and *ab initio* quantum mechanical calculations. Empirical scoring use an energy function based on weighted energy terms, which are parametrized through repro-

duction of crystal structure ligand conformations. Knowledge-based scoring uses energy potentials between interacting atoms, which are derived from favorable interactions in protein-ligand complexes and thus tries to reproduce these in the proposed binding conformation. Consensus scoring combines the different approaches with the aim to balance out their shortcomings. Despite the undertaken efforts for optimization of scoring functions, performance varies significantly for different proteins-ligand complexes and the general correlation with ligand binding affinity is poor [181]. Therefore, an alternative to the native scoring functions of docking tools is the evaluation of the generated docking poses in respect to specific interactions, which have been identified as crucial for target binding, by 3D pharmacophores or protein-ligand interaction fingerprints (PLIFS) [150]. In this thesis, the docking tool GOLD v 5.0.2 by the Cambridge Crystallographic Data Center (Cambridge, UK, Europe), which uses a genetic algorithm for pose generation, was used as it performs comparably well, offers different scoring functions and allows for covalent docking, which was mandatory for the pursued approach [182]. Covalent docking is performed in GOLD through the definition of a linkage atom present in both, the protein and the ligand, which therefore has to be transformed into the covalently bound state prior to the docking. A small molecule substructure defines the moiety of the ligand covalently bound to the protein. During the docking run, this substructure is placed upon the covalent bond forming atoms of protein and ligand. This way, the covalent bond is simulated while the distances and angles are held constant throughout the docking.

3.1.3 Shape-based virtual screening

Shaped-based approaches are based on the idea of complementarity between the protein active site and a ligand binding to it. Virtual screenings for molecules similar in shape and electrostatics to known biologically active ligands have led to the identification of inhibitors for a multitude of relevant targets [183–185]. The gold standard tool for shape-based screening is ROCS (Rapid Overlay of Chemical Structures) [186], which defines the molecular volume of a ligand through a Gaussian function. Chemical functionalities such as hydrogen bond donor/acceptor capacities or ionized groups are described by a so-called color-score. This simplicity makes shape-based screenings computationally very efficient. Furthermore, the independence from structural information enables VS in the absence of crystal structures.

3.1.4 Molecular dynamics simulations

Molecular dynamics (MD) simulations have become a popular method to investigate proteins or protein-ligand complexes in respect to the inherent conformational flexibility of proteins [187]. Hereby, the atoms within a biomolecular system are treated as particles connected through springs based on Newtonian mechanics and the law of motion [188]. The inter- and intramolecular interactions between the atoms are described by an additive potential energy function consisting of simple energy terms for the covalent and non-covalent interaction of the atoms in terms of bond distances, angles and electrostatics [189]. The parameters used in the potential energy function are derived from quantum mechanical calculations. Together they form the biomolecular force field used by an MD tool for the simulation. The most common force fields are general AMBER force field (GAFF) [190], CHARMM [191], GROMOS [192] and OPSL-AA [193]. During an MD simulation, the force field is used to derive coordinates for the individual atoms trough integration of Newton's laws of motion for a predefined time period. This yields a MD trajectory, which contains the positions of all particles throughout the MD simulation. The statistical distribution of atoms can be used to investigate dynamic process such as the binding of a ligand or conformational changes in the protein structure and are thus valuable to identify putative binding modes of a ligand or the structural changes that underlines protein functions.

3.1.5 Chemical structures & transformations

The typical way to computationally represent a chemical structure is the use of a graph, which consists of nodes connected by edges. The nodes refer to the atoms, whereas the edges represent the covalent bonds between them. The most common molecular file format is the Molecular Design mol format [194]. Linear notations are an alternative, more simple and thus more compact way to represent chemical structures. The SMILES notation is the most popular one [195]. Hereby, atoms and their connections within a molecule are described by alphanumeric characters in a linear fashion by going through the molecule from one end to the other. Atom types are represented by the respective atomic symbol. Single bonds are implicated, whereas a double bond is indicated by '=' and a triple bond by '#'.

The SMARTS notation is an extension to SMILES, which has been developed for substructure searches [196]. It can be used to assemble focused molecule sets based on knowledge about key moieties that have to be present in the ligands. Fur-

thermore, it can also be used to identify and exclude molecules containing undesirable functionalities from the screening library. The advantage of the SMARTS annotation is that it allows logic operators to define the environment of the described substructure.

A further extension is the SMIRKS language, which is designed to describe chemical transformation of molecules [196]. Hereby, two chemical structures described in SMILES format, extended through SMARTS-based environment definitions, are connected by a '$\gg$'. Mapping the corresponding atoms at either side indicates the transformation to be made, whereas unmapped atoms will not be transformed. This can be used to transform single molecules but also to simulate a chemical reaction between two or more molecules, and therefore represents an annotation for *de novo* design of molecules *in silico*.

3.2 Experimental methods

3.2.1 Enzyme kinetic assays

For evaluation of enzyme inhibition by small molecule biochemical assays are used, which are based on the catalytic reaction of the enzyme Hereby, a substrate, often artificially modified, is used to monitor its consumption in the presence of an inhibitor [197]. Biochemical assays can be differentiated based on the detected signal and the time frame of readout [198]. This can be either (i) continuous, (ii) discontinuous or (iii) at an endpoint. A continuous assay can be advantageous as the progress curve of the enzymatic reaction is monitored, which is exceptionally useful in the context of time-dependent inhibition as faced in this thesis. In dead-end and discontinuous assays, the enzymatic reaction is quenched and analyzed for the whole mixture or an aliquot. This can be necessary for enzymatic reactions, which can not be monitored directly and is advantageous as the monitoring and required instrumentation time is reduced [50]. The appropriate readout depends on the target enzyme and the catalyzed reaction. Fluoresce-based methods are among the most common in protease assays [199]. Thereby three different types can be differentiated: (i) Chemically quenched dyes, (ii) fluorescence resonance energy transfer (FRET) and (iii) fluorescence polarization. Chemically quenched dyes are linked to the P1 residue and get released through the catalytic reaction [200]. FRET-based assays make use of the Förster resonance energy transfer, which describes the energy transfer between two light-sensitive molecules (fluorescent dyes) without the emission of light [201]. For this, the emission wavelength of one dye must overlap with the excitation wavelength of the other one and bot dyes must be in close spatial proximity (< 10 Å). This is achieved by placing the dyes at either end of a substrate peptide chain. Upon cleavage, the energy transfer can no longer take place. In this thesis, a FRET system was utilized, in which the fluorescent dye EDANS (5-((2-aminoethyl)amino)naphthalene-1-sulfonic acid) is connected to the C-terminus of the peptide KTLEALFQGPPVYE-NH_2 and the quencher DABCYL (4-(4- dimethylaminophenylazo)benzoyl) is bound to the N-terminus (Figure 3.3). In fluorescence polarization the ratio of polarized to non-polarized fluorescent light emitted by a fluorescent dye is measured. The fluorescent dye is linked to a large substrate. Upon cleavage the polarization is reduced due to the higher rotational flexibility of the substrate fragments.

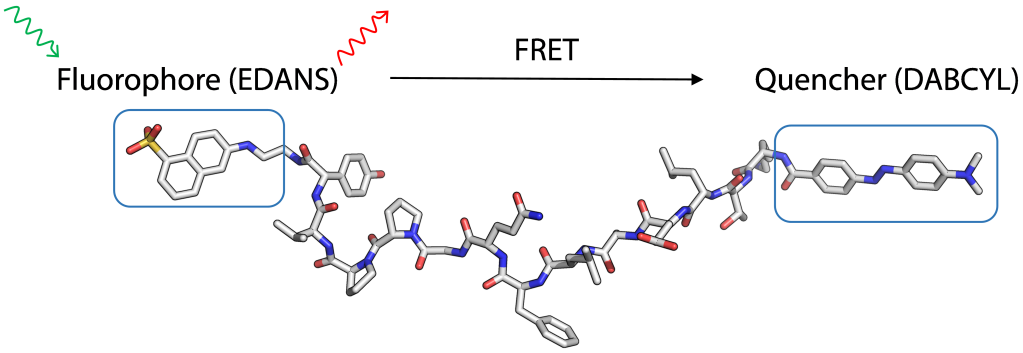

Figure 3.3: FRET substrate utilized in 3C protease kinetic assay.

3.2.2 Protein mass spectrometry

In mass spectrometry (MS) chemicals are ionized and characterized by the mass to charge ratios $\mathrm{m/z}$ of formed ions [202]. For ionization several methods have been developed, which include (i) electron ionization [203], (ii) electrospray ionization (ESI) [204] and (iii) matrix-assisted laser desorption/ionization (MALDI) [205]. Electron ionization is a so-called hard ionization technique, which refers to the potential to induce fragmentation in the ionized molecule. This can be very useful as fragmentation provides additional information about the analyzed molecules as this proceeds via specific fragmentation patterns. In contrast, ESI and MALDI are soft ionization techniques, which cause little fragmentation. The separation of the ions based on the $\mathrm{m/z}$ is done by applying an electromagnetic field, which affects the subsequent path of the ion to the detector. In sector instruments the path of the ions is altered and the degree of deflection correlates with the $\mathrm{m/z}$ [206]. In time of flight (TOF) MS the ions are accelerated toward the detector with a constant current, whereas the duration depends on the $\mathrm{m/z}$ [207]. In quadrupole mass analyzers an oscillating electric field is generated between four parallel rods, which allows only ions with a specific $\mathrm{m/z}$ to pass at a time [208]. This is also used in ion trap-based mass selectors. Hereby, the ions are trapped within an electric field and sequentially passed on to the detector [209]. The detector is an electron multiplier in most cases, which detects either the charge or the induced current. The advantages of MS are that at one hand the detected $\mathrm{m/z}$ of an analyzed molecule or its fragments allow for a direct link to the chemical structure. On the other hand, MS is compatible with chromatographic methods such as liquid- or gas chromatography. In such LC/MS or GC/MS systems, MS is used for detection after chromatographic separation of the sample components. In drug design, MS of proteins is utilized to detect the binding of ligands, which is especially suit-

able for irreversibly binding ligands [210]. Hereby, MS is used as an orthogonal screening method. From the detected $\mathrm{m/z}$ after deconvolution the composition of the protein-ligand complex can concluded in terms of covalent linkage and ratio, which is key to assess the specificity of a covalent ligand.

Chapter 4

Results Part I: Covalent Binders

4.1 3D Pharmacophore model for the $3C^{pro}$

To enable a rational approach for the fragment-based design of $3C^{pro}$ inhibitors a 3D pharmacophore model comprising the essential protein-ligand interactions necessary for target binding was developed. Thereby, fragments can be presampled *in silico* so that only an enriched set has to be evaluated in the more time and material consuming enzymatic assay.

The applicability of 3D pharmacophores for this target has been enabled by my preceding master thesis entitled "*Design of Novel Covalent Coxsackievirus B3 3C Protease Inhibitors*" [169]. Herein a novel chemical feature type has been introduced into the pharmacophore modeling tool LigandScout [164] representing covalent protein-ligand interactions. This enables the rational identification of covalently binding ligands from screening collections, which is an essential part of the aspired *in silico* fragment screening.

Therefore, the first part of this thesis, described in this section, was to elucidate the structural characteristics of the $3C^{pro}$ active site in order to identify the key interactions essential for substrate and inhibitor binding. This enabled the subsequent generation of a 3D pharmacophore model from $3C^{pro}$ co-crystal structures, which was validated through the use of bioactive ligands reported in the literature and decoy molecules.

4.1.1 3C Protease structural analysis

The structural investigations of the $3C^{pro}$ is based on protein-ligand crystal structures available in the Protein Data Bank (PDB) [211]. Thereby, two main questions were addressed: (i) Which interactions are formed between protein and substrate or inhibitors, respectively, and (ii) which differences do exist for $3C^{pro}$ of the different enterovirus species?

To compare the closely related enterovirus species only those that are pathogenic for human were considered and for which a structure with a co-crystallized ligand was available in the PDB. This includes the four human-pathogenic enteroviruses A-D as well as the rhinovirus species A and B as listed in Table 4.1. At first, a sequence alignment was performed using Clustal Omega [212]. The alignment revealed an overall identity of 29 % of the amino acids and a similarity of 50 %, whereas the rhinovirus and the enterovirus species have the highest accordance. In addition to the overall high similarity, three regions are exceptionally well conserved. These comprise the residues 81-86, 142-150 and 161-167, respectively (Fig-

Table 4.1: 3C^{pro} structure comparison I. 3C^{pro} crystal structures of the representative genotypes for the comparison of the different enterovirus species with the following abbreviations: EV, enterovirus; CV, coxsackievirus; PV, poliovirus; HRV, human rhinovirus.

Family	Genus	**Species**	**Genotype**	**PDB ID**
Picornaviridae	Enterovirus	Enterovirus A	EV 71	4GHT
		Enterovirus B	CV B3	3ZZA
		Enterovirus C	PV 1	4DCD
		Enterovirus D	EV D68	3ZVE
		Rhinovirus A	HRV A2	4CQQ
		Rhinovirus B	HVR B14	2B0F

ure 4.3). The relevance of these conserved residues was further elucidated through structural exploration of the 3C^{pro}. This was done by superimposing the protein structures in MOE, version 2014.09. The α-carbon root-mean-square-deviation (RMSD) was calculated for the total structure and the active site residues with an average value of 1.23 Å and 0.95 Å, respectively (Figure 4.1A). The small structural deviations are reflected by the superimposed backbones, which only differ at loop regions. This is in accordance with the high sequence similarity. In addition, the binding mode of the co-crystallized ligands, which belong to the class of peptidic 3C^{pro} inhibitors described in subsection 1.2.3, are virtually the same (Figure 4.1B). Two regions with sequences of conserved residues are part of the active site architecture and form key interactions with the ligands. These are the residues 142-150, which are involved in formation of the S1 pocket and the oxyanion hole and the residues 159-167, which are involved in binding P1 to P4 residues of substrates as will be described in more detail below. The third conserved sequence is located far off from the active site and is implicated in RNA binding and RNA replication as described in subsection 1.2.3.

The structural features of the 3C^{pro} specificity were evaluated through catalytic cleavage of artificial peptides, which mimic the different junction in the viral polyprotein as reported in literature [97, 213]. Table 4.2 lists the determined catalytic efficacies of the 3C^{pro} from EV D68 and EV 71 for a small variety of peptides. The crystal structure 3ZZA of the CVB3 3C^{pro} was used to model the binding of a short peptide based on the sequence of the substrate utilized in the biochemical assay (Figure 4.2). Thereby, the protein substrate interactions and the specificity indicated by the differences in catalytic efficacy were investigated. The most important feature of the 3C^{pro} substrate interactions is the hydrogen bond pattern in the S1 pocket, comprising three hydrogen bonds, formed between the terminal

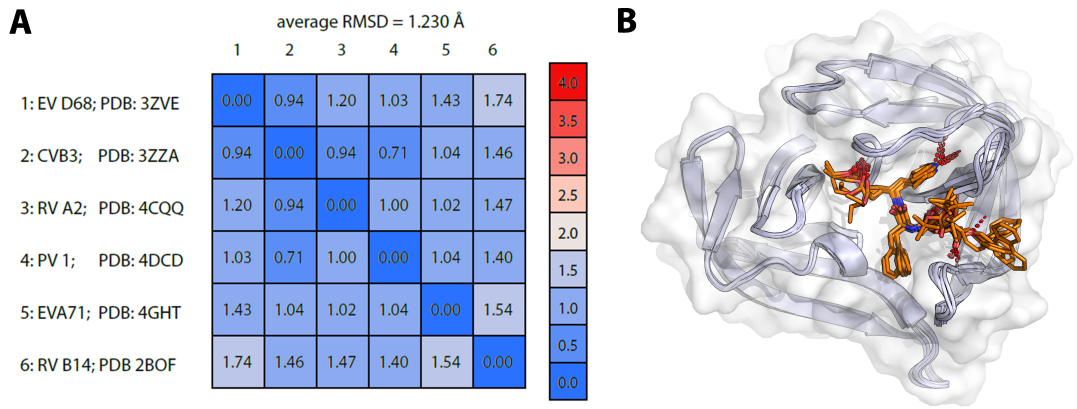

Figure 4.1: 3C^{pro} structure comparison II. (A) α-carbon RMSD of the superpositioned 3C proteases of the different enterovirus genera for the overall structure. (B) Superpositioned 3C^{pro} crystal structures of the different genera with the co-crystallized ligands depicted in orange.

Table 4.2: 3C^{pro} catalytic efficiency. Catalytic efficiency of the 3C^{pro} from EV D68 and EV71 for different artificial peptides as reported by Lu *et al.* and Tan *et al.*, respectively [97, 213].

Genotype	Peptide	Catalytic efficiency
EV D68	EALFQ–GPPQF	200 µM/h
	TAKVQ–GPGFD	200 µM/h
	FAGIQ–GPYTG	75 µM/h
	HAITQ–GVPTY	50 µM/h
EV 71	FAGLRQAVTQ–GFPTEL	5.08 µM/min
	FAGLRQAV**F**Q–GFPTEL	14.20 µM/min
	FAGLR**A**AVTQ–GFPTEL	4.11 µM/min
	FAGLRQAV**E**Q–GFPTEL	3.65 µM/min
	FAGLRQAVTQ–**S**FPTEL	0.21 µM/min
	FAGLRQ**L**VTQ–GFPTEL	0.00 µM/min

amide moiety of the P1 glutamine with the side chain imidazole of H161 and the side chain hydroxyl moiety of T142 as well as the backbone carbonyl group. In addition, the substrate recognition and cleavage depends on multiple residues as reflected by the length of the utilized peptides with a minimum of ten residues. However, P1 glutamine is most important as cleavage only occurs after this amino acid and preferably in conjunction with glycine. The substitution of glycine for serine results in a dramatic drop in catalytic efficacy as shown for the peptide substrates evaluated using the EV 71 3C^{pro} (Table 4.2). This is due to the lack of a specific pocket to accommodate the side chain of P1' residues. The S1' pocket hosts the P2' residue, which is preferably proline but also phenylalanine and other

amino acids with small lipophilic side chains are tolerated due to possible hydrophobic contacts with F25. In general, the specificity for this position as for the primed site in general is little as the 3C^{pro} active site does not extend to this site. The oxyanion hole is formed by residues A144, G145, Q146 and C147, whereas the backbone nitrogen of C147 and G145 engage in hydrogen bonding with the oxyanion formed during catalysis (see subsection 1.2.2). The S2 pocket, formed

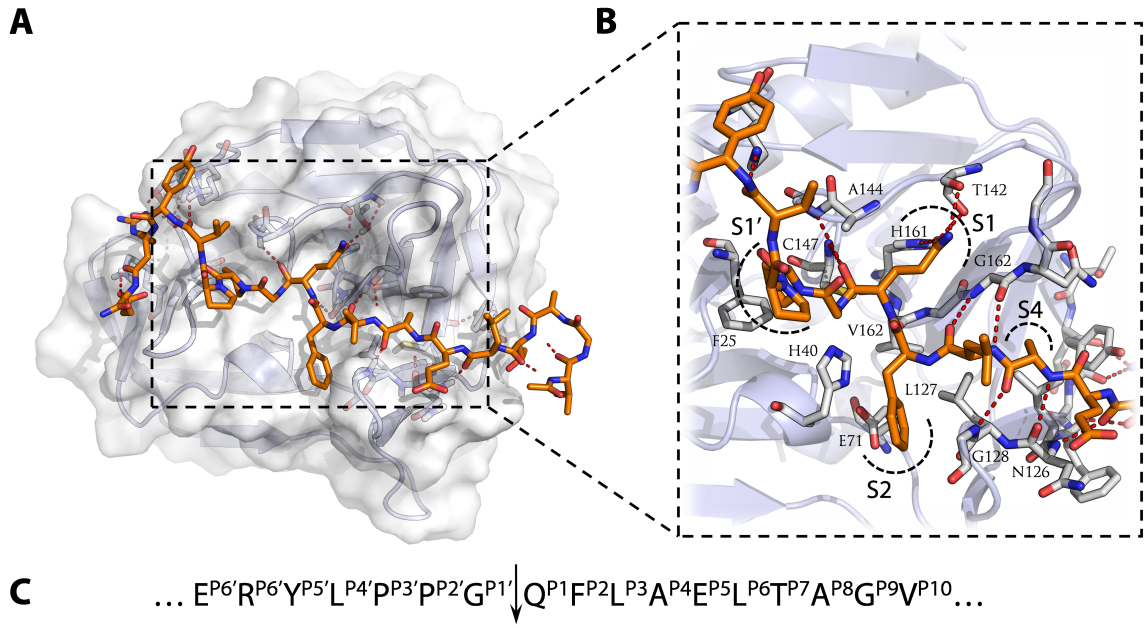

Figure 4.2: 3C^{pro} substrate interactions. (A) Substrate bound to active site of the CVB3 3C^{pro}. (B) Close up on the key interactions in the active site pockets between a substrate and the CVB3 3C^{pro}. Hydrogen bonds are depicted as red dotted lines. (C) Amino acid sequence of the representative substrate with cleaving site indicated by an arrow.

by the side chains of the catalytic residues H40 and E71 on one side and L127 on the other, tolerates a variety of residues, whereas lipophilic contacts can be formed with L127, which increase the binding affinity as exemplified by the increase in catalytic efficiency through the introduction of phenylalanine as P2 residue. The P3 position also shows little specificity, which is due to the fact that a putative side chain of a P3 residue would point away from the active site and no S3 pocket exists. Instead, the key interactions at this position are hydrogen bonds formed by the backbone atoms in a β-sheet like interaction pattern to G162. This is continued by the subsequent residue at the P4 position by hydrogen bonding to G128 and N126. However, this position is very specific for alanine as the S4 pocket is only large enough to accommodate the small methyl group in the alanine side chain. Peptides, in which the alanine is substituted for amino acids with bigger

```
-GPEEEFGMSLIKHNSCVITTENGKFTGLGVYDRFVVVPTHADPGKEIQVDGITTKVIDS   59
-GPNTEFALSLLRKNIMTITTSKGEFTGLGIHDRVCVIPTHAQPGDDVLVNGQKIRVKDK   59
MGPSLDFALSLLRRNVRQVQTDQGHFTMLGVRDRLAVLPRHSQPGKTIWIEHKLVNILDA   60
-GPGFDYAVAMAKRNIVTATTSKGEFTMLGVHDNVAILPTHASPGESIVIDGKEVEILDA   59
MGPAFEFAVAMMKRNSSTVKTEYGEFTMLGIYDRWAVLPRHAKPGPTILMNDQEVGVLDA   60
-GPGFDFAQAIMKKNTVVARTEKGEFTMLGVHDRVAVIPTHASVGETIYINDVETKVLDA   59
 **  ::. :: ::*     *. *.** **: *.  ::* *:. *  : ::     : *

YDLYNKNGIKLEITVLKLDRNEKFRDIRRYIPNNEDDYPNCNLALLANQPEPTIINVGDV   119
YKLVDPENINLELTVLTLDRNEKFRDIRGFISEDLE-GVDATLVVHSNNFTNTILEVGPV   118
VELVDEQGVNLELTLITLDTNEKFRDITKFIPENISTASDATLVINTEHMPSMFVPVGDV   120
KALEDQAGTNLEITIITLKRNEKFRDIRPHIPTQITETNDGVLIVNTSKYPNMYVPVGAV   119
KELVDKDGTNLELTLLKLNRNEKFRDIRGFLAKEEVEVNEAVLAINTSKFPNMYIPVGQV   120
CALRDLTDTNLEITIVKLDRNQKFRDIRHFLPRYEDDYNDAVLSVHTSKFPNMYIPVGQV   119
  * :    :**:*::.*. *:*****  .:       :  * : :..     : ** *

VSYGNILLSGNQTARMLKYSYPTKSGYCGGVLYKIGQVLGIHVGGNGRDGFSAMLLRSYF   179
TMAGLINLSSTPTNRMIRYDYATKTGQCGGVLCATGKIFGIHVGGNGRQGFSAQLKKQYF   178
VQYGFLNLSGKPTHRTMMYNFPTKAGQCGGVVTSVGKVIGIHIGGNGRQGFCAGLKRSYF   180
TEQGYLNLGGRQTARTLMYNFPTRAGQCGGVITCTGKVIGMHVGGNGSHGFAAALKRSYF   179
TEYGFLNLGGTPTKRMLMYNFPTRAGQCGGVLMSTGKVLGIHVGGNGHQGFSAALLKHYF   180
TNYGFLNLGGTPTHRILMYNFPTRAGQCGGVVTTTGKVIGIHVGGNGAQGFAAMLLHSYF   179
.  * : *..  * * : *.: *::* ****:   *:::*:*:**** .**.* * : **

T-----180     RV A2
VEKQ--182     RV B14
ASEQLE186     EV  71
TQSQ--183     PV 1
NDEQ--184     CV B3
TDTQK-184     EV D68
```

Figure 4.3: 3C^{pro} sequence alignment of the analyzed enterovirus species. Residues forming the active site are marked by a green line. Regions of highly conserved residues are marked by a blue line.

side chains such as leucine (Table 4.2), are hardly cleaved by the 3C^{pro}. The subsequent residues of the non-primed site do not bind to specific pockets but continue the backbone hydrogen bond pattern and thereby contribute to the correct positioning of the aforementioned residues.

The described interactions and residues involved are based on the Coxsackievirus B3 (CVB3) 3C^{pro} as it is used as representative 3C protease in this thesis. However, the interactions can be found for each 3C^{pro} of the investigated enterovirus species as the residues are either identical or the interactions occur through backbone atoms. The high similarity of the 3C^{pro} structures and substrate interactions exemplifies a conserved feature of the 3C^{pro} in respect to the otherwise high variability of viral proteins. Therefore, the 3C^{pro} is an attractive target for the development of broad spectrum enterovirus inhibitors.

4.1.2 3D Pharmacophore model generation & validation

Coxsackievirus B3 $3C^{pro}$ was chosen as representative $3C^{pro}$ for the pursued approach as a biochemical assay for the evaluation of inhibitors exists in-house. For the development of a structural 3D pharmacophore model crystal structures of the CVB3 $3C^{pro}$ available from the PDB served as starting point. However, only eight out of 18 structures listed in the PDB could be used for the development of the 3D pharmacophore model as the others fell short due to at least one of the following issues: (i) Apo protein structure, (ii) non-covalent ligand or (iii) structure quality issues. The utilized crystal structures are listed in Table 4.3. The co-crystallized ligands, depicted in Figure 4.4C, belong to the class of peptidic $3C^{pro}$ inhibitors, which mimic the natural substrate as described in subsection 1.2.3. The interactions formed between the ligands and the protein are therefore highly similar to what has been described in the previous subsection. The major difference is the covalent bond formed between an α,β-unsaturated ester of the co-crystallized ligand and the catalytic cysteine.

Table 4.3: PDB crystal structures of the CVB3 $3C^{pro}$ utilized for 3D pharmacophore model generation.

PDB ID	Ligand ID	Resolution [Å]	Mutations
3ZZ6	**G75**	2.05	-
3ZZ7	**G81**	1.80	-
3ZZ8	**G82**	1.85	-
3ZZ9	**G83**	1.90	-
3ZZA	**G84**	1.80	-
3ZZB	**G85**	2.10	-
3ZZC	**G83**	2.10	T68A, N126Y
3ZZD	**G85**	2.10	T68A, N126Y

For each crystal structure the interactions formed between the ligand and the protein were identified and transferred into a 3D pharmacophore feature pattern using LigandScout. Hereby, the interactions are represented in generalized way through pharmacophore features in a specific spatial arrangement (Figure 4.4A,B). The covalent bond is represented by the orange sphere of the newly introduced covalent bond feature. The pharmacophore feature patterns obtained from the different crystal structures were aligned and a shared feature pharmacophore was generated. This model comprises the features present in all of the individual interaction patterns and reflects their spatial variance. In order to elucidate, which

of the interactions are essential for the inhibition of the $3C^{pro}$ and to optimize the exclusion volume spheres representing the shape of the active site this model had to be validated.

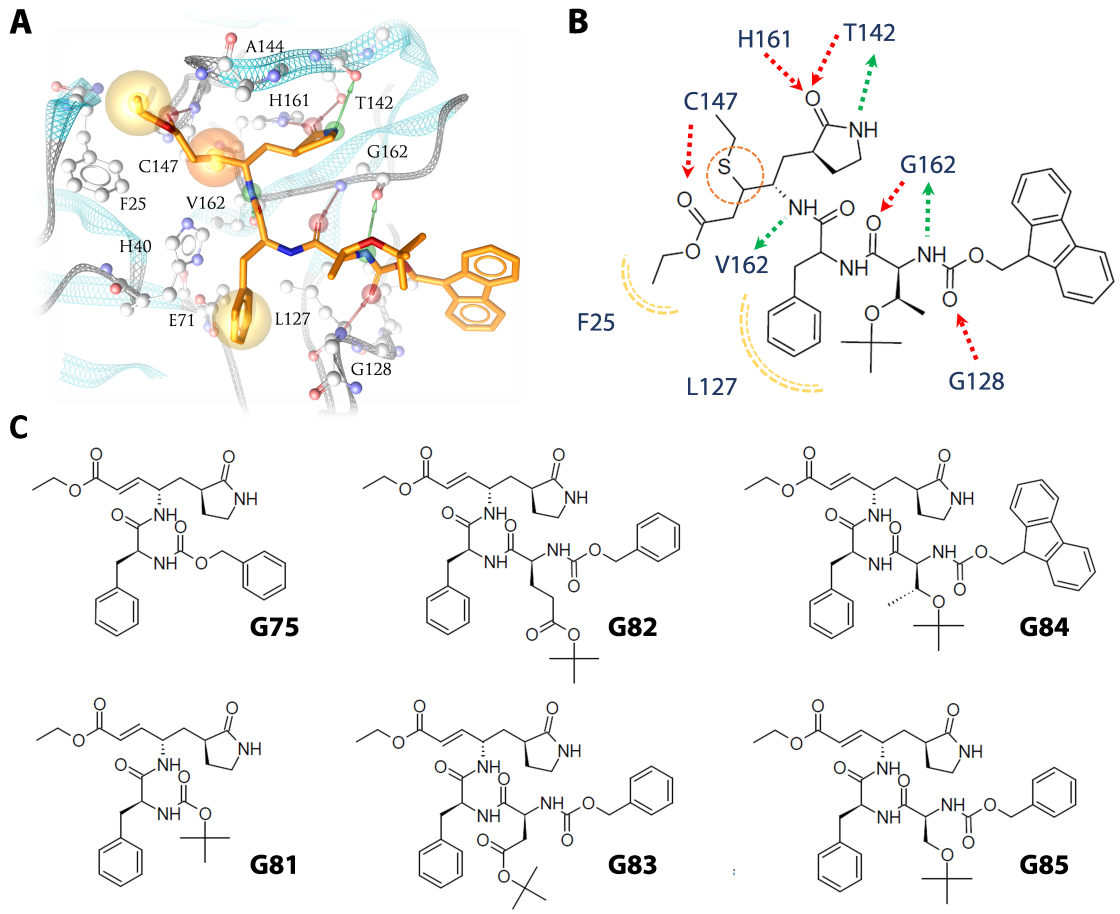

Figure 4.4: $3C^{pro}$ structure-based 3D pharmacophore. Protein-ligand interactions of **G84** co-crystallized with the CVB3 $3C^{pro}$ of PDB ID 3ZZA represented by pharmacophore features in 3D (A) and 2D (B). An orange sphere indicates a covalent bond, yellow spheres lipophilic contacts, red arrows hydrogen bond acceptors and green arrows hydrogen bond donors, respectively. (C) Chemical structures of co-crystallized ligands utilized for 3D pharmacophore model generation.

For the validation of the pharmacophore hypothesis ligands with known bioactivity against the $3C^{pro}$ were required. However, inhibitors of the CVB3 reported in the literature belong to one of the following categories: (i) Peptidic scaffold, (ii) unknown or questionable mode of action or (iii) only evaluated in cell-based experiments. Either one of those categories made the inhibitors not applicable for the desired purpose. The specific covalent interaction with the $3C^{pro}$ is a prerequisite to match the 3D pharmacophore model, whereas peptidic inhibitors adhere to the same interactions formed by the co-crystallized ligands used to build the model. Since the aim is the identification of non-peptidic inhibitors this would not

be conducive. To overcome this limitations, non-peptidic inhibitors of the human rhinovirus HRV14 were used. The high structural similarity of the $3C^{pro}$ from different species assessed before assures activity of these compounds to the CVB3 $3C^{pro}$ as well as conservation of the binding mode. Through a literature search, $3C^{pro}$ inhibitors representing three different scaffolds were identified as useful bioactive ligands for validation. This includes benzamides reported by Reich *et al.*, which bind to the S1 pocket by the name giving structure and form a covalent bond to the catalytic cysteine through a Michael acceptor moiety [98] (Figure 4.5A). Johnson et al reported a set of peptidomimetic inhibitors of the $3C^{pro}$, which mimic the glutamine by a γ-lactam moiety [100] (Figure 4.5A). The covalent bond formation also occurs via an α,β-unsaturated ester. In contrast, the isatines (1H-indole-2,3-dione) reported by Webber *et al.* form a reversible thiohemiketal through the cyclic ketoamide moiety [99] (Figure 4.5A). The amide substituent forms the crucial hydrogen bonds in the S1 pocket. From each set only those actives with k_{inact}/K_I values greater than 100 $M^{-1}s^{-1}$ or K_i values less than 1 µM were selected. This strict activity cut off ensures inhibition across 3C proteases of different enterovirus species and led to a total of 33 inhibitors in the active set (appendix A.1).

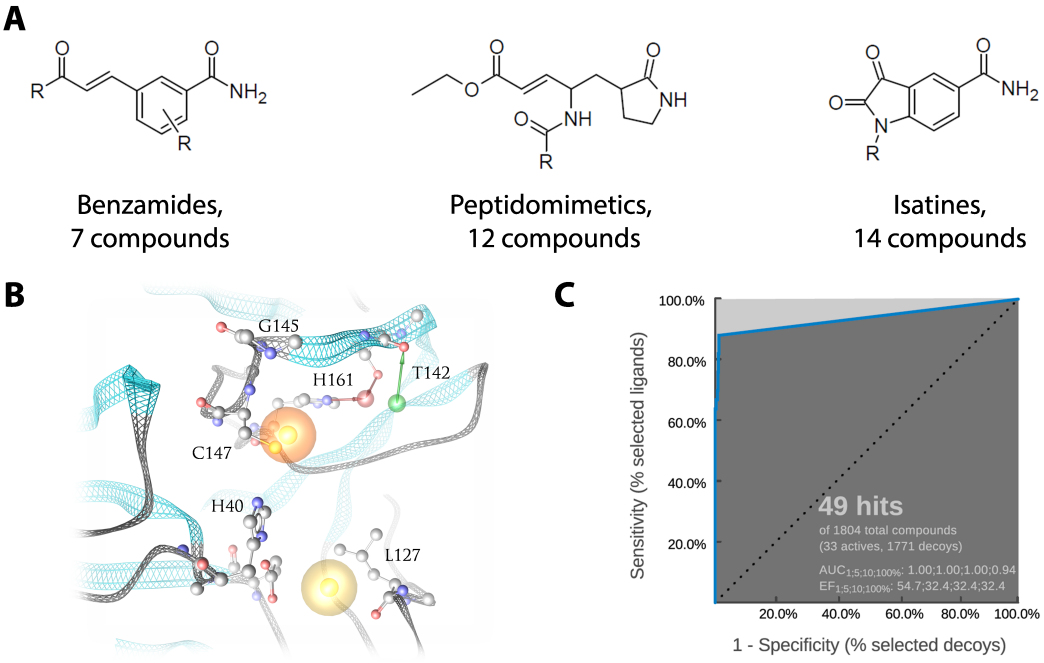

Figure 4.5: 3D pharamcophore model of CVB3 $3C^{pro}$. (A) Active set scaffolds and number of utilized compounds per scaffold. (B) Validated 3D pharmacophore model in active site of CVB3 $3C^{pro}$. (C) ROC curve obtained for 3D pharmacophore model validation using the actives and decoys.

The selection of suitable decoys for the validation of the pharmacophore model was exceptionally challenging for this case. On the one hand, due to the covalent bond feature being part of the 3D pharmacophore and on the other hand, the fact that compounds, which contain a warhead moiety necessary to match the respective pharmacophore feature, are often removed from screening collections to circumvent putative reactivity-based non-specific binding. Therefore, the decoys utilized for validation were generated *de novo* through a virtual molecule generation workflow developed in thesis. The primary objective for the development of the *de novo* design application, which is described in section 4.5, was the optimization of identified fragment hits but it is also beneficial for the generation of decoys. In brief, molecules are generated virtually by combining tangible chemical building blocks through synthesis tailored reaction rules. This ensures the generated compounds to be chemically reasonable and synthetically accessible. For the generation of useful decoys, this virtual molecule generation was advantages as the incorporation of the chemical moieties essential to match the pharmacophore could be ensured. The generation and final selection of the decoys, which is described in more detail in subsection 4.5.6, lead to a final number of 1771 useful decoys.

By the use of the collected active and decoy sets, the initial pharmacophore model, based on the merged 3D pharmacophore interaction patterns, was validated. The essential interactions formed by the actives include only three regions of the active site (Figure 4.5B). The covalent bond to the cysteine and the hydrogen bonds in the S1 pocket are formed by the scaffold of each respective active set as described before. The variable substituents match a lipophilic feature representing hydrophobic contacts in the S2 pocket with the side chain of L127. Additional interactions derived from the co-crystallized ligands in the non-primed site are hydrogens bonds formed by their peptidic backbone. Therefore, these are not matched by the active set ligands and were thus removed from the model. Features within the primed site were also removed from the pharmacophore model since the ability to match these interactions depends primarily on the nature of the warhead and thus would make the pharmacophore model more restrictive. The final model depicted in Figure 4.5B, is able to identify 29 of the 33 true actives, whereas 1751 of the 1771 decoys are rejected, which results in a sensitivity of 0.87 and a specificity of 0.99 as shown by the ROC curve depicted in Figure 4.5C [172]. This excellent performance is benefiting from the high degree of binding mode similarity within the active set molecules but also from the specific phar-

macophore composition of the covalent feature in close proximity to the hydrogen bond features representing the essential S1 binding pattern. This feature composition makes the 3D pharmacophore quite restrictive by itself. The alignment of a representative ligand from each set of actives is depicted in Figure 4.6A-C.

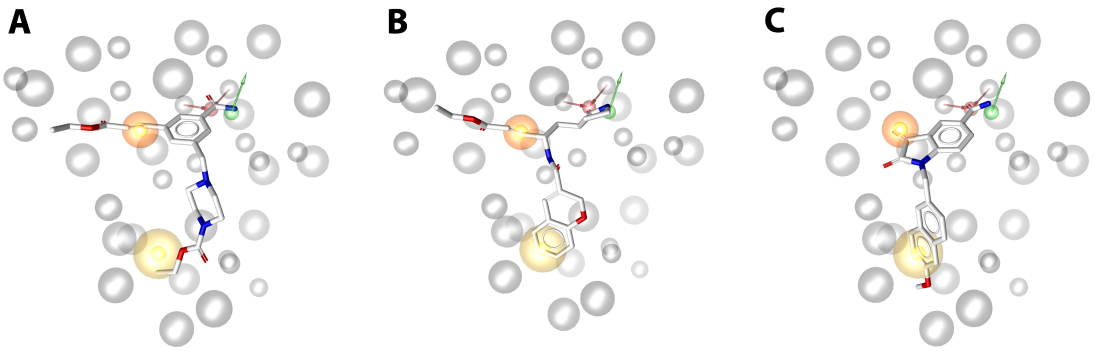

Figure 4.6: Active set molecules aligned to 3D pharmacophore model of CVB3 3C^{pro}. Representative structures for the benzamides (A), peptidomimetics (B) and isatines (C) aligned to the 3D pharmacophore model.

The validated 3D pharmacophore model comprises the essential features for binding of non-peptidic inhibitors to the CVB3 3C^{pro}. Therefore, it served as basis for virtual screening-based identification of putative 3C^{pro} binding fragments. The incorporation of a pharmacophore feature representing the covalent interaction enables the aspired rational design of covalently binding ligands.

4.2 Exploring covalent binding characteristics

The generated 3D pharmacophore model and the pursued approach of the rational design of covalently binding inhibitors was first challenged by the work of Daniel Becker in the collaborating group of Prof. Rademann. They were interested in the rationalization of the biological data trough structural models for covalent inhibitors of the $3C^{pro}$ that have been developed as part of his PhD thesis. The undertaken work raised several questions, which led us reconsider and investigate basic principles of the covalent protein-ligand interactions through a computational study described in this section.

4.2.1 Covalent ligands through templated assembly

In his PhD thesis, Daniel Becker was also interested in the development of covalent CVB3 $3C^{pro}$ inhibitors through a templated assembly approach, in which fragment-sized molecules are ligated in the presence of the targeted protein. This is analogue to the previously described linking of fragment but the formation of favorable fragment combinations is promoted by the templating effect of the protein. Additivity in binding of the fragment combinations can be directly observed through the biological readout. A library of amine bearing fragments was screened in combination with the bis-electrophile **1** (Figure 4.7). On the one hand, **1** can covalently bind to the catalytic cysteine of the $3C^{pro}$ via its epoxide moiety and on the other hand, it can form a Schiff base with the amine group of the fragments. Fragment **2** shows over additivity in binding in combination with **1**, which was reasoned to be based on the ability to bind to the crucial S1 pocket as its pyrazolone moiety is bioisosteric to the amide group of the P1 glutamine (Figure 4.7). Several combinations of **1** and **2** were synthesized, based on the hypothetical Schiff base being formed as ligations product, of which compound **3** was the most potent with a k_{inact}/K_I of 606 $M^{-1}s^{-1}$ (Figure 4.7).

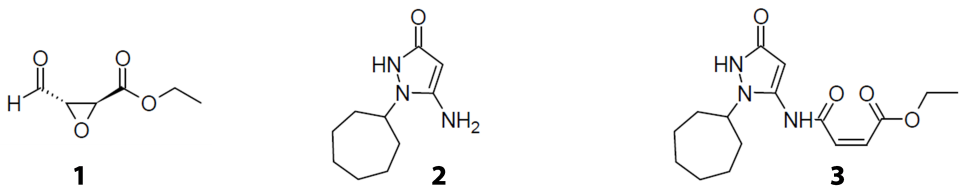

Figure 4.7: Chemical structures of compounds utilized in the templated assembly approach.

In order to elucidate the binding mode of compound **3**, two assumptions were made: (i) A covalent bond is formed via the Michael acceptor warhead to the catalytic cysteine and (ii) the lactam moiety of the pyrazolone forms the crucial hydrogen bonds in the S1 pocket. Kinetic and protein mass spectrometry (MS) data confirmed the formation of a covalent bond, whereas the relevance of the S1 pocket for ligand binding is based on structure-activity data described in the previous section and utilized to build the 3D pharmacophore model. Putative binding poses for compound **3** were generated through covalent protein-ligand docking. However, in none of the obtained poses the pyrazolone moiety was positioned in the S1 pocket. This prompted the question on the contributions of the individual moieties to the overall binding. Does the pyrazolone moiety in deed bind to the S1 pocket but the respective conformation is subsequently perturbed upon covalent bond formation? This question can be put into a wider perspective considering the general scheme for the interaction of covalently binding ligands introduced in section 1.3. How much does the binding mode change during the transition from the initial non-covalent complex (EI) to the covalent complex (E-I) or how similar are these two states? This was not only relevant for binding mode investigations of the ligation products but also for our pursued approach as the prior generated 3D pharmacophore model is based on the covalent complex derived from crystal structures and thus comprises the interaction of the covalently bound ligand in the covalent complex. However, it was supposed to identify ligands from screening libraries, which comprise the ligands in the pre-reaction state as in the initial non-covalent complex.

The specificity of covalent drugs, which is assumed to be the result of the proper spatial arrangement of the warhead described in section 1.3, makes dramatic changes of the binding mode unlikely. On the other hand, through the formation of a covalent bond, a ligand can undergo major alterations in its chemical structure (i) due to the elimination of big leaving groups, (ii) opening of cyclic warheads or (iii) changes in the geometry of the warhead. This make conservation of the initial binding mode less likely.

The challenge of investigating the initial complex lies in its transient nature. In order to provide a proper approximation, one would either need to synthesize an analogue of the covalent ligand missing the warhead and solve the crystal structure of it binding to the target or by generating a modified version of the protein, which can no longer interact covalently with the ligand. For either case, the binding mode has to be unaffected by the changes in order to draw relevant

conclusions. In addition, an extensive amount of time and resources would be necessary as this would have to be done for a representative number of covalent protein-ligand pairs. Therefore, we envisioned an *in silico* approach for the investigation of the initial complex, which will be described in the following subsection.

4.2.2 Reconstructing the initial complex

In order to investigate the initial non-covalent complex, computational methods were employed to generate and compare it to the covalent complex, which has been experimentally determined through X-ray crystallography. Thereby, we did not aim for the generation of the 'true' initial complexes as this could not be verified without the above described experiments but we wanted to assess if it is similar to the covalent complex. Therefore, protein-ligand docking was utilized to sample possible binding conformations of the non-covalently bound ligand. Their similarity to the covalent complex was assessed through comparison of binding conformations and interactions formed to the protein. Heavy atom RMSD was calculated for comparison of the covalently bound conformer of the crystal structure and the non-covalently bound conformations generated through protein-ligand docking. Since the chemical structure of the ligand in both states differs due to the reaction of the warhead with the protein, only the core part of the ligands was taken into account (Figure 4.8A). The interactions of the covalent complex in the crystal structure were transferred into a pharmacophore feature pattern to which the docking based conformers were compared. This is reported as relative feature fit (Fit), which describes the quality and the number of features (interactions) matched by the ligand in a respective docking pose in a normalized way.

At first, a test set of covalent protein-ligand crystal structures was assembled, which covers 24 different proteins listed in Table 4.4. To prevent a bias from proteins overrepresented in the PDB, the number of utilized crystal structures per protein was restricted to four. Furthermore, inhibitors, which belong to the class of enzyme activated suicide inhibitors, were discarded as the mechanism of their covalent interaction is out of the scope of this study. For each of the crystal structures in the test set, the ligand structure in the covalent-bound and in the initial state was generated (see appendix A.2), the ligand core was defined and the interaction of the covalent protein-ligand complex was transferred into a 3D pharmacophore feature pattern.

The ligands were covalently docked in order to generate the covalent complex of the crystal structure and to evaluate whether the utilized docking tool

Table 4.4: Proteins with covalent crystal structures used for the generation of the initial complex.

Protein targets			
Adenovirus 2 protease	Cruzain	Factor XIa	HCV-Protease
Autotaxin (E-NPP 2)	DPP IV	FGFR-4	Proteasome
Calpain-1	T-1 Dehydroquinase	β-Lactamase	K-Ras
Caspase-3/8	EGFR	HCV-Polymerase	RSK2
Cathepsin A, K, L, S	ERK-2	PPARγ	Sortase A
CRM1/XPO1	FAAH	3C/3CL Protease	Transthyretin

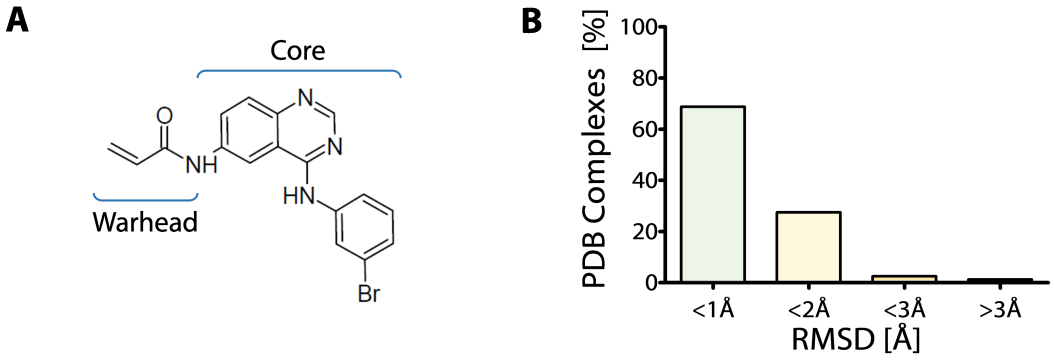

Figure 4.8: Covalent redocking results. (A) Definition of the core structure for a representative covalent ligand of the test set. (B) Distribution of the RMSD values from the covalent redocking of the ligands from the crystal structures in the test set.

GOLD is able to generate biologically relevant binding conformations for the protein ligand complexes in the test set and if the applied similarity metrics are able to retrieve these from the ensemble of generated binding conformations. The obtained docking poses were compared to the feature patterns and the RMSD of the top scored poses was calculated. For over 90 % of the complexes, docking poses were obtained, which deviated less than 2 Å from the conformation in the crystal structure (Figure 4.8B). For two out of three of the complexes, the RMSD is even below 1 Å, which represents almost perfect alignment of the binding conformations. This indicates that the utilized docking tool is able to generate the correct binding conformation and that the applied similarity metrics are appropriate to enrich them for most of the protein ligand complexes in the test set.

Next, the ligands were docked non-covalently into the respective crystal structure to generate an ensemble of possible binding conformations representing the initial complex. The obtained poses were scored by using the feature pattern of the respective covalent complex. For the top scored poses, the RMSD to the con-

formation of the core of the co-crystallized ligand was calculated. Around 60 % of the analysed poses show a RMSD below 1.5 Å and 20 % below 3 Å. This indicates a general agreement between the conformation in the crystal structure and the selected docking pose (Figure 4.9A). Relaxed binning criteria were used in comparison to the covalent redocking as two different chemical entities were compared. The results for the relative feature fit are similar. 60 % of the docking poses have a Fit of 0.8 or higher, which implies that 80 % of the interactions formed in the crystal structure are equally formed in the selected docking pose of the non-covlaently bound ligand (Figure 4.9B).

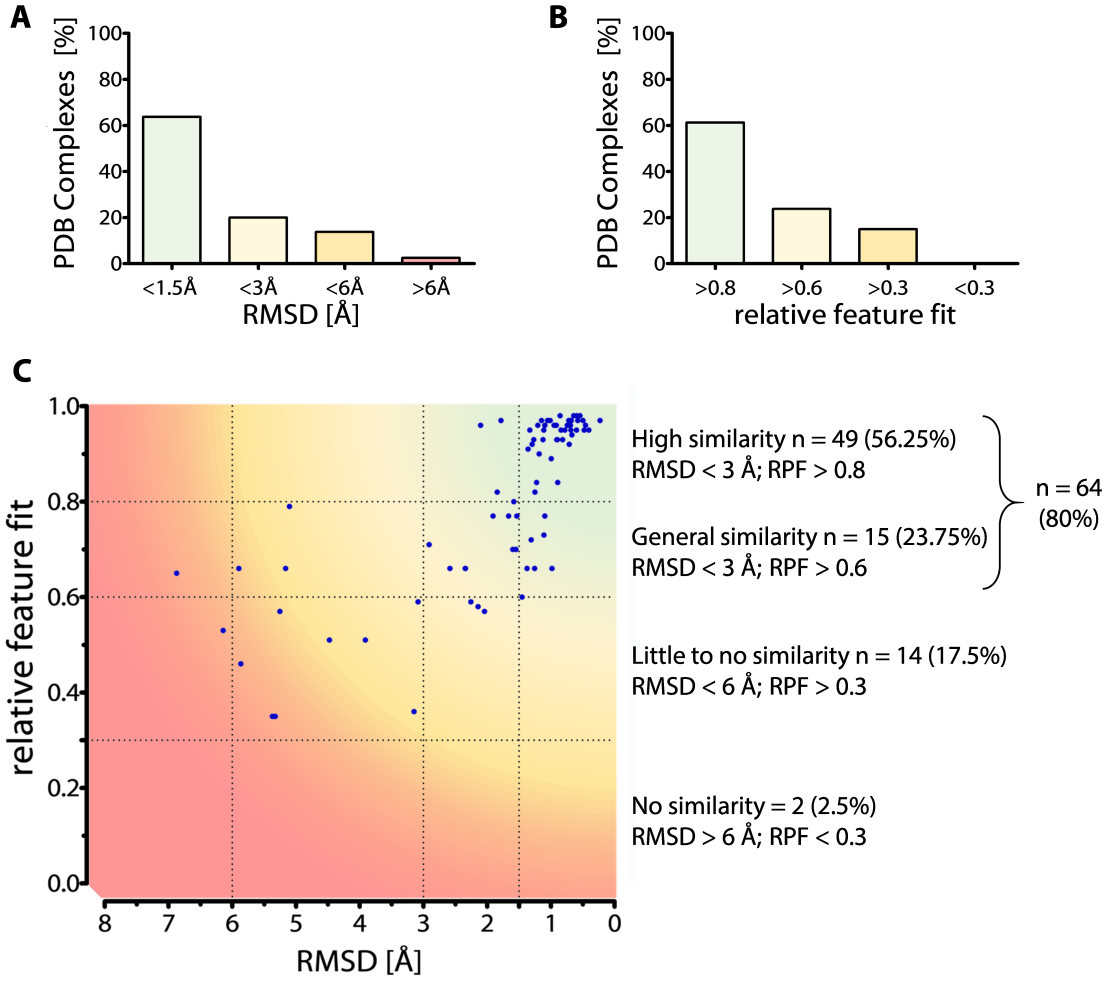

Figure 4.9: Initial complex similarity to covalent complex. (A) Distribution of RMSD values of the selected docking pose of each ligand in the test set. (B) Distribution of the relative feature fit of the selected docking pose of each ligand in the test set. (C) 2D Matrix for the selected docking pose of each ligand in the test set based on its RMSD and the relative feature fit. Color coding indicates degree of similarity, which reduces from green (high similarity) to red (no similarity).

The overall agreement of the two applied similarity metrics was a first indication for a general binding mode conservation during covalent bond formation.

However, the metrics do not necessarily correlate as the similar spatial positioning of a functional group does not inevitably imply formation of the same interactions. Vice versa, a high Fit value does not require a conformational similarity as the respective interactions can also be formed by a different moiety of the ligand. Therefore, both metrics were applied jointly. The selected docking pose of the initial complex for each investigated protein ligand complex was plotted on a 2D matrix based on its RMSD and Fit value as depicted in Figure 4.9C. The obtained plot shows a general agreement between RMSD and feature fit for those poses with an RMSD below 3 Å. This includes poses with an RMSD <1.5 Å, which form more than 80 % of the interactions and thus show very high similarity to covalent complex. More than 50 % of the analyzed crystal structures can be categorized into this group. The second group of docking poses showing reasonable similarity, resulting in an RMSD between 3-6 Å and a Fit over 0.6 and contains about 25 % of the poses. Therefore, for over 75 % of the analyzed crystal structures a docking pose of the non-covalently bound ligand was generated, which is similar to the covalent complex in terms of its binding conformation and the interactions formed with the protein. Representative examples are depicted in Figure 4.10 illustrating the general agreement between the binding conformations of both states.

Around 20 % of the docking poses show an RSMD above 3 Å and a relative feature fit from 0.3 to 0.8. This indicates that there is no conformational similarity to the covalently bound ligand and the interactions are putatively formed by a different moietiy of the ligand. For these 20 % of the analysed crystal structures no similarity of the docking poses representing the initial non-covalent complex with the covalent complex could be determined. However, this is only the case for a minority of investigated complexes in the test. Further analyses, in order to indentifiy common characteristics among these, were inconclusive as neither a protein, a class of proteins nor a warhead type is enriched in this set of analyzed crystal structures. A more decisive factor for these structure may have been the pursued rigid docking approach, in which the inherent flexibility and motion of proteins are not taken into account. But these could play a key role to accommodate the ligand prior to the covalent bond formation.

Nonetheless, for more than three-quarter of the test set protein-ligand complexes non-covalent docking poses were obtained through protein-ligand docking, which show high similarity to the covalent complex. This suggests that in general the binding mode of a covalent ligand does not change during covalent

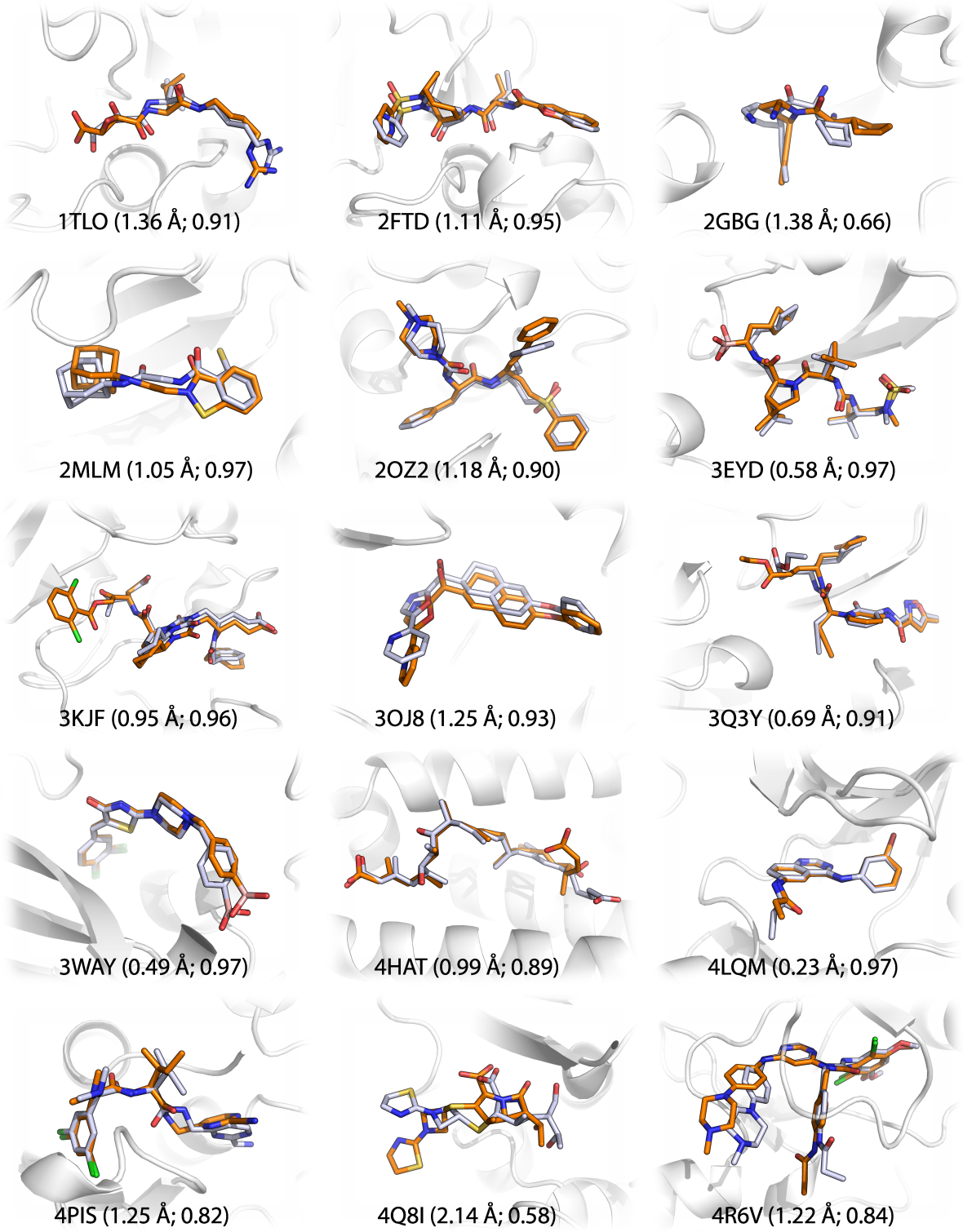

Figure 4.10: Recreated initial complex of covalent ligands. Selected docking pose of the putative initial complex (orange) superimposed to the covalently bound ligand in the respective crystal structure (lightblue) for a subset of the investigated protein-ligand complexes. The PDB ID of each crystal structure, the RMSD and the relative feature fit are given.

bond formation, which is further indicated by the generated poses for those ligands with spacious leaving groups or which undergo ring-opening reactions. These warheads, which would make alterations of the binding mode mandatory, are generally well accommodated in the active site. This exemplifies the specificity of covalent ligands through the accommodation and precise positioning of the warhead during the formation of the initial complex. This is therefore a specific characteristic of covalent ligands and can thus be utilized for their rational design.

With these results in mind, we were able to rationalize the binding of the templated assembly-based $3C^{pro}$ inhibitors, which initially prompted the investigations of the initial complex. The binding mode elucidation was further enabled by additional biological data obtained by Daniel Becker. At the one hand, protein-ligand MS of the covalently bound ligation product showed that the mass is higher than it would be for the assumed Schiff base. On the other hand, incubation of **1** and **2** in different solvents showed that ligation can occur at the carbonyl oxygen and the carbon-4 position in addition to the primary amine, which is in good agreement with the known reactivity of pyrazolones [214]. Based on these experimental findings possible ligation products were virtually generated and covalently as well as non-covalently docked into the $3C^{pro}$ active site (see appendix A.3). By using the 3D pharmacophore model for $3C^{pro}$ binding, the obtained docking poses were analyzed with respect to key interactions and accordance of the binding modes between the covalent and the non-covalent complex. For the secondary alcohol **4** binding conformations were retrieved, in which the pyrazolone forms the crucial hydrogen bonds in the S1 pocket in the initial complex (Figure 4.11A,B). The epoxide moiety is thereby placed in close proximity to the catalytic cysteine enabling covalent complex formation, in which the initial binding mode is maintained (Figure 4.11C). The computational simulations indicate that **4** is the most probable ligation product, which was further confirmed by the synthesis of the desoxy-analogue **5**, which also shows biological activity (Figure 4.11A). The binding mode of the potent Michael acceptor analogue **3** was determined through X-ray crystallography using its chemically more stable amide derivative. The crystal structure shows that the S1 pocket is occupied by a water molecule, which bridges a hydrogen bond from the ligand to the H161. The biological implications of this binding mode is uncertain as parts of the ligand electron density are not fully resolved.

The analysis of the covalent and the initial complex enabled the rationalization of the binding mode for a series of inhibitors developed by Daniel Becker through

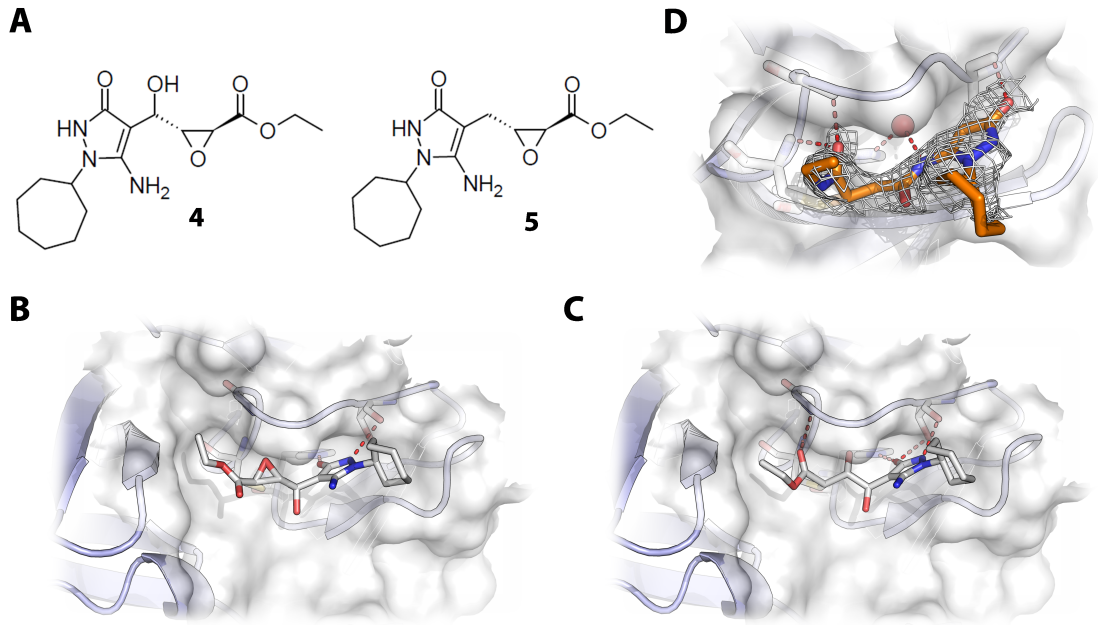

Figure 4.11: Dynamic ligation product and 3C^{pro} binding. (A) Chemical structure of the putative ligation product **4** and its synthesized desoxy-analogue **5**. Binding conformation of the putative ligation product **4** non-covalently (B) and covalently (C) bound to the 3C^{pro}. (D) Crystal structure of the amide analogue of **3** bound in the active site of the 3C^{pro} with its electron density shown as gray mesh. Hydrogen bonds are depicted as red dashed lines.

a templated assembly approach. Furthermore, the implicated conservation of the binding mode during covalent bond formation is a key characteristic of covalent ligands, which can be utilized for their rational design and it suggests the applicability of 3D pharmacophores derived from covalently bound crystal structures for this purpose.

4.3 Fragment screening

The 3D pharmacophore model, fostered by the deepened understanding of the covalent binding obtained from the systematic analysis, was utilized in the next step, the fragment screening. Here, we aimed for the identification of fragments as covalent inhibitors of the CVB3 $3C^{pro}$ through a combined computational and biological screening cascade. The following section describes the performed virtual screening of a fragment collection using the 3D pharmacophore model, which led to an enriched set of fragments and their subsequent biological evaluation in the $3C^{pro}$ enzyme kinetic assay.

4.3.1 Virtual fragment screening

The generated and validated 3D pharmacophore model represents the key interaction necessary for binding to the $3C^{pro}$. This enabled the aspired virtual screening-based pre-sampling of compound libraries to identify those fragments comprising the essential chemical functionalities to match the pharmacophore model and thus to be putatively active against the $3C^{pro}$. As source of fragments, the small-molecule collection of the group of Prof. Rademann was used. This library has not been purely assembled as a fragment collection but rather as set of drug-like small-molecules by applying a maximum common substructure approach [215]. Its primary purpose is the templated assembly approach, introduced in the previous section, which has been successfully applied for several targets by the group of Prof. Rademann [216–220]. To enable the use of the molecules collection for the pursued fragment screening approach a set of physicochemical and substructure filters were applied. Size (MW $\leq$ 325 Da, HA $\leq$ 20) and lipophilicity (cLogP $\leq$ 3) cut offs ensure fragment-likeness, whereas compounds containing substructures from known pan-assay interference compounds were removed (PAINS) by using substructure descriptions in SMARTS notation (see section 4.5) to prevent artificial assay read out [145]. These initial selection steps reduced to number of molecules from 4298 to 3004. Since we were aiming for covalently binding ligands, an additional filtering step was performed to eliminate fragments, which do not contain a warhead. This led to a final set of 397 fragments, which were subsequently used for the virtual screening.

The fragments were docked into the active site of the $3C^{pro}$, covalently as well as non-covalently in order to generate the putative initial and the covalent complex. The obtained poses were scored using the 3D pharmacophore model. Thereby,

fragments were identified, which are able to form the essential interactions with the $3C^{pro}$, covalently as well as non-covalently bound. This way both states occurring during the covalent binding of ligands were assessed and taken into account. Docking poses were obtained for 23 fragments, in which the initial and the covalent complex form the necessary interactions in the active site of the $3C^{pro}$ according to the 3D pharmacophore model. Visual inspection of the binding poses with careful investigation of the ligand geometries and interaction angles led to the removal of four fragments. The remaining 19 fragments represent 12 chemotypes. From each of these, a fragment was selected for biological testing (Figure 4.12). The final selection of fragments comprises especially ketone and nitrile moieties as warheads. The prevalence of these less reactive warheads reflects the drug-likeness of the prepared fragment collection. The crucial hydrogen bonds in the S1 pocket are formed by a variety of functional groups in the selected fragments, of which not all engage in all three hydrogen bonds to H161 and T142. This is due to the limited chemical space represented in the fragment library. However, we also want to assess the ability of different chemical groups, bioisosteric to the amide moiety of glutamine, to bind to the S1 pocket and thereby to increase the chemical diversity within the set of selected fragments.

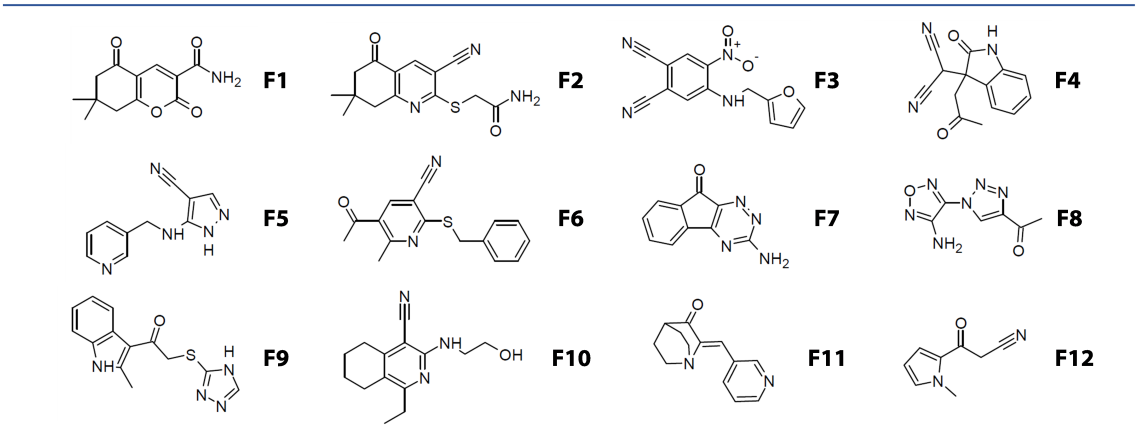

Figure 4.12: Virtual screening-based fragment selection for biological evaluation.

4.3.2 Biological evaluation & fragment hit analysis

In order to determine the biological activity of the selected fragments against CVB3 $3C^{pro}$, they were evaluated in a FRET-substrate based enzyme activity assay. At first, each fragment was measured at a single concentration to assess if it shows activity after all. At 650 µM, six fragments (**F1**, **F4**, **F6**, **F7**, **F8**, **F9**) inhib-

ited the $3C^{pro}$ by more than 30 % (Figure 4.13). These were further analyzed in concentration-response experiments. Only fragment **F1** showed reasonable dose-response characteristics with a hill slope close to unity and an apparent $IC_{50}{}^{*}$ of $\sim$ 350 µM (Figure 4.14A). In addition, the enzyme reaction progress curves level off, which indicated a time-dependent activity for the inhibition of the $3C^{pro}$ by **F1** (Figure 4.14B). This is a typical characteristic of covalent inhibitors as described in subsection 1.2.4 and a first indication that we have had identified a covalently binding fragment. From the predicted binding mode of **F1** and its chemical structure, we surmised a covalent interaction to most likely occur between the ketone moiety and the thiol of the catalytic C147, which would lead to the formation of a reversible thiohemiketal (Figure 4.14D-F). Therefore, equation 1.20 was used to calculate k_{obs} values from the enzyme reaction curves. The obtained plot of k_{obs} as a function of the concentration of **F1** shows a straight line, which does not intersects at the origin and yields an apparent K_i of 393 ± 17 µM by using equation 1.22 (Figure 4.14C).

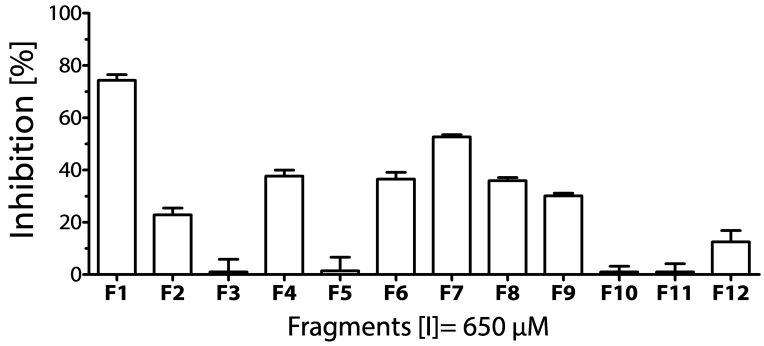

Figure 4.13: Inhibition of the $3C^{pro}$ by VS-based fragment selection. Activity of the fragments determined at a concentration of 650 µM.

Further analysis of the time-dependent activity of **F1** was performed through incubation experiments. Hereby, the $3C^{pro}$ was incubated with **F1** for different time periods prior to the start of the enzymatic reaction through the addition of the substrate. Time-dependent activity leads to an increased enzyme inhibition for each prolonged incubation period. The enzyme reaction progress curves for different incubation durations of **F1** are plotted in Figure 4.15A. At the beginning of the reaction, **F1** is more active with each increased incubation duration. In the course of the enzymatic reaction, the progress curve for the non-incubated sample levels off as observed before but for the incubated samples, the enzyme activity increases over time. This suggests that **F1** acts through a time-dependent mode of action to form an enzyme-inhibitor complex, which is putatively not stable and

thus decays. The two most plausible reasons are either a displacement of **F1** from the enzyme inhibitor complex through competition with the substrate or chemical instability of **F1**.

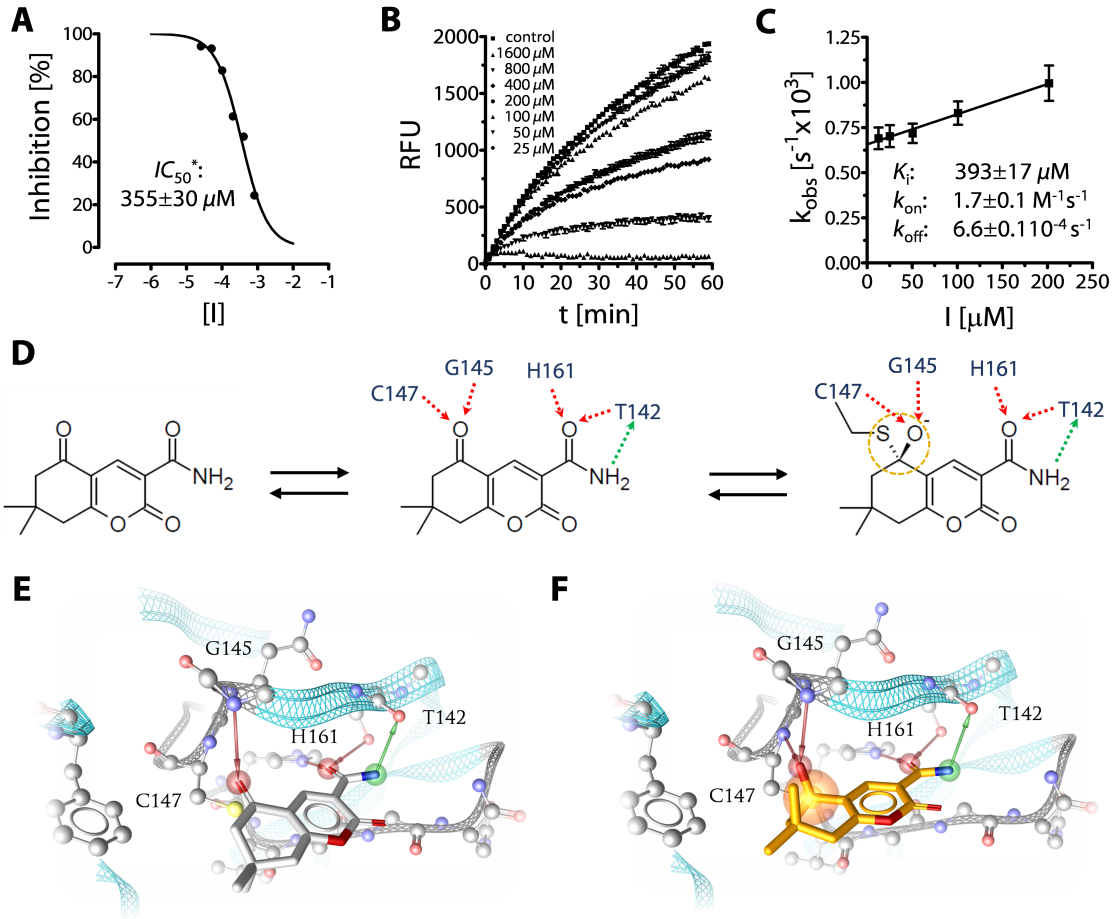

Figure 4.14: Biological activity and binding characteristics of fragment F1. (A) Dose-response curve for **F1**. (B) Progress curves for the enzyme reaction in the presence of different concentrations of **F1**. (C) Plot of k_{obs} as a function of the concentration of **F1** and kinetic values. (D) Different steps of the interaction of **F1** with the $3C^{pro}$ depicted in 2D. From left to right: unbound **F1**, initial complex and covalent complex. Interactions are colored as follows: Orange sphere indicate covalent binding, green and red arrows indicate hydrogen bond donor or acceptor functionalities. (E,F) Binding mode of **F1** in the initial complex and the covalent complex. Interactions are represented by pharmacophore features. An orange sphere indicates a covalent bond, red arrows hydrogen bond acceptors and green arrows hydrogen bond donors, respectively.

To investigate the stability of **F1**, it was incubated in the assay buffer. LC/MS measurements after representative time frames revealed an instability of **F1** (Figure 4.15B). Analysis of freshly prepared samples showed a single peak ($\geq$ 95 %), which had a tiny shoulder. This observation provided a first indication for the presence of a second compound. Interestingly, the shoulder grew over time, whereas the peak of **F1** diminished, which indicates a direct conversion of **F1** into another

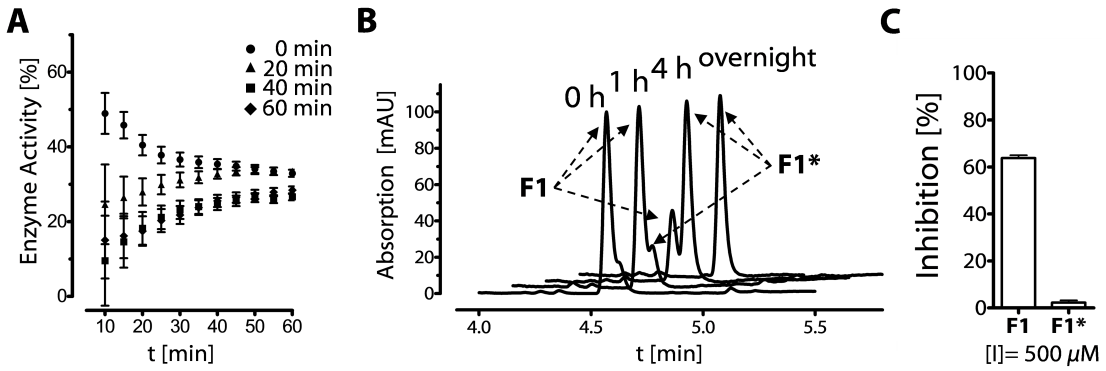

Figure 4.15: F1 time-dependent activity and stability analysis. (A) Relative enzymatic activity of the $3C^{pro}$ in the progress of the enzymatic reaction after incubation with **F1** at 500 µM for 60 min (diamond), 40 min (square), 20 min (triangle) or 0 min (circle). (B) Superimposed HPLC chromatograms of **F1** after incubation in assay buffer for 0 h, 1 h, 4 h or overnight. (C) Comparison of $3C^{pro}$ inhibition by **F1** and a equally concentrated sample of **F1** incubated in the assay buffer overnight, in which **F1** has been completely converted into the instability reaction product **F1***.

molecule, which is referred to as **F1***, rather than an unspecific decay. This conversion was completed after overnight incubation as LC/MS analysis showed only one peak with a retention time slightly larger than for **F1**. Enzyme kinetic analysis after conversion of **F1** to **F1*** showed no activity against the $3C^{pro}$, which indicates that this chemical instability is the reason for the decrease in activity detected in the incubation experiments (Figure 4.15C). This analysis showed that the activity measured in the assay did result from **F1**. However, the decrease in activity, detected through the incubation experiments, resulted from the conversion of **F1** into **F1*** in the aqueous assay buffer. This rendered **F1** less attractive as starting point for the design of $3C^{pro}$ inhibitors. But understanding of the instability reaction and the responsible chemical moiety, allows to search for analogues of **F1** in commercial fragment libraries, which are devoid of this weakness.

Therefore, the instability reaction was further investigated. The conversion of **F1** into **F1*** is slower in a $1:1$ mixture of water and acetonitrile used as LC/MS solvent, than in the aqueous assay buffer. This indicates a hydrolysis to take place. Apparently, the lactone of the tetrahydrobenzopyran-2,5-dione system in **F1** is the hydrolytically most vulnerable moiety. Additional information was drawn from the detected ions in the respective MS spectra. The major ion detected for **F1** in the positive mode had a mass to charge ratio (m/z) of 236.0934 corresponding to the $[M+H]^+$. Ions detected after overnight incubation had a m/z of 237.0751 in the positive mode. Furthermore, a high abundance was detected in the negative mode with a m/z of 235.0612 indicating the formation of an acid. The similarity of the

retention times as well as in the $\mathrm{m/z}$ for **F1** and **F1*** indicate a high degree of chemical similarity. The initially proposed structure for an instability reaction product

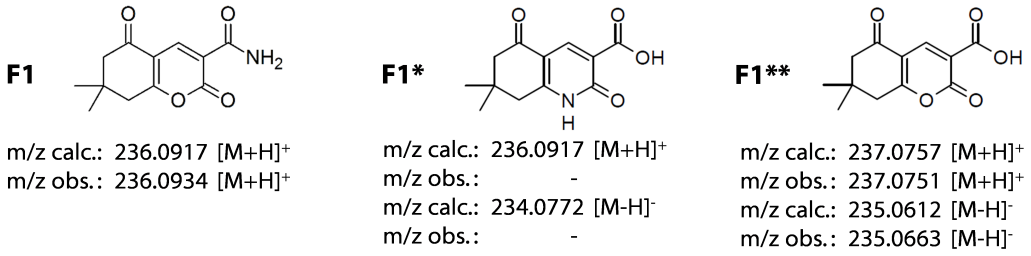

Figure 4.16: Putative chemical structures of F1 instability reaction products with calculated and observed m/z.

was the cyclic amide analogue (**F1***) (Figure 4.16). The greater hydrolytic stability of amides would explain the observed unidirectional conversion of **F1** to **F1***. A possible mechanism for the formation of the lactam is given in Figure 4.17A. Hydrolysis of the ester leads to the open-chain intermediate, which can exist in multiple tautomeric states. This enables the entropy-driven recyclization to occur via the primary amide to form the proposed lactam containing product **F1***. However, the calculated ions for the lactam do not fit those detected in the MS spectra as they are the same as for **F1** (Figure 4.16). In contrast, the negative and positive ions detected for the instability reaction product show a $\mathrm{m/z}$ increased by one. Therefore, the acid lactone (**F1****) was proposed as putative reaction product since the exchange of a hydroxyl group to a an amine moiety explains the mass difference. Consequently, the calculated $\mathrm{m/z}$ fit those detected in the MS spectra (Figure 4.16). A possible mechanism for the formation of **F1**** is given in Figure 4.17B. The direct hydrolysis of the primary amide is very unlikely. Therefore, the hydrolysis of the lactone is once more assumed as initial step. In contrast to the proposed formation of the lactam from the hemiaminal intermediate for **F1***, the hemiaminal proceeds to the lactone via the elimination of ammonia. This reaction appears also unidirectional as the hydrolysis of the lactone and subsequent recyclization leads to the same molecule. Furthermore, the proposed mechanism for the formation of **F1**** is consistent with the scheme proposed for the synthesis of **F1** as reported by Vodolazhenko *et al*. [221]. The synthesis starts with the condensation of an α-CH acidic ketone **1** with dimethoxymethyl-dimethylamine **2** (DMDMA), which is further reacted with an α-CH acidic nitrile **4** to give the enolate **5**. Subsequent cyclization leads to the iminolactone **6**, which is converted via the hemiaminal **7** under acidic conditions through deamination to the lactone

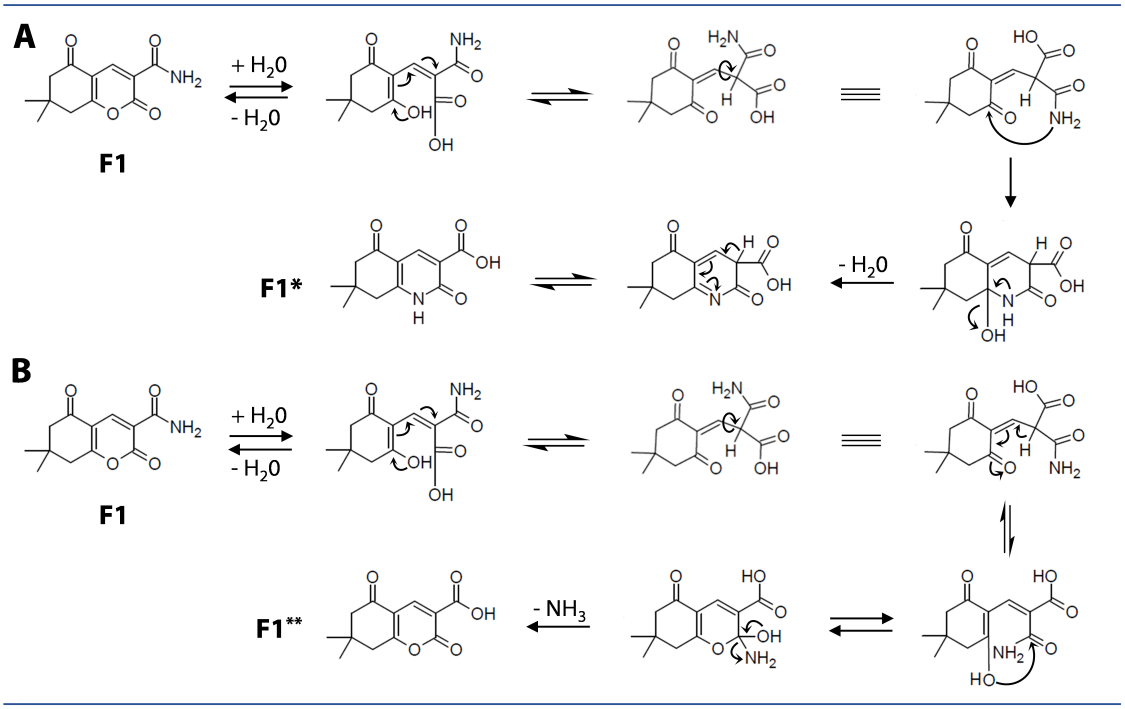

Figure 4.17: Proposed schemes for the instability reaction of F1. (A) Reaction scheme for the formation of the lactam **F1***. (B) Reaction scheme for the formation of the lactone **F1****.

8 (Figure 4.18). This step is similar to the proposed formation of **F1**** and provides an additional rationale for the proposed scheme of the instability reaction and the product **F1****.

The stability analysis revealed that **F1** is most likely instable due to its lactone moiety and is putatively converted in the aqueous assay buffer into its acid analogue **F1****. This provides an explanation for the abolished activity against the $3C^{pro}$ after incubation of **F1** as the carboxylic acid is less tolerated in the crucial S1 pocket reducing the activity by a factor of 50 [82, 99]. This instability further indicates that **F1** inhibits the $3C^{pro}$ specifically through the formation of a reversible covalent bond, as the $3C^{pro}$ activity increases alongside the decay of **F1**. The reversibility was further confirmed through protein MS analysis of **F1** and the $3C^{pro}$ as no irreversible adduct could be identified.

From the biologically evaluated fragments, **F1** is the only one, which was proposed to bind to the crucial S1 pocket via an amide moiety, with exception to **F4**. The other fragments may be also able to form hydrogen bonds to H161 and T142 but the missing biological activity indicates that the S1 pocket is exceptionally discriminative for the amide moiety. This is also apparent from the comparison of **F1** and **F2**, which are structurally similar but the inactive **F2** comprises a nitrile moiety instead of the amide in **F1**. Furthermore, the distance between the

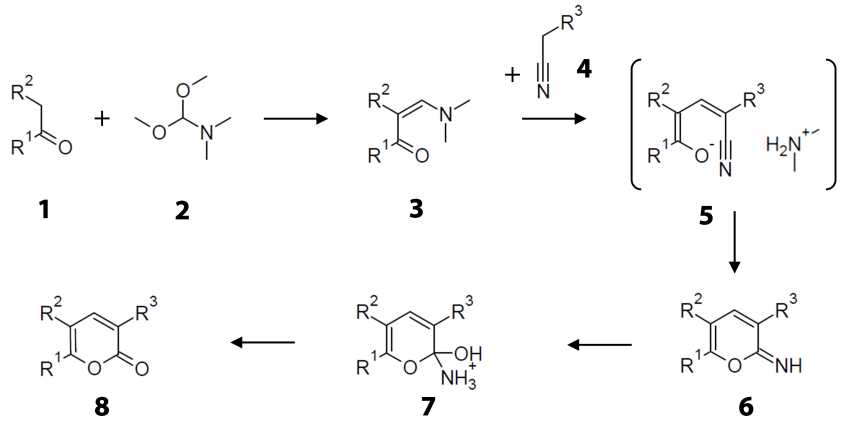

Figure 4.18: Chemical synthesis procedure for F1 scaffold as reported by Vodolazhenko et al. [221].

carbonyl carbon of the covalently binding ketone moiety of **F1** and the carbonyl carbon of its amide moiety is five single atoms in a linear connection. This is also true for the scaffolds of the active set molecules used for the 3D pharmacophore model validation and for the P1 glutamine. This distance descriptor represents the necessary spatial arrangement of the two key moieties for $3C^{pro}$ binding and the favorable conformational alignment of them. In **F1**, the covalently binding ketone moiety and the hydrogen bond forming amide group are combined within the rigid tetrahydrobenzopyran-2,5-dione scaffold, which renders it an attractive structure for a $3C^{pro}$ inhibitor, despite its chemical instability. A second drawback of the scaffold is the lack of an apparent synthetic handle for fragment optimization. Based on the proposed synthesis route, only a few variations, marked by R-groups, are accessible (Figure 4.18). This include the essential ketone and amide positions, which have to be untouched to preserve the proposed binding mode. Thus, little room for variations and optimization remains.

4.4 Scaffold hopping

The fragment screening for $3C^{pro}$ inhibitors through a combined computational and biological approach led to the identification of the fragment hit **F1**. The performed biological investigations provide evidence for the proposed covalent binding mode in the $3C^{pro}$ active site by mimicking the P1 glutamine residue. However, the chemical instability rendered **F1** unattractive as starting point for further optimization. Yet, its structural properties serve as basis for further exploration. This was performed through a computational scaffold hopping approach described in this section. Libraries of commercially available fragments were virtually screened. Fragments of promising scaffolds were selected for biological evaluation.

4.4.1 Fragment hit exploration

Fragment hit **F1** provides an attractive structural basis for $3C^{pro}$ inhibitors despite its apparent drawbacks, (i) chemical instability and (ii) lack of synthetic handle. Therefore, a search in commercially available fragment space was pursued to identify analogue scaffolds, which are devoid of the structural issues of **F1** and will serve as starting point for fragment optimization. Our in-house virtual libraries of commercially available compounds containing more than five million molecules were used. At first, the same physicochemical and substructure filters were applied as for the fragment library generation described in the previous section. This led to a reduced set of around 500,000 fragment-like molecules. The elucidated descriptor for the arrangement of the ketone and amide moiety in **F1** as key elements for $3C^{pro}$ binding was used to further filter the assembled fragment collection. Therefore, this substructure was described by a SMARTS-pattern (see subsection 3.1.5). Only 80 molecules remained after the final filtering step. These comprise the essential substructure and represent six distinct scaffolds: (i) tetrahydroquinoline-2,5-diones, (ii) tetrahydroindole-4-ones, (iii) phenones, (iv) 2-aminoindolizines, (v) five-membered heterocycles and (vi) pyridines (Figure 4.19). The fragments were non-covalently and covalently docked into the active site of $3C^{pro}$. The 3D pharmacophore model was used to evaluate the obtained docking poses and to assess the ability of the fragments to bind in accordance to **F1**. For fragments of three scaffolds, suitable docking poses for the covalent and the initial complex were generated. Hereby, the crucial hydrogen bonds in the S1 pocket were formed by the amide moiety, whereas the ketone is placed

in close proximity to the cysteine thiol (initial complex) or is covalently bound by forming a thiohemiketal (covalent complex). These fragments belong to the tetrahydroquinoline-2,5-dione, the five-membered ring heterocycle or the phenone scaffold. The latter contains three molecules, in which the ketone and the amide moiety are either connected by an alkyl chain or a benzene ring. The former implies high degree of conformational freedom whereas the latter is highly similar to the benzamide-based $3C^{pro}$ inhibitors, which were used for 3D pharmacophore model validation. Thus, this scaffold is already known and not attractive. In contrast, the other two scaffolds were further explored to selected suitable fragments for purchase and biological evaluations.

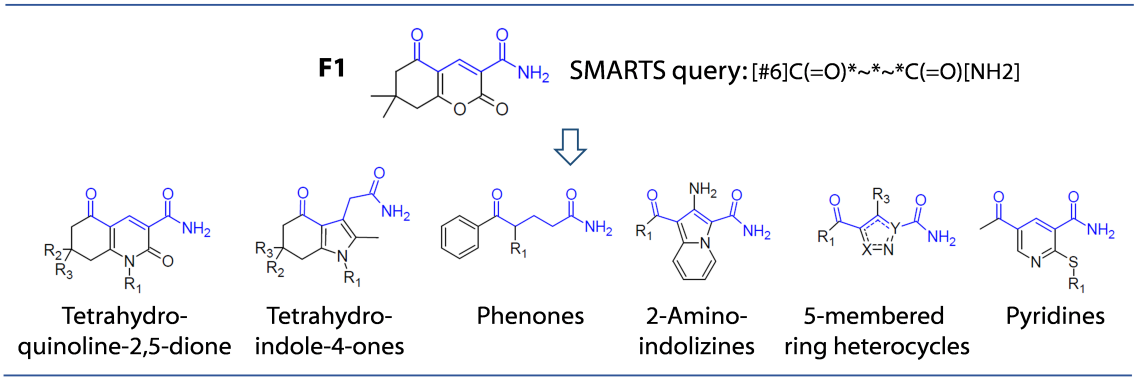

Figure 4.19: Scaffold hopping hit scaffolds. (Top) Chemical structure of **F1** and SMARTS-based substructure pattern. (Bottom) Identified scaffolds from the substructure-based search in commercial fragment space. Key substructure is colored in blue.

4.4.2 Quinoline scaffold exploration

At first, the fragments of the tetrahydroquinoline-2,5-dione scaffold were investigated more closely. They share a chemical substructure, 7,8-Dihydro-1H,6H-quinoline-2,5-dione. An amide moiety at position 3 enables the formation of the S1 pocket hydrogen bonds and thus $3C^{pro}$ binding in a similar way as **F1**. The only difference to the tetrahydrobenzopyran-2,5-dione scaffold is the substitution of the lactone to a lactam moiety. This make the tetrahydroquinoline-2,5-diones more attractive than **F1** in two ways. First, the lactam should be more stable against hydrolysis and secondly, the nitrogen can bear an additional substituent and thus offer a synthetic handle for optimization with an attractive vector toward the S2 pocket of the $3C^{pro}$ active site (Figure 4.20A). These characteristics render the quinoline-based fragments promising. In order to select fragments for purchase and subsequent biological evaluation, the commercially available frag-

ments of this scaffold were analyzed more closely. The major variations include N1-substitution, primary amide substitution and C-7 substitution (Figure 4.20A). Several compounds contain phenyl subtituents at the lactam nitrogen, which putatively provide additional occupation of the S2-pocket and thus some initial optimization. Substitution of the primary amide is clearly unfavorable as it would disturb the crucial S1 pocket hydrogen bonding due to the dominance of the *cis*-conformation in non-cyclic amides. However, an interesting exception to this was found, in which the amide was exchanged for a pyrazolone system annelated to the pyridine. The pyrazolone as putative amide bioisosteric moiety is also present in the template assembly based $3C^{pro}$ inhibitors described in section 4.2. The car-

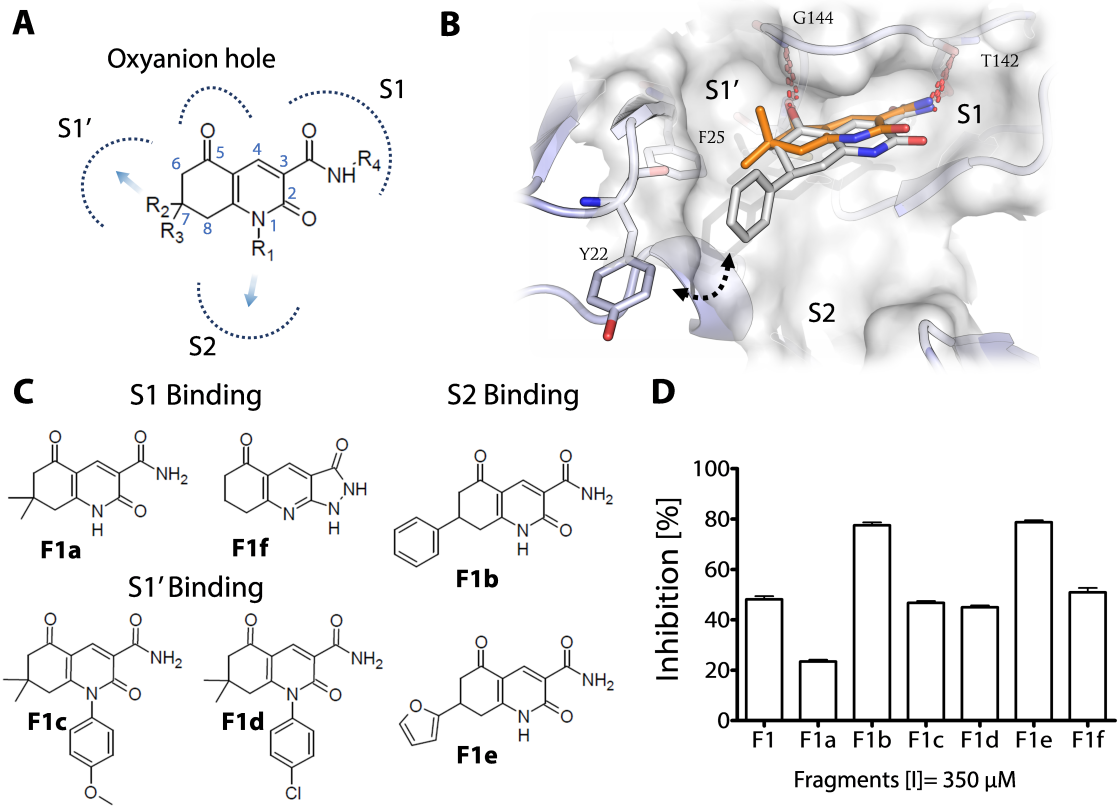

Figure 4.20: Tetrahydroquinoline-2,5-dione scaffold-based F1 analogues. (A) 2D depiction of the tetrahydroquinoline-2,5-dione scaffold in the active site of the $3C^{pro}$. Available variations are indicated by R-groups and putative subpocket occupation potential. (B) Docking poses of 7-phenyl substituted quinolin-2,5-dione (white) and 7-dimethyl-quinolin-2,5-dione (orange) bound in the active site of the $3C^{pro}$. The arrow indicates a possible steric clash with Y22. (C) Chemical structures and IDs of tetrahydroquinoline-2,5-diones selected for biological testing. (D) Inhibition of the $3C^{pro}$ by the selected fragments at 350 µM.

bon at position 7 is most frequently substituted with two methyl groups similar to **F1** but also mono- and unsubstituted analogues are available. Small substituents such as the methyl group do not affect binding, but larger and rigid substituents

could. Docking poses of quinolines with aromatic substituents like benzene indicated steric clashes with the S2 pocket forming residues, especially the side chain of Y22. In the low energy conformation used as input for the docking, the phenyl substituent is in an equatorial position and thus in the same plane as the quinoline ring. This makes the molecule too long to fit into the active site properly, especially for covalent bond formation (Figure 4.20B). However, the partly saturated quinoline ring allows for multiple conformations, which were generated by using CORINA and evaluated for their ability to bind to the $3C^{pro}$ active site (Figure 4.21). The obtained docking poses revealed that the molecule can fit nicely with its phenyl substituent placed in the S1' pocket in a half-chair conformation. The axial position of the substituent may be energetically less favorable but due to the apparent good occupation of the S1' pocket the respective fragments were also included in the final selection.

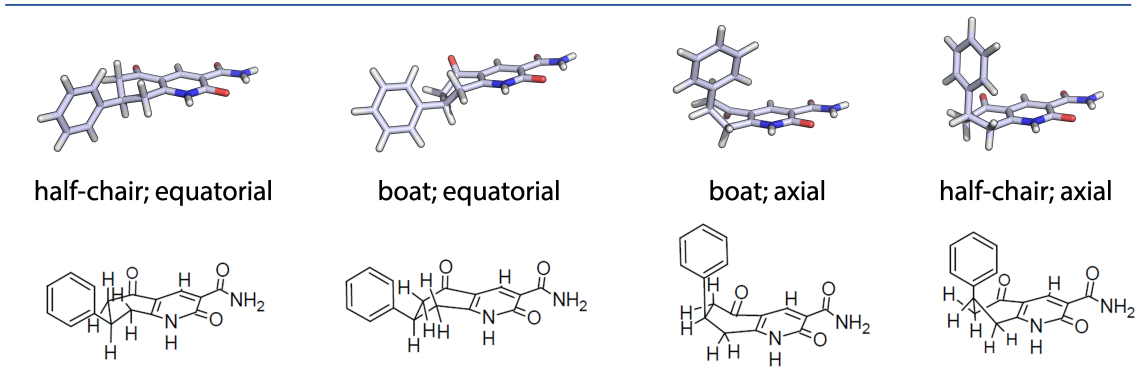

Figure 4.21: Conformational exploration of the 7-phenyl substituted tetrahydroquinoline-2,5-dione derivative in 3D (top) and 2D (bottom).

For each of the described favorable variation of the tetrahydroquinoline-2,5-dione scaffold representative fragments were selected and purchased if available, which led to a final set of six fragments (Figure 4.20C). These were biologically evaluated for $3C^{pro}$ inhibition. At first, single dose experiments were performed and revealed biological activity for all of the fragments (Figure 4.20D). However, **F1a**, the aza-analogue of **F1**, is significantly less active than **F1** as it only shows $\sim$ 25 % inhibition at 350 µM, which indicates that this scaffold is in general less active than the tetrahydrobenzopyran-2,5-dione of **F1**. The difference in the scaffolds may only be in one position but the exchange of the oxygen for the nitrogen could affect the ketone moiety as it is vinylogous to the quinoline nitrogen so that it has a slight amide-like character, which would make it less reactive. For the fragment **F1f** this is not possible due to the annelated pyrazolone ring, which leads to an imine nitrogen vinylogous to the ketone moiety. This may explain the higher

activity of **F1f** in comparison to **F1a**, which is similar to the initial hit **F1**. The nitrogen substituted quinoline derivatives **F1c** and **F1d** are active in the similar activity range, whereas the fragments **F1a** and **F1e** show more activity than **F1**. This indicates that substitution in position 7 is more favorable than at the quinoline nitrogen, which was unexpected as the substituted fragments are of similar size and the structural model suggested that the substituents at C-7 have to be in the energetically less favorable axial orientation for $3C^{pro}$ binding.

In order to elucidate the structure activity relationship (SAR) of the fragments additional biological evaluations were performed (Table 4.5). The dose-response experiments of **F1b** and **F1e**, which have a phenyl, respectively a furyl substituent at C-7, show reasonable concentration-dependent activity in the high two-digit micromolar range and indication of time-dependent activity as the progress curves show the characteristic flattening (Figure 4.22A-C). The pyrazolone **F1f** has an IC_{50} of 408 ± 41 µM, which is similar to the initial hit **F1** and further confirms the potential of the pyrazolone as amide mimetic for S1 pocket occupation (Figure 4.22D). The data points obtained for the fragments **F1c** and **F1d** do not cover the full dose-response curve due to their limited activity as well as solubility issues at higher concentrations (Figure 4.22E,F). Consequently, a poor fit to equation 1.16 led to an IC_{50} of 530 ± 64 µM for **F1c** and 592 ± 92 µM for **F1d**. In addition, their enzyme reaction progress curves do not level off, which indicates no time-dependent activity.

Table 4.5: Substitution pattens and inhibition values for quinoline-based fragments. R-groups are numbered according to Figure 4.20. Substituents are denoted as follows: Ph for phenyl, Me for methyl, Fu for Furyl. Time-dependency indicates an increase in activity over time. Values marked by 'nd' have not been determined due to limited activity of the respective fragment.

ID	R_1	R_2	R_3	R_4	IC_{50} [µM]	time-dependency
F1a	H	H	CH_3	CH_3	nd	nd
F1b	H	H	C_6H_5	H	84 ± 5	yes
F1c	4-CH_3-O-C_6H_5	H	CH_3	CH_3	530 ± 64	no
F1d	4-Cl-C_6H_5	H	CH_3	CH_3	592 ± 92	no
F1e	H	H	C_4H_3O	H	109 ± 3	yes
F1f	-	NH	H	H	408 ± 41	yes

Time-dependency of the fragment activity was further investigated through incubation experiments. The progress curve of **F1b** shows reduced activity of the $3C^{pro}$ with increased incubation time (Figure 4.23A,B). On the other hand, no significant difference is apparent for the different incubation times for **F1c** with respect to assay variance, which confirmed the time-independent activity (Fig-

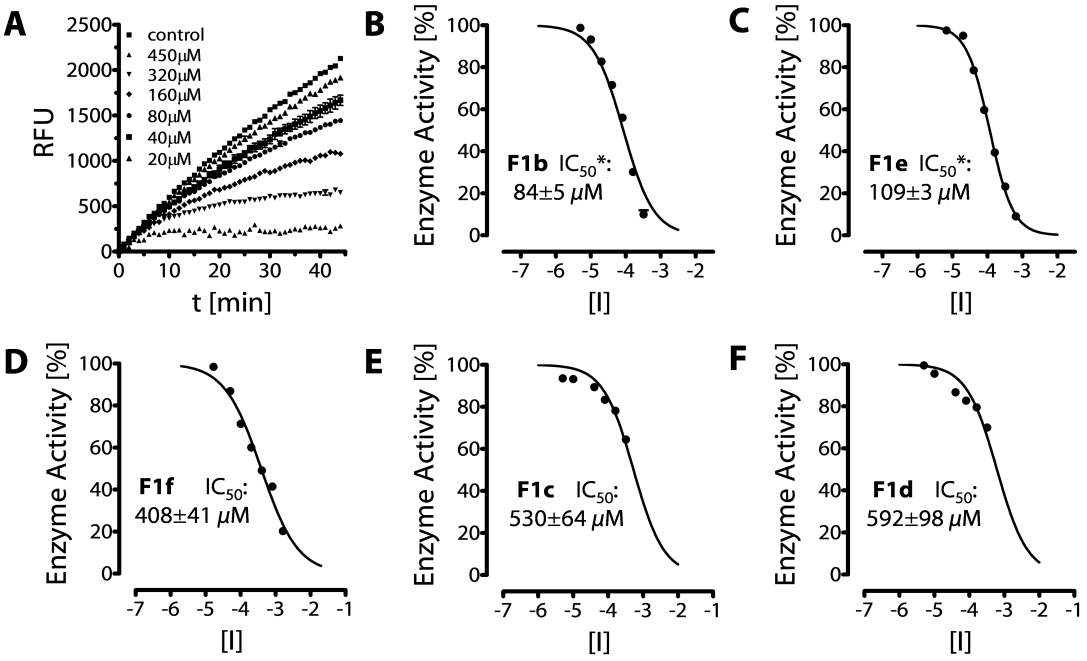

Figure 4.22: Activity of quinoline-based fragments against CVB3 3C^{pro}. (A) Progress curves for the enzymatic reaction in the presence of different concentrations of **F1b**. (B-F) Dose-response curves and IC_{50} for **F1b**, **F1e**, **F1f**, **F1c** and **F1d**, respectively.

ure 4.23A,B). From these results, it is apparent that the position of substitution not only affects the binding activity but also the covalent interaction with the 3C^{pro}.

In order to rationalize these experimental observations, the structural binding models that led to the selection of the fragments were reinvestigated with emphasis on the comparison of the initial and the covalent complexes as this is the point in which the fragments differ according to the obtained enzyme kinetic data. **F1c** and **F1d** only form the initial complex, whereas **F1b** and **F1e** proceed to the covalent complex as indicated by the time-dependency of their activity against 3C^{pro}. We surmised that the non-covalent binding of **F1c** and **F1d** could have two possible reasons: Either the binding conformation is different to the proposed one, so that the ketone moiety is not in spatial proximity to facilitate covalent binding or the formation of the covalent complex is energetically unfavorable. The ability of the selected tetrahydroquinoline-2,5-dione fragments to adhere to the 3D pharmacophore model comprising the key interactions for 3C^{pro} binding provides sufficient rationale for the proposed binding conformations. Yet, without experimental verification through X-ray crystallography uncertainty remains. The docking poses for **F1c** and **F1d** indicate that their substituents occupy the S2 pocket not as well as initially suspected. This is due to the aromatic nature of the substituents

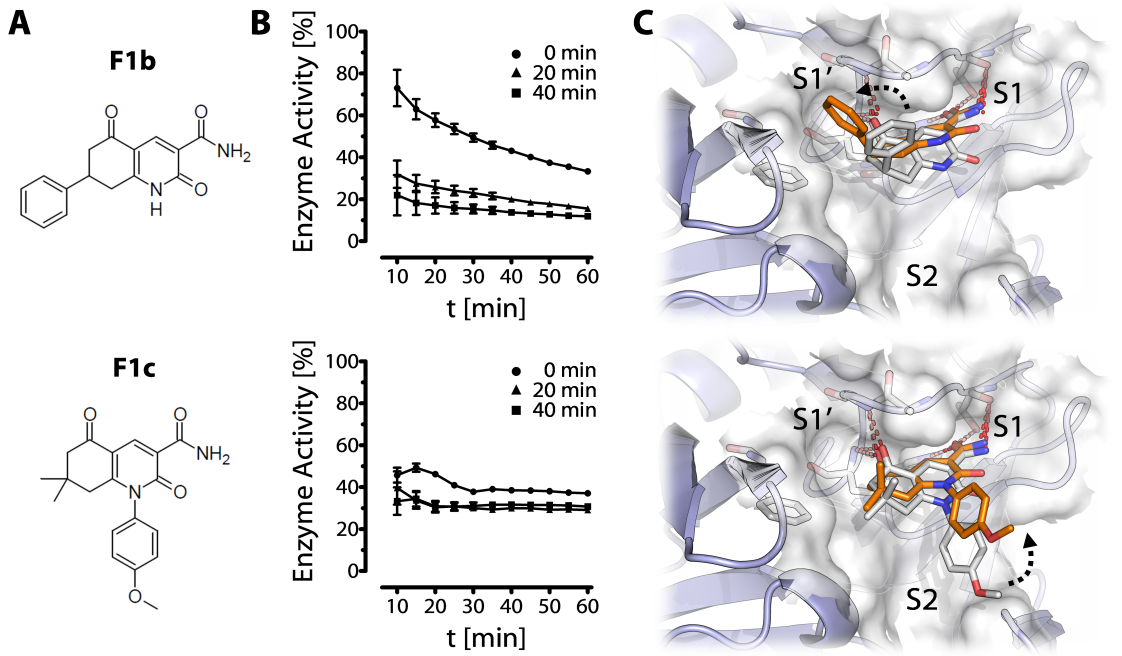

Figure 4.23: Time-dependent activity of quinoline-based fragments. (A) Chemical structures of **F1b** (top) and **F1c** (bottom). (B) Relative enzymatic activity of the $3C^{pro}$ in the progress of the enzymatic reaction after incubation with **F1b** (top) or **F1c** at 200 µM and 300 µM, respectively, (bottom) for 40 min (square), 20 min (triangle) or 0 min (circle). (C) Putative binding modes of **F1b** (top) and **F1c** (bottom) for the initial (white) and the covalent complex (orange) in the active site of the $3C^{pro}$. Conformational differences between the initial and the covalent complex are indicated by a dashed arrow. Key interactions are indicated by red colored dashed lines.

and the resulting planar conformation of the fragments. But for the sufficient occupation of the S2 pocket, a tilted orientation of the substituents is required. Therefore, the phenyl substituents of **F1c** and **F1d** are not well buried in the $3C^{pro}$ active site (Figure 4.23C). Comparison of the proposed conformations for the initial and the covalent complex showed that the substituents are even more solvent exposed in the latter as the covalent bond formation induces a perpendicular orientation of the quinoline ring to the active site. The entropic loss due to the solvent exposure could circumvent the enthalpic gain of the covalent bond formation and thus provides an explanation for the biological data. In contrast, **F1b** and **F1e** can bind in the half-chair conformation in both states in the active site of the $3C^{pro}$, whereas the substituents occupy the S1' pocket apparently better in the covalent complex (Figure 4.23C).

The surmised chemical stability of the tetrahydroquinoline-2,5-dione was suggested by the incubation experiments as no reduction in activity was detected as for **F1**. This was further confirmed by long-time incubation in the assay buffer and subsequent LC/MS analysis (Figure 4.24). After 24 h, no reduction in the

fragment concentration could be detected. Interestingly, a small impurity (< 5 %) was identified. The m/z of the respective ions indicate that this impurity is the acid analogue as suggested for **F1**. However, for the quinoline-based fragments no conversion can be detected, which confirms their chemical stability. The presence of the acid analogues results from the synthesis, which is done in analogy to the scheme reported for the synthesis of **F1** [221].

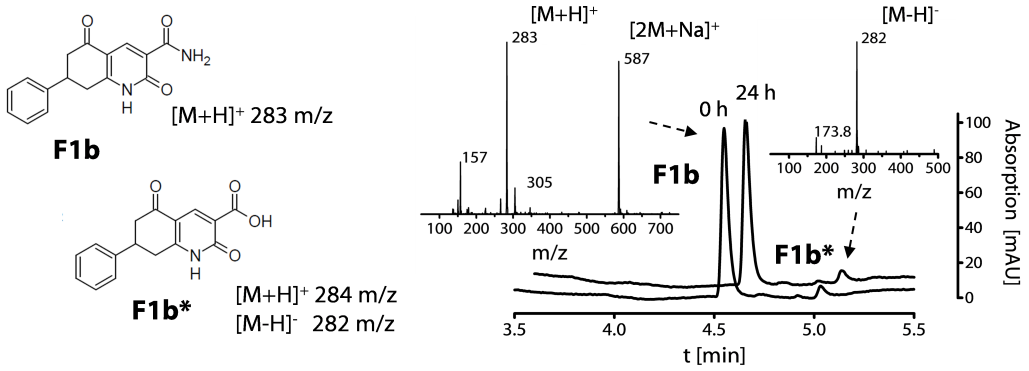

Figure 4.24: Chemical stability of quinoline-based fragments. Representative stability results for **F1b**. Chemical structures and calculated m/z for relevant ions for **F1b** and acid analogue **F1b***. Superimposed HPLC chromatograms of **F1b** after incubation in assay buffer for 0 h and 24 h. MS spectra for the main peak in the positive mode and the small impurity in the negative mode are labeled with the respective ion species.

The biological activity of the fragments as well as the verification of the assumed chemical stability renders this scaffold attractive for further optimization. Especially the fragments **F1b** and **F1e**, which are five times more active than **F1** with equal ligand efficacy, were promising. The substituents of these fragments putatively occupy the S1' pocket, which leaves room for optimization at the more important non-primed site of the $3C^{pro}$. However, this made **F1f** less attractive, despite its activity similar to **F1**, as the lack of the lactam moiety diminishes the potential for optimization to address the non-primed active site part.

4.4.3 Pyrrole-based fragments

The second scaffold, which resulted from the exploration of hit fragment **F1**, comprises several five-membered ring heterocycles, like pyrrole, pyrazole and isoxazole. These contain the essential amide and ketone moieties in a meta-substitution pattern, which enables them to adopt the proposed binding mode for $3C^{pro}$ inhibition (Figure 4.25A). Substituents of the ketone moiety could occupy the S1' pocket, which was shown to be favorable by the biological evaluation of the quinoline-based fragments. From the initially identified fragments adhering to this scaffold,

seven were commercially available and thus were purchased for biological evaluation (Figure 4.25B). Single concentration measurements revealed activity at 650 µM for only two fragments, the isoxazole **C1** and the pyrrole **C5** (Figure 4.25C). The fragments **C3**, **C4** and **C6** only show limited inhibition ($\leq$ 20 %), whereas the fragments **C2** and **C7** could not be evaluated due to fluorescence and solubility issues, respectively. Additional characterization of **C1** showed that its mode of action is different to what has been described so far and what was the initial aim of this project. Therefore, **C1** and the additional investigations that it triggered will be described separately in chapter 5 of this thesis.

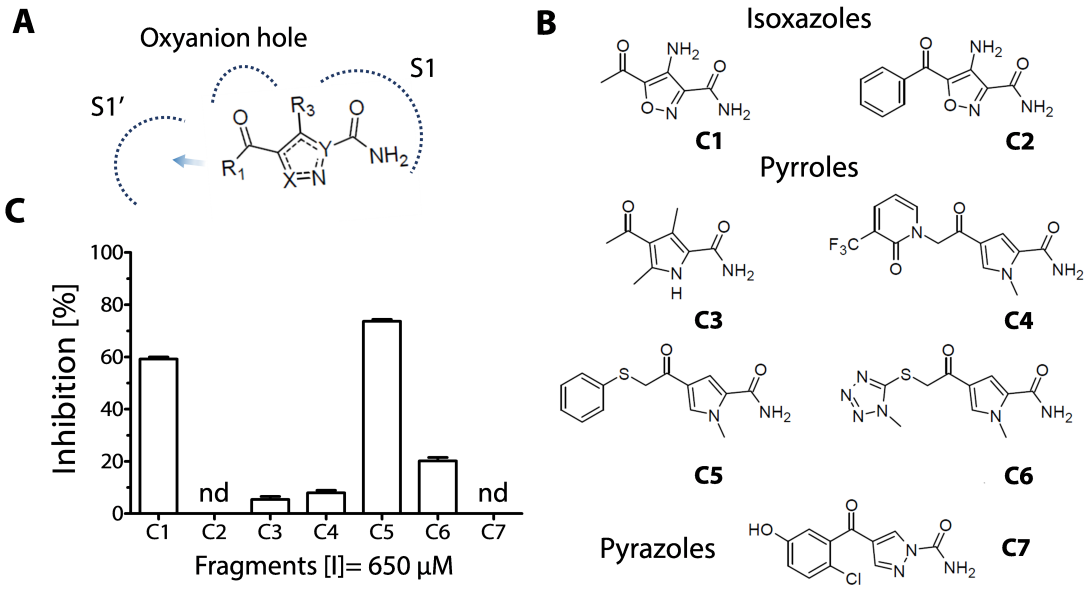

Figure 4.25: F1 analogues with five-membered ring heterocycle scaffold. (A) 2D depiction of the scaffold in the active site of the $3C^{pro}$ with different variations and subpockets indicated. (B) Chemical structures and IDs of selected fragments for biological testing. (C) Inhibition of the $3C^{pro}$ by the selected fragments at a single concentration. Activity of fragments marked by nd were not determined due to fluorescence and solubility issues, respectively.

Further characterization of fragment **C5** was done due to the promising biological activity but also to investigate the apparent activity cliff for the pyrrole-based fragments. Dose-response measurements showed reasonable concentration and time-dependent activity. The apparent IC_{50}* was determined to be 127 $\pm$ 13 µM and thus in similar range as the most active quinoline-based fragments (**F1b**, **F1e**) (Figure 4.26A). The time-dependency of **C5** was further confirmed by incubation experiments, which show the characteristic enzyme reaction progress curves (Figure 4.26B). Determination of the activity for slow binding inhibitors showed better fit by using equation 1.24 for irreversible binding and the plot of k_{obs} as a function

of **C5** concentration also proceeds through the origin with a k_{off} equal to zero (Figure 4.26C). This was unexpected as **C5** and the other selected fragments only contain a ketone moiety for covalent interaction with the $3C^{pro}$. Thus, the putative irreversible mode of action of **C5** provides an explanation for the difference in biological activity but not necessarily due to a specific mode of action as it does not contain an apparent irreversibly binding warhead.

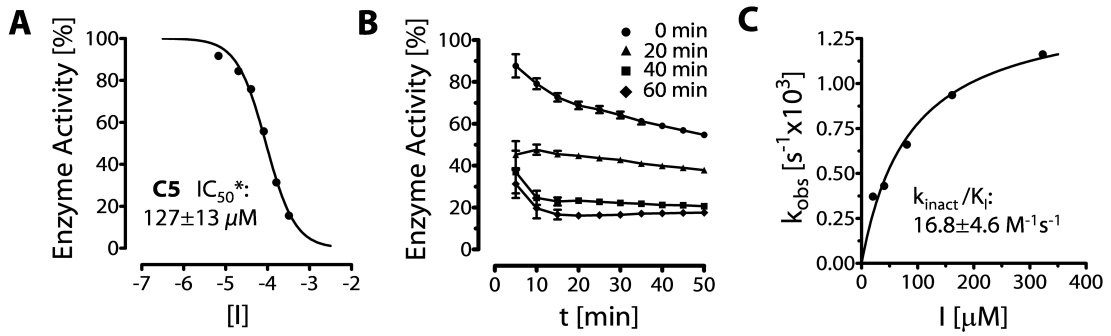

Figure 4.26: Biological characterization of C5. (A) Dose-response curve for **C5** with IC_{50}* value. (B) Relative enzymatic activity of the $3C^{pro}$ in the progress of the enzymatic reaction after incubation with **C5** at 50 µM for 60 min (diamonds), 40 min (square), 20 min (triangle) or 0 min (circle). (C) Plot of k_{obs} as a function of **C5** concentration and inactivation rate.

Protein mass spectrometry was performed to shed light on the mode of action of fragment **C5**. The obtained mass spectra were deconvoluted, which revealed three components (Figure 4.27A). One with the mass of the apo $3C^{pro}$ (21456 Da) as determined by control measurements and two components with greater masses, 21620 Da and 21562 Da, respectively. Whereas, the former shows the highest abundance and the latter the least of the three. But the mass shift for neither of them equals the molecular weight of **C5**, which nourished suspicion concerning the specificity of inhibition. However, we were able to propose a putative mode of action through detailed analysis of the structural differences of the fragments and the increase in mass for the different adducts detected. The major difference to the other fragments of this series is the substituent of the α-methyl ketone moiety, which is a thiophenol group in **C5**. This has a molecular weight of $\sim$ 109 Da, which is the mass shift detected for the minor adduct. Furthermore, the major component in the LC/MS spectrum has an addition in mass of 165 Da, which equals the fragment **C5** mass (275) except the 109 Da of the thiophenol. Based on this observation, we reasoned that **C5** may react similar to other α-substituted methyl ketones, such as α-halomethyl ketones, which are known as covalent modifiers (section 1.3) according to the following mechanism (Figure 4.27B). The interac-

tion of **C5** with the 3C^{pro} leads initially to a reversible covalent complex through the formation of a thiohemiketal by the ketone moiety as described before, which can further react through the elimination of the thiophenolate. This is stabilized through the formation of an intermediary thiiranium ring, which rearranges to a thioether of the cysteine at the α-methyl ketone of **C5**. This irreversible reaction product has the same addition in mass as the major product (+165). In addition, the eliminated thiophenolate can react with the cysteine through the formation of a disulfide, which has the addition in mass detected for the minor product (+109). The exact mass of the apo 3C^{pro} and the protein-ligand adducts were determined through high resolution mass spectrometry and confirmed the proposed reaction mechanism (Figure 4.27C).

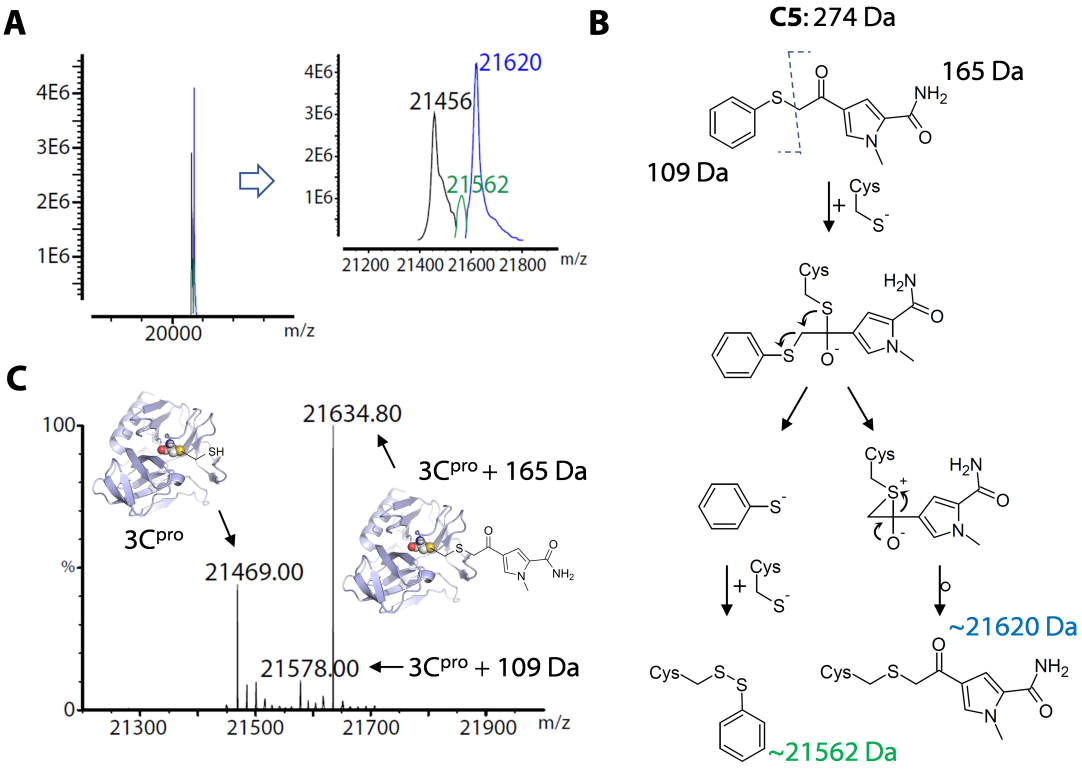

Figure 4.27: Protein mass spectrometry for fragment C5. (A) Deconvoluted ESI mass spectrum of the 3C^{pro} after incubation with **C5**. (B) Putative reaction mechanism for the formation of the 3C^{pro}-**C5** adducts. (C) Deconvoluted high-resolution MALDI-TOF spectra of the 3C^{pro} after incubation with **C5**. The main peak of 21634.8 Da corresponds to the schematically depicted thioether adduct formed after reaction of **C5** with the 3C^{pro}.

The MS data and the derived mode of action provide a rationale for the determined biological activity of **C5** as an irreversible 3C^{pro} inhibitor as well as for the difference in the fragment series. However, it was uncertain in how far the structural model, which led to the selection of the fragment, was still valid. Therefore,

the proposed reaction mechanism for the covalent interaction of **C5** with the $3C^{pro}$ was modeled and compared to the initial model (Figure 4.28). At first, the non-covalent complex is formed in which the amide moiety of **C5** forms the crucial hydrogen bonds to H161 and T142 in the S1 pocket. The thiophenol group is nicely accommodated in the S1' pocket and the ketone moiety is placed in close proximity to the catalytic C147. This enables the subsequent formation of the covalent bond through the nucleophilic attack of the cysteine thiol at the ketone carbon, which leads to the thiohemiketal intermediate stabilized by the hydrogen bonds in the oxyanion hole. To this point, the initially proposed binding mode was in accordance to the biological data but it did not contain the subsequent elimination of the thiophenol suggested by the mass spectra and irreversible binding kinetic. The α-methyl thioether with C147, formed after the elimination of the thiophenol, is still capable of forming the crucial hydrogen bonds in the S1 pocket. This indicates that the binding of **C5** is in accordance to the generated binding model for $3C^{pro}$ inhibition and it shows the typical binding mode conservation during covalent bond formation. But it also exemplifies the challenges in the rational design of covalent inhibitors as reactivity is difficult to predict.

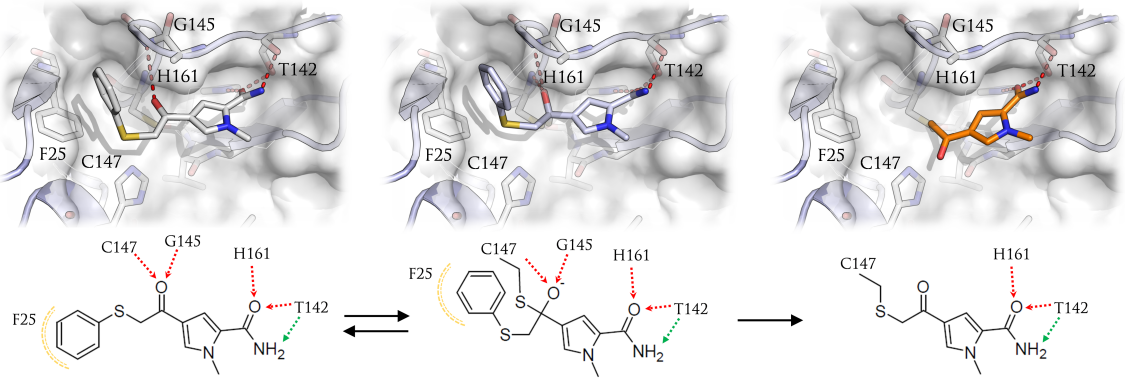

Figure 4.28: Proposed covalent binding mode of C5. 3D (top) and 2D (bottom) depictions for the different steps of the interaction between **C5** and the $3C^{pro}$. Interactions are described as pharmacophore features. Yellow spheres indicate lipophilic contacts, orange sphere indicates covalent bond formation, green and red arrows indicate hydrogen bond donor or acceptor functionalities.

The biological evaluation of the second scaffold revealed that the comprised fragments are hardly active. Exceptions are only **C1** (see chapter 5) and **C5**, which is active with an apparent $IC_{50}{}^{*}$ around 125 µM and a k_{inact}/K_I value of 16.8 $\pm$ 4.6 $M^{-1}s^{-1}$. The biological activity results from an interesting irreversible mode of action through its α-thiophenolmethyl ketone moiety unknown at that time. Choong *et al.* have reported α-thiomethyl ketones as covalent inhibitors for caspase-3, but these bind reversibly as determined by X-ray crystallography [222]. The

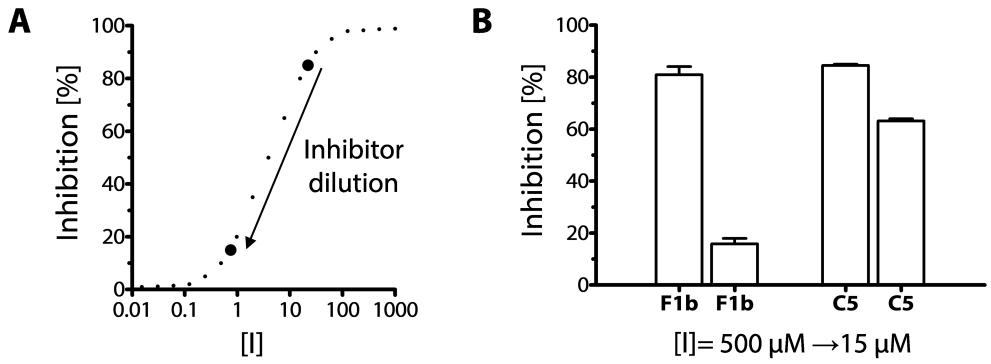

Figure 4.29: Reversibility of 3C^{pro} inhibition. (A) Schematic depiction of the effect of dilution on enzyme inhibition for an ideal inhibitor. (B) 3C^{pro} inhibition by **F1b** and **C5** before and after dilution.

biological activity similar to the most active quinoline-based fragments but especially the interesting mode of action render **C5** attractive as starting point for optimization.

The different covalent mode of actions displayed by the two identified scaffolds were further confirmed through dilution experiments. Hereby, the enzyme is incubated with the inhibitor at a concentration, well above the IC_{50}. Subsequently, the mixture is heavily diluted, which results in inhibitor concentration well below the IC_{50}. This change spans the steep part of the concentration-response curve as depicted in Figure 4.29A. Therefore, the enzyme inhibition dramatically decreases for a reversible inhibitor upon dilution, whereas an irreversible inhibitor is less affected. In the conducted experiment, **C5** as well as **F1b**, representing the quinoline-based fragments, were incubated with the CVB3 3C^{pro} at 500 µM each, which results in 80 % 3C^{pro} inhibition. After incubation they were diluted 30-fold and the inhibition was determined (Figure 4.29B). **F1b** shows only 15 % inhibition after dilution, whereas **C5** inhibits the 3C^{pro} by 60 %. This result is in accordance with the biological data obtained for the respective fragment series. The recovery of the 3C^{pro} activity after dilution confirms the reversible nature of the covalent binding mode of **F1b** and the other tetrahydroquinoline-2,5-diones with a time-dependent activity. The irreversible 3C^{pro} inhibition by **C5** is reflected by the missing recovery of the 3C^{pro} after dilution. This is not only in accordance with the kinetic data obtained before but also confirms the proposed thioether as reaction product of **C5** with the 3C^{pro} due to the irreversible nature of the formed covalent bond.

4.5 De novo design

To enable rational optimization of the identified fragment hits a computational *de novo* design approach was pursued. Thereby, analogues of the hit fragments are virtually generated and evaluated for target binding based on the generated structural models of the CVB3 $3C^{pro}$. This allows to virtually sample thousands or millions of possible fragments analogues of which an enriched set of the most promising is selected for chemical synthesis and biological evaluation.

In the following section, the utilized *de novo* design application and the established workflows for the virtual generation of molecules are described. In addition to the optimization of fragment hits, the *de novo* design workflow was extended for the preparation of ligands for covalent docking and the generation of decoy molecules for the validation of the 3D pharmacophore model, which is described below.

4.5.1 Reactor application for chemical transformations

De novo design is referred to as virtual generation of chemical structures from scratch [223]. This is beneficial as even the largest tangible collections of chemicals cover only a tiny fraction of theoretically existing drug-like molecules [224]. Therefore, *de novo* design can be used to cover greater portions of the chemical space filled with novel chemotypes. However, *de novo* design approaches typically face two major challenges: (i) Synthetic feasibility of the generated molecules and (ii) complexity of the chemical space. The synthesis of a *de novo* designed compound is essential to enter the test tube and thus make use of it. Secondly, the unknown biological activity of the virtual compounds hardly justifies extensive synthesis efforts. Despite the grown computational power, an exhaustive sampling of possible molecules through *de novo* design is still not feasible so that the represented chemical space is limited.

To address the above stated issues of *de novo* design, our pursued approach ensures synthetic feasibility by assembling molecules from tangible building blocks according to synthesis-tailored chemical transformation rules. Furthermore, the primary objective was the design of molecules based on biologically active fragments, which limits the sampled chemical space to the sweet spot around the respective hit fragment chemotype. The biological relevance of the fragment hit-based virtual compounds make their synthesis highly attractive.

A variety of *de novo* design tools have been developed of which some are free

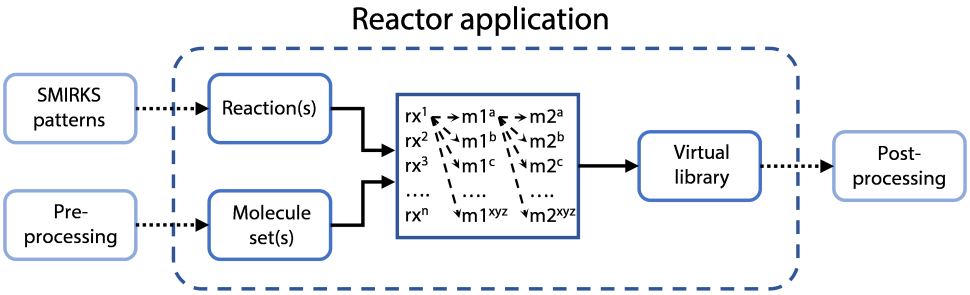

Figure 4.30: Schematic depiction of the function of the Reactor application.

of charge. However, we decided to build up our own application, which enables full control and ensures applicability for the intended purpose. Based on the open JChem API from ChemAxon (Budapest, Hungary) a reaction tool was built by Dr. Susanne Dupré as a command line-based java application. The tool reads chemical transformation rules encoded as SMIRKS patterns, applies these on the input molecules and generates a virtual library of the transformed entities. Depending on the SMIRKS pattern, either single molecule transformations or reactions between molecules are performed. For the latter, two or more input molecule sets are required as the reaction tool performs the transformations combinatorial (Figure 4.30). Each input molecule and chemical transformation is characterized by an unique identifier, which is used to tag the generated product molecules. This provides the necessary information to track every *de novo* generated compound back to the structures it originates from and the chemical transformation applied in the making. This is essential for the transition from the virtual to the chemical synthesis.

The *de novo* design of virtual compounds through the reaction tool was utilized for multiple purposes in the thesis. It was implemented within a specific workflow comprising the preparation of the input molecules sets, the SMIRKS pattern of the utilized chemical transformations and the processing of the virtual molecules.

4.5.2 Chemical reactions as SMIRKS patterns

The SMIRKS notation (see subsection 3.1.5) was used to encode the chemical transformations applied during the *de novo* design [196]. SMIRKS are a straightforward way to describe chemical transformations in terms of (i) generation of the patterns and (ii) compatibility with a wide range of computational tools. Despite the public availability of patterns describing chemical synthesis reactions [225], we decided to generate our own patterns due to better control of the results. SMIRKS

[H][NX3;!$([NH0]);!$(N~[#7,#8,#15,#16]);!$(NC[#7,#8,#15,#16]);!$(NC=[#7,#8,#16]);!$(NC=CC(=O));!$(N#*);!$(N*
#*);!$(N=*):1].[#6:3][CH1;!$(C-#7,#8,#16]); $(CC(=[O,C,N,S]));!$(C*[Cl,Br,I]);!$(CC(O)O):2](=O)>>[C:2]([#6:3])[N:1]

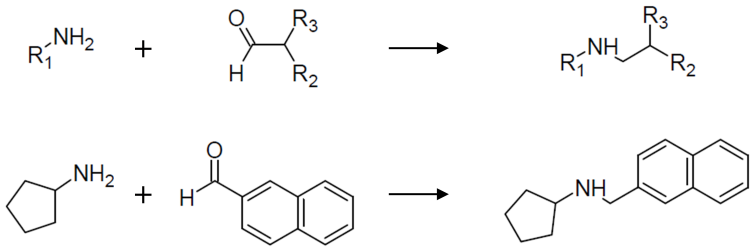

Figure 4.31: Chemical reaction encoding through SMIRKS patterns. SMIRKS pattern of a chemical reaction (top), Markush representation (middle) and an example for the application (bottom).

patterns represent a heuristic way to describe chemical reactions by defining substructures in a Markush representation, which enables the description of a reaction in a generalized pattern (Figure 4.31). The major challenge of SMIRKS patterns to mimic chemical reactions is the proper description of the substructure environment to ensure chemical reasonability and thus synthetic feasibility of the product molecules generated through the respective pattern. In a SMIRKS pattern only the educt(s) and the main product are considered. Leaving groups, conditions or intermediates can not be specified. Substructure definitions include the description of the respective chemical group like an amine, which can be defined as a nitrogen with at least one hydrogen atom, single-bonded to a carbon atom. However, other chemical groups comprising amine-like nitrogen atoms such as an amide have to be excluded. This is not only essential to prevent the generation of chemically meaningless products, but can also be used to control the reaction with respect to the reacting moieties. Thereby, the formation of certain functional groups in the products can be suppressed, which are either synthetically less accessible, undesirable in terms of drug-likeness or stability.

The selection of the chemical reactions for the pursued *de novo* design approach was essential as it determines the chemical nature of the designed compounds but also their synthetic accessibility. The major aspects, which determined the selection of the chemical reactions, resulted from (i) own considerations, (ii) *de novo* design approaches reported in literature and (iii) close communications with chemists from the collaborating group of Prof. Rademann. These included (i) relevance of the reaction for the synthesis of drug-like molecules, (ii) robustness, (iii) availability of the utilized building blocks and (iv) explicitly defined reaction product. Roughley and Jordan analyzed the reactions utilized in the chemical synthesis of compounds within drug design campaigns published by pharmaceutical

companies [226]. From their analysis, it becomes apparent that only a small number of reactions accounts for most of the applied transformations. These include (i) N/O-alkylation and -arylation, (ii) N-acylation, (iii) Pd-catalyzed C-C bond formation and (iv) deprotection. Eight chemical reactions of the first three chemical transformations types were selected (Figure 4.32) and encoded as SMIRKS patterns (see appendix A.4).

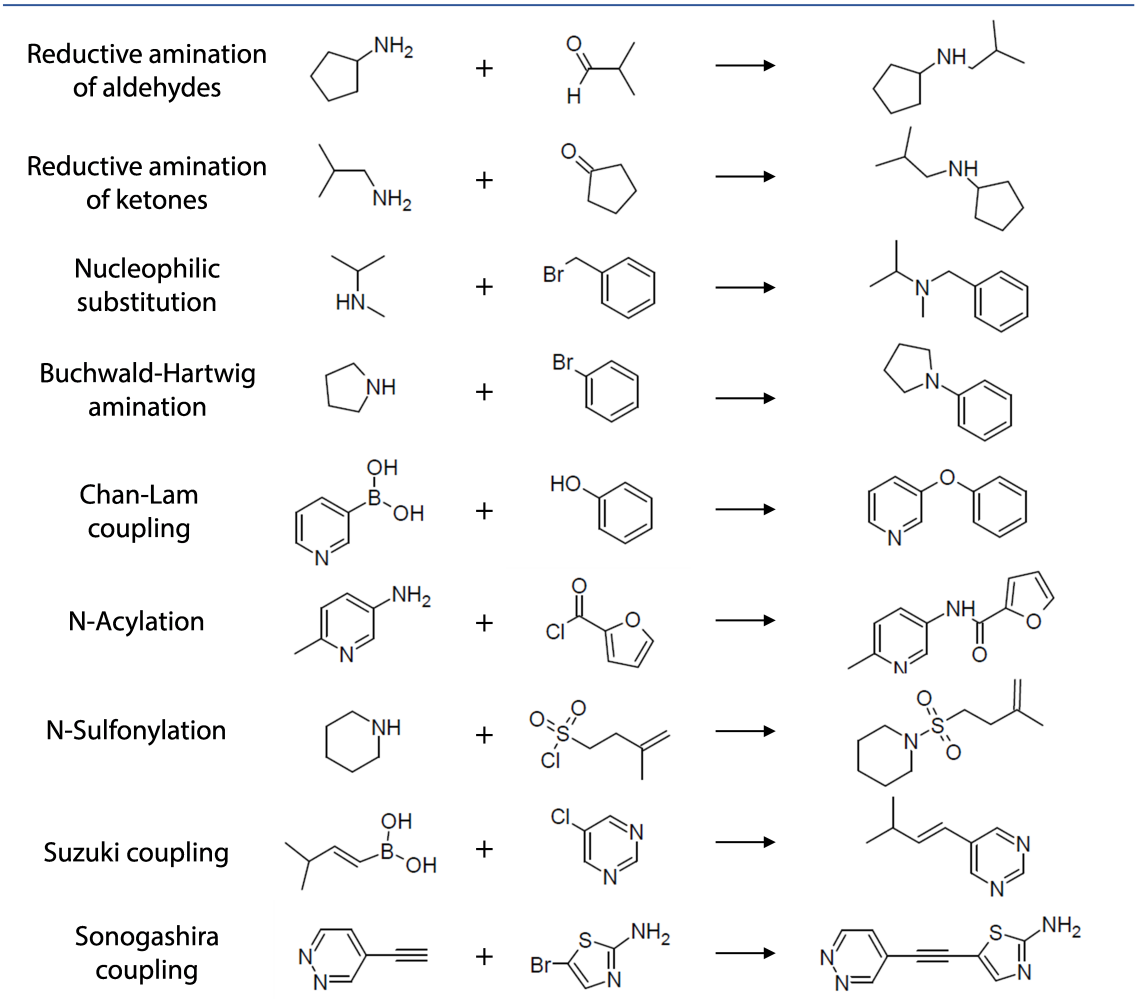

Figure 4.32: Chemical reactions utilized in the de novo design.

The selection was based on the most prominent members for each reaction type: (i) reductive amination and nucleophilic substitution with alkyl halides for N-alkylation,Buchwald-Hartwig and Chan-Lam coupling for heteroatom arylation, (ii) N-acylation and N-sulfonylation with acyl halides and sulfonyl chlorides, respectively, (iii) C-C bond formation via Suzuki and Sonogashira coupling. These reactions make up for 73 %, 81 % and 58 % of the applied transformations within their respective type of reaction. Their significance results from the robustness to

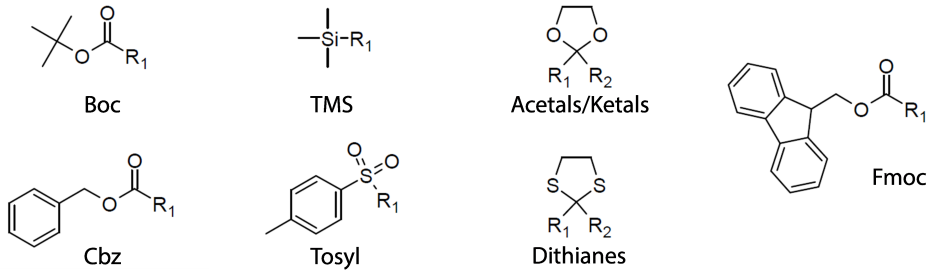

Figure 4.33: Protection groups removed during processing of de novo designed compounds.

be applicable for a wide range of educts and conditions but also from a general availability of the required building blocks. The frequency of deprotection reactions shows the practical relevance of protection/deprotection steps in chemical synthesis. Furthermore, commercial suppliers offer a variety of building blocks, which already incorporate protection groups. In our approach, deprotection is performed as part of the processing of the *de novo* designed molecules. Therefore, SMIRKS patterns for the removal of the most prominent protection groups depicted in Figure 4.33 were generated. The deprotection not only complies with the chemical practice, but also enables the reaction of previously protected functional groups in a subsequent virtual synthesis run. Furthermore, undesirable protection groups in terms of drug-likeness or stability are removed. Less often applied chemical transformations include heterocycle formations, reductions and functional group interconversion. The latter two are of little applicability for the desired purpose as we were aiming for the optimization of fragment hits via *de novo* design and these reaction would rather change the hit fragment structure than providing additional interactions or occupations of the protein active site and were thus omitted. Heterocycle formations are key in drug design as these provide the core scaffold of a compound. More than half of the reactions used in the approach of Hartenfeller *et al.* are for the synthesis of heterocycles [225]. This is reasonable for their intention to generate drug-like molecule from scratch. But in our case, the fragment hits already represent the scaffold so that heterocycle forming reactions were not necessary and thus were not included. Prominent reactions such as aldol reactions or Friedel-Crafts acylation were also not considered as their products can not be explicitly defined.

4.5.3 Building block collection

The second major aspect of the *de novo* design approach was the use of tangible chemical building blocks, which make the virtually generated molecules not only theoretically but also practically feasible. As the building blocks are provided by local chemical vendors, they can be obtained within days, which enables a quick transition of the virtual synthesis to the bench and subsequently to the biological evaluation.

Sigma Aldrich was chosen as primary source for building blocks (BB) due to (i) the large collection of regents, (ii) general availability and (iii) fast delivery of the majority of reagents. The chemical structure of the building blocks was directly obtained from the Sigma Aldrich web page and assembled in a large collection. However, before it could be used as input for the reactor application it had to be thoroughly filtered to ensure, on the one hand, the chemical reasonability of the performed virtual transformations and on the other hand, drug-likeness of the generated products. Therefore, a sequence of physicochemical property and substructure-based filters were applied (Figure 4.34). The physicochemical filters

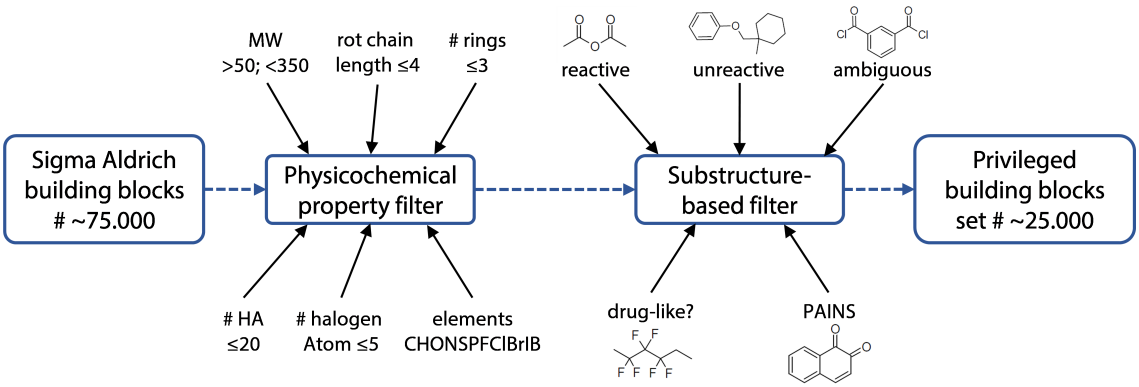

Figure 4.34: Filter applied during the preparation of the building blocks.

removed BB, (i) which are too big in term of molecular weight or heavy atom count, (ii) too flexible, (iii) contain too many halogen atoms, (iv) too many or too large ring systems and (v) non drug-like elements. The substructure-based filtering was applied to remove BB due to (i) reactive moieties, which are not part of the selected reactions, (ii) unreactive, non drug-like moieties, (iii) known assay interfering substructures, (iv) ambiguous BB due to multiple instances of the same moieties used in the reactions, which would not lead to an explicit product and (v) BB, which lack a chemical group taking part in at least one of the selected reactions. The applied filtering reduced the initial collection from 75.000 to 25.000 BB.

Additional preparation steps performed included (i) the standardization of the representation of functional groups, (ii) addition of explicit hydrogens to ensure correct treatment of the stereochemistry and (iii) deletion of unnecessary information associated with the BB to reduce the required size of the collection.

The assembled BB library represents a ready-to-use collection of tangible reagents that are suitable for the *de novo* design of drug-like molecules either by using the entire collection or specifically generated subsets.

4.5.4 De novo design workflow

The reactor application, the reaction patterns and the assembled set of BB provide the foundation for the pursued *de novo* design approach. The primary purpose was the optimization of the fragment hits. Therefore, it has been implemented in a convenient workflow, which takes care of the required processing of the *de novo* generated compounds in order to use them in a subsequent virtual screening or as input for additional *de novo* design cycles (Figure 4.35). The BB from our collection, which were split into two sets according to the specific requirements for fragment optimization as will be described in the next section, serve as input. The molecules in the two sets are combined according to the SMIRKS reaction patterns, which yields a virtual library of *de novo* designed compounds. At first, all duplicates are eliminated. Next, all molecules, which contain an untransformed reactive moiety are separated. The clean molecules are deprotected according to the rules described above. From the obtained deprotected molecules, those that contain a functional group, which can react in one of the utilized reactions, are combined with the prior separated reactive molecules. Together, these molecules can be used as potential input for an additional *de novo* design cycle. All deprotected, clean molecules are filtered based on their physicochemical properties to remove once more all non drug-like compounds. For the remaining molecules ionization states at physiological pH are assigned and 3D conformers are generated.

4.5.5 Covalent product formation workflow

The reactor application was additionally applied for the preparation of ligands for covalent docking. The utilized docking tool GOLD requires the ligands to be in the covalent bound state of the covalent enzyme inhibitor complex. Therefore, SMIRKS patterns, which mimic the reaction of the most common warheads with

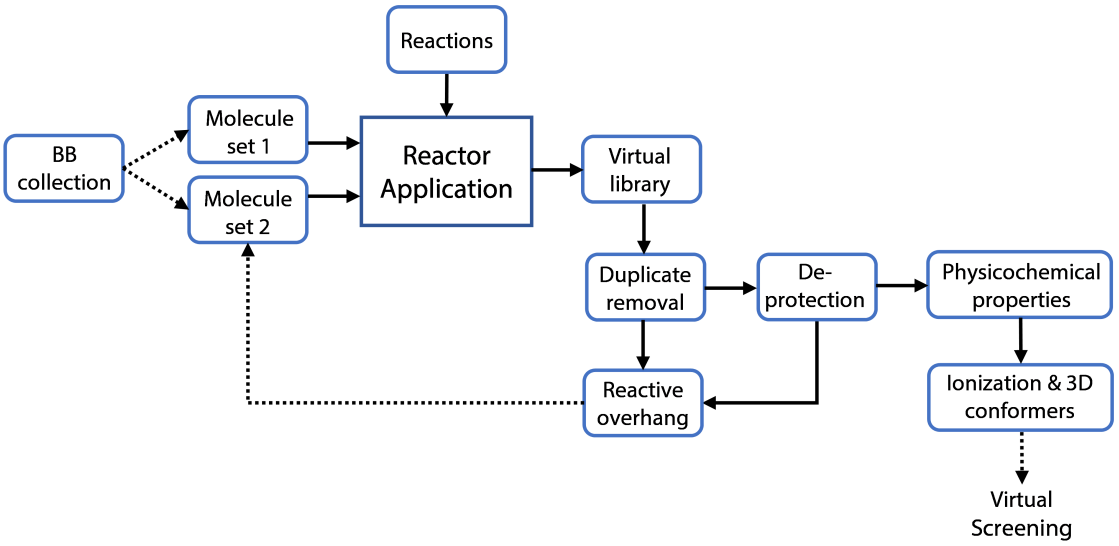

Figure 4.35: Workflow utilized for the de novo design of fragment hit analogues.

a cysteine nucleophile, were generated (Figure 4.36) (see appendix A.4). For some warheads, the covalent bond formation involves the generation of one or more stereocenters. For these cases, pattern for all possible stereoisomers were generated.

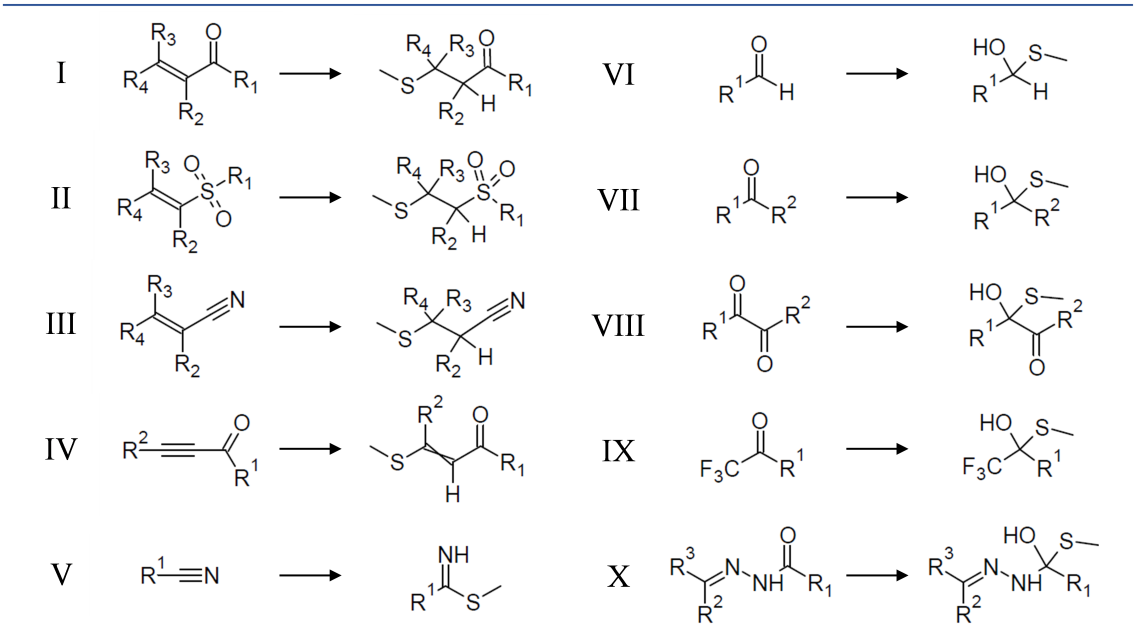

Figure 4.36: Warhead cysteine reactions encoded as SMIRKS patterns for automatic covalent product generation for covalent docking. I α,β-unsaturated carbonyl compounds, II vinyl sulfones, III vinyl nitriles, IV propargyl carbonyl compounds, V nitriles, VI aldehydes, VII ketones, VIII dicarbonyl compounds, IX trifluoromethyl ketones, X semicarbazones.

The use of these reaction patterns in combination with the reactor application

allows the preparation of large molecule collections for covalent docking circumventing the tedious manual preparation of each individual molecule. This enables to use covalent docking as convenient virtual screening method. The workflow for the preparation of ligands for covalent docking consist of the following steps, depicted in Figure 4.37: (i) Selection of the ligands for covalent docking, (ii) generation of the covalent bound state through the reactor application and the appropriate SMIRKS pattern, (iii) generation of the 3D conformation for the reaction application output, which is in smiles format.

Several docking programs incorporate the warhead transformation within their covalent docking procedure. However, this is often limited to certain warheads and also error prone due to simplified transformation patterns [227, 228]. As already mentioned for the patterns mimicking chemical synthesis reactions, the proper description of the environment of the warhead is key to ensure chemical reasonability. Therefore, we utilized once more our own patterns, which, in addition, allowed to better reflect biological data like for the thiophenolmethyl ketone fragment **C5** as described in the previous section or to include challenging warheads such as epoxides, which has been done for the covalent docking of the epoxide-based $3C^{pro}$ inhibitors synthesized by Daniel Becker (see section 4.2).

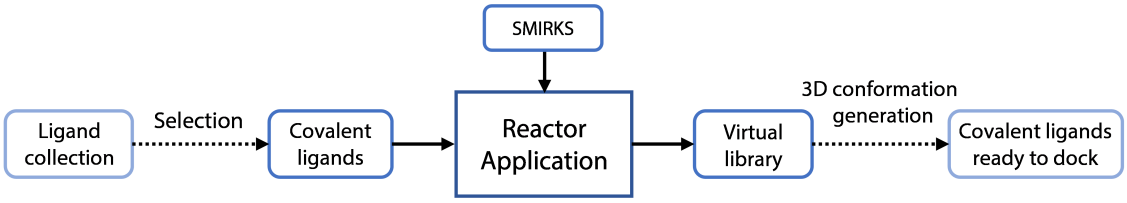

Figure 4.37: Workflow established and applied for the preparation of ligands for covalent docking in GOLD.

4.5.6 Decoy generation workflow

For the validation of the pharmacophore model for the CVB3 $3C^{pro}$ described in section 4.1, decoys were generated using the above described *de novo* design workflow (Figure 4.38). The primary objective was the generation of appropriate decoys due to the incorporation of warhead moieties, which lack molecules from typical sources of putative decoys [170]. This was achieved by preselecting BB as input, which comprise a covalent binding substructure. Through substructure-based filtering, a focused set of BB was assembled and reacted with a diverse subset of the entire BB collection with the reactor application according to the SMIRKS reaction patterns. The obtained virtual molecule library was processed by using

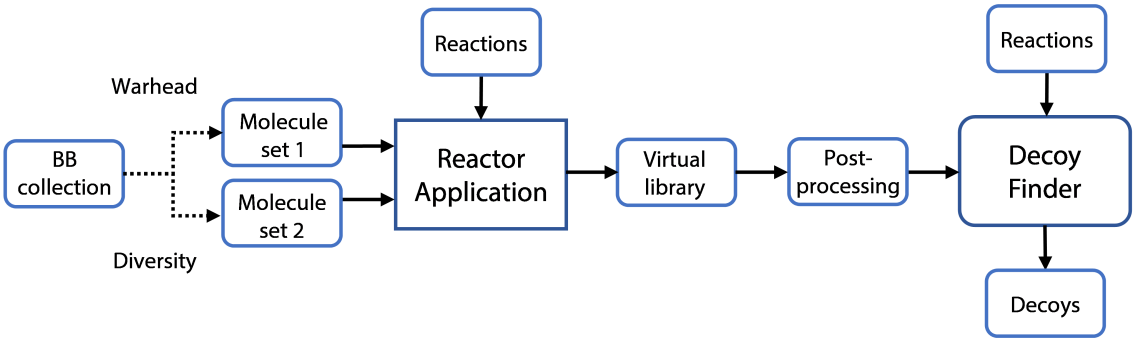

Figure 4.38: Workflow established for the generation and the selection of decoys utilized in the validation of the 3C^{pro} 3D pharmacophore model.

the same *de novo* design workflow described above with a slight modification to ensure the incorporation of a warhead moiety. Product molecules, in which the warhead was transformed during their generation, were discarded. The prepared collection containing around 5 million virtual compounds served as source for the selection of appropriate decoys.

This was done through the in-house KNIME decoy finder workflow developed by Dr. Susanne Durpré. Hereby, molecules from a large collection, such as the one generated *de novo*, are compared to the biological active compounds used in the model validation. For each of the molecules in the active set up to 50 similar molecules are selected, which serve as decoys. The similarity is defined by physicochemical properties, which are (i) molecular weight, (ii) calculated partition coefficient clog P, (iii) number of rotatable bonds and (iv) number of hydrogen bond donors and acceptors, respectively. Furthermore, the Tanimoto coefficient for the fingerprint similarity is utilized as selection criteria [229]. From the virtual compound library 1771 molecules were selected as useful decoys based on their similarity to the active set.

4.6 Fragment hit optimization

The scaffold hopping led to the identification of additional fragments hits for the CVB3 3C^{pro}, which serve as starting point for optimization. Therefore, the established *de novo* design workflow was utilized to generate virtual analogues. The most promising ones were selected through structure-based virtual screening for subsequent chemical synthesis and biological evaluation.

4.6.1 Fragment growing strategy

Scaffold hopping-based exploration of the initial hit fragment **F1** led to the identification of additional 3C^{pro} inhibiting fragments, which could be used as starting points for subsequent optimization (Figure 4.39). The fragment **F1b** represents the tetrahydroquinoline-2,5-dione scaffold, which is similar to **F1** but chemically stable. The additional occupation of the S1' pocket improved the biological activity in a ligand efficient manner, which is in the high two-digit micromolar range with a reversible covalent mode of action. In contrast, **C5**, which is active in a similar range, forms an irreversible covalent bond with the catalytic cysteine of the 3C^{pro} through its thiophenolmethyl ketone moiety. The binding mode elucidation through protein mass spectrometry not only revealed an interesting mode of action but also provided further evidence for the proposed binding conformation, which is key for structure-based optimization in the absence of crystallographic data. Therefore, **C5** was chosen as starting point for chemical optimization.

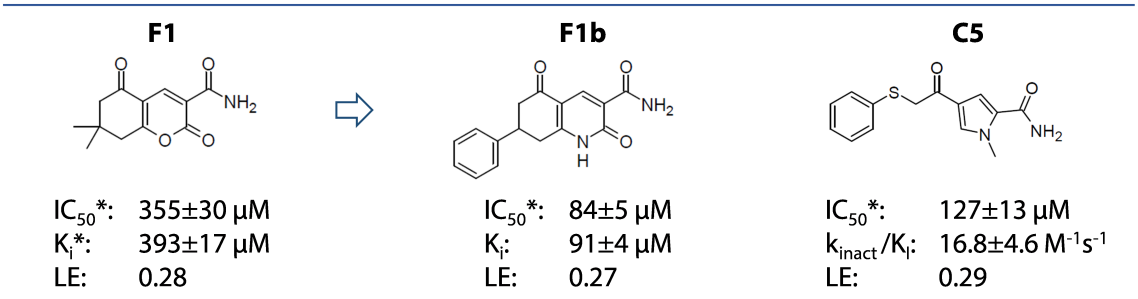

Figure 4.39: Fragment hits identified as 3C^{pro} inhibitors with activity values.

The most promising fragment optimization strategy was fragment growing as all identified fragment hits putatively bind to the same part of the 3C^{pro} active site, the S1 to S1' pocket area. At first, growing vectors for **C5** were defined based on the proposed binding conformation. The pyrrole nitrogen represents a potential synthetic handle for optimization as its methyl substituent does not contribute to

$3C^{pro}$ binding and thus could be exchanged for other chemical groups. Through the incorporation of additional moieties at this site, **C5** could be grown according to the prior defined vectors toward the adjacent pockets S2 and S4 (Figure 4.40). The *de novo* design workflow, described in the previous section, was utilized to generate putative analogues of **C5** that should be able to occupy these pockets. Thereby, the fragment growing space of **C5** could be effectively sampled through the virtual generation of all synthetically feasible analogues of **C5** with respect to the defined growing vectors. To enable a virtual screening-based enrichment of

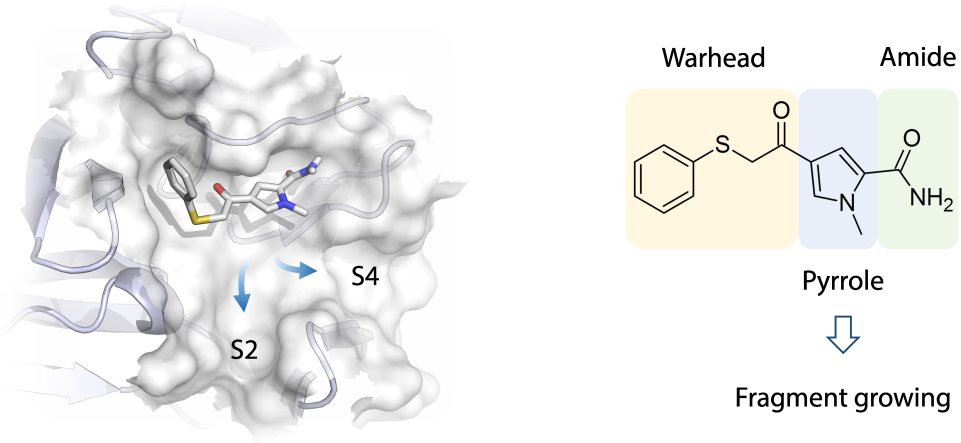

Figure 4.40: Fragment growing vector and design strategy for C5 optimization.

promising analogues able to occupy the adjacent pockets S2 and S4, respectively, the structure-based 3D pharmacophore model was extended as it only comprises the interaction essential for binding to the S1 and S2 pocket. The characteristics of the active site regions, which bind the P2-P4 substrate residues have been described in detail in section 4.1. The S2 pocket is unrestrictive, whereas lipophilic moieties are favorably bound due to lipophilic contacts with the neighboring side chain of L127. The S4 pocket is very small and thus restrictive for alanine and its methyl side chain. The key interactions of the P3,4 substrate residues are a β-sheet like hydrogen bonding pattern with the backbone atoms of G164 and G128 (Figure 4.41A). This pattern is maintained in the peptidic and peptidomimetic inhibitors developed for the $3C^{pro}$(Figure 4.41B). Therefore, the peptidic ligands in the $3C^{pro}$ crystal structures utilized for initial structure-based 3D pharmacophore model generation, described in section 4.1, form this interaction pattern as well. However, the respective features were removed from the pharmacophore model due to the focus on the non-peptidic inhibitors, which do not bind to this part of the $3C^{pro}$ active site. Therefore, a second 3D pharmacophore model was generated, which comprises the β-sheet like hydrogen bonding pattern of the P3,4

binding area. For model validation 52 highly potent ($k_{inact}/K_I > 10^5\ M^{-1}s^{-1}$) peptidic inhibitors of the HRV $3C^{pro}$ reported by Dragovich *et al.* were used [81–86, 89]. However, the peptidc nature of these inhibitors with a median of 21 rotatable bonds provides a significant challenge for the generation of multiconformer libraries used for 3D pharmacophore-based screening. Even the generation of 500 conformations per ligand did not provide any match to the model. Therefore, protein-ligand docking was utilized to generate structurally relevant ligand conformations. Docking poses were obtained for 22 of the ligands, which matched the interactions represented by the 3D pharmacophore model (Figure 4.41C). These poses are highly similar to the binding conformations of the ligands verified by X-ray crystallography, which shows the ability of the 3D pharmacophore model to identify biologically relevant binding conformations.

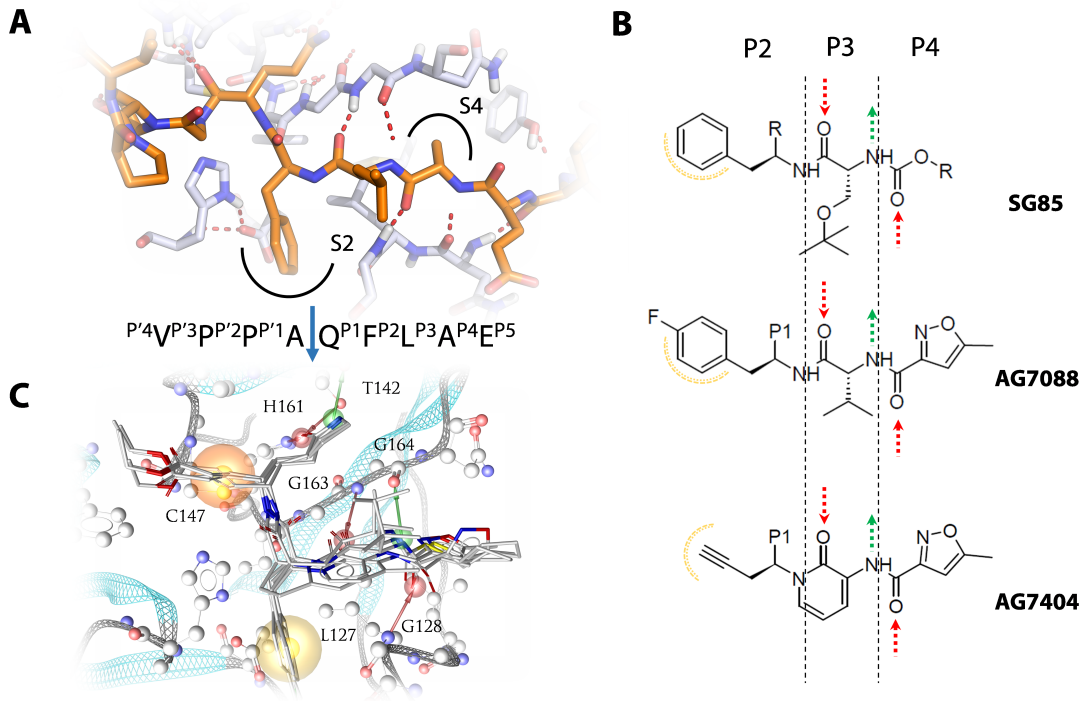

Figure 4.41: CVB3 $3C^{pro}$ S2,S4 pocket characteristics. (A) P2-4 of a representative peptide binding to the $3C^{pro}$ active site. Hydrogen bonds are depicted as red dashed lines. (B) P2-4 mimicking substructures of representative inhibitors of the $3C^{pro}$ reported in literature. Lipophilic contacts in the S2 pocket are indicated by dashed yellow lines. Hydrogen bonds of the β-sheet like interaction pattern are depicted as green and red arrows for hydrogen bond donor or acceptor functionality, respectively. (C) Depiction of the extended 3D pharmacophore model in the active site of the CVB3 $3C^{pro}$ with representative docking poses of matched peptidic inhibitors. Orange sphere indicates covalent bond, yellow spheres lipophilic contacts, red arrows hydrogen bond acceptors and green arrows hydrogen bond donors, respectively.

4.6.2 De novo design & virtual screening

For the virtual growing of **C5** a suitable core structure had to be selected, which is applicable for the *de novo* design workflow and enables the subsequent chemical synthesis of the virtually generated compounds. Therefore, the prepared BB library was searched leading to several building blocks of which the 4-acetyl-1H-pyrrole-2-carboxylic acid was the most convenient one since the amide as well as the α-thiophenol ketone moiety could be introduced in a few reaction steps (Figure 4.42). This was done manually for the virtual building block to enable subsequent virtual screening for the generated fragment growing analogues. This prepared **C5** core structure served as input for the *de novo* design-based fragment growing workflow described in the previous section, in which it was combined with the BB of the prepared library according to the generated set of chemical transformations. But only those reactions were used, in which neither the amide nor the ketone moiety were transformed, as these had to be maintained to ensure biological activity of the generated product molecules. The generated virtual library comprised ∼2000 analogues of **C5**, which exclusively resulted from N-alkylation of the pyrrole (Figure 4.42). These were subsequently screened by using the extended 3D pharmacophore model. However, no match was retrieved by using the whole model. Therefore, either the features for the interactions in the S2 pocket or in the P3,4 binding area were omitted, so that subsets of ligands specific for either part of the active site were obtained. 118 ligands matched the 3D

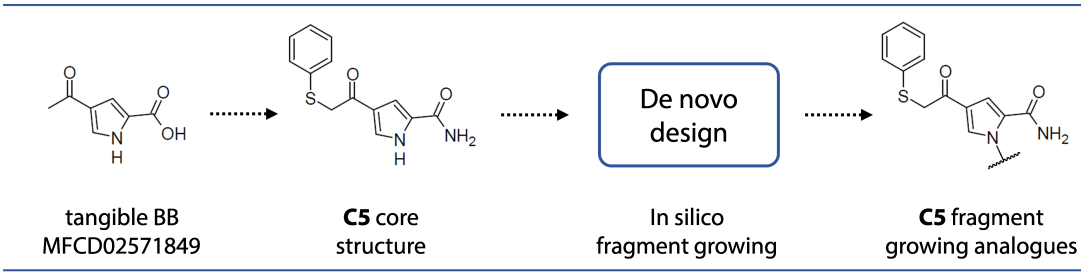

Figure 4.42: De novo design-based fragment growing of C5. From left to right: Chemical structure of putative synthetic precursor, input structure for *de novo* design workflow, generalized structure of **C5** fragment growing analogues with variable pyrrole substituents.

pharmacophore model comprising the lipophilic feature in the S2 pocket. Subsequently, each ligand was covalently and non-covalently docked into the $3C^{pro}$ active site. As the proposed binding mode of **C5** involves the formation of two covalent complexes, the thiohemiketal and the thioether, the ligands were docked in both states. The obtained poses were evaluated through the 3D pharmacophore

model to identify ligands able to form the crucial interactions in the S1 as well as the S2 pocket in all three states (Figure 4.43A). This was the case for 38 **C5** analogues. Further reduction resulted from visual inspection of the binding conformations as well as consideration of availability and price of the respective building blocks. The final selection comprised 11 promising **C5** analogues, which are able to bind to the $3C^{pro}$ in accordance to **C5**. In addition, they could occupy the S2 pocket through their lipophilic pyrrole substituents, which range from flexible phenyl alkyl ethers to the rigid methyl-biphenyl moiety (Figure 4.43B). The variability in the utilized building blocks results from the exposed and unrestrictive nature of the S2 pocket.

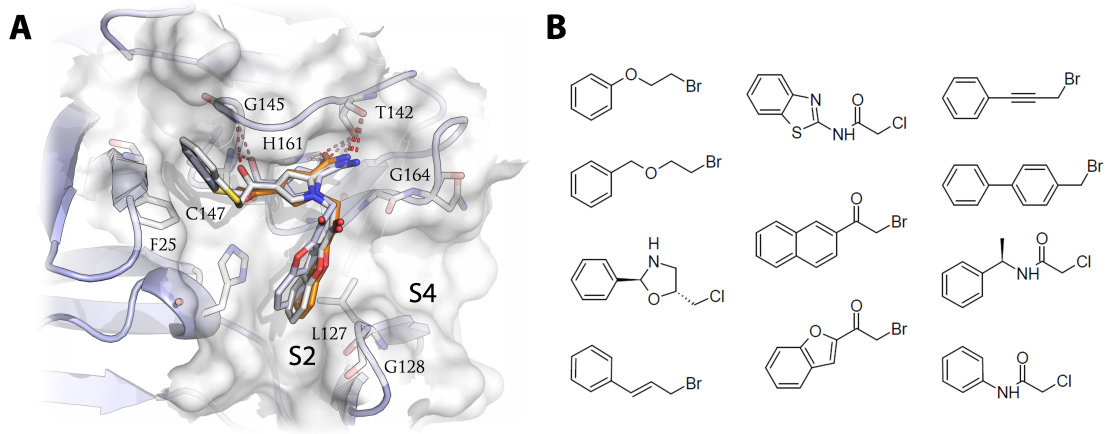

Figure 4.43: Fragment growing toward $3C^{pro}$ S2 pocket. (A) Docking poses of a representative fragment growing analogue of **C5** occupying the S2 pocket of the $3C^{pro}$. All three states are depicted: initial complex (white stick coloring), covalent thiohemiketal intermediate (light blue stick coloring) and final thioether (orange coloring). Hydrogen bonds are indicated by red dashed lines. (B) Chemical structures of the BB utilized in the *de novo* design of the **C5** fragment growing analogues in the S2 subset.

The growing toward the P3,4 area was more challenging due to the specific interaction pattern required for binding to this area. The 3D-pharmacophore-based virtual screening of the initial set of *de novo* designed ligands did not yield any promising structures. Therefore, building blocks from the chemical vendor ENAMINE were additionally included. The ENAMINE collection of BB exceeds the one of Sigma Aldrich and thus could contain BB, which comprise the necessary chemical functionalities. The preparation of the ENAMINE BB was done in accordance to the Sigma Aldrich collection. The final collection contained over 50,000 structures, which were not directly reacted with the **C5** core through the *de novo* design workflow but screened with the 3D pharmacophore model to assess the ability of the BB to bind the P3,4 area. Thereby, those BB were selected, which

are able to form the β-sheet like hydrogen bonding pattern. These were combined with the **C5** core through the *de novo* design tool, which resulted in 2442 fragment growing analogues. The 3D pharmacophore-based screening and subsequent docking yielded 12 ligands able to form at least two of the three hydrogen bonds in the P3,4 area (Figure 4.44A). Visual inspection and availability check of the BB reduced the number by 50 %. This led to a final selection of six fragment analogues of **C5**, which are able to occupy the P3,4 binding area. The pyrrole substituents of the virtual **C5** analogues consist in general of a flexible linker, which comprises the hydrogen bond forming moiety and connects the pyrrole with a terminal lipophilic moiety. These can form hydrophobic contacts in the S4 pocket. Ortho-dimethyl substituted benzene moieties appeared to be especially suitable as the methyl groups induce an out of plane conformation of the planar aromatic system allowing for occupation of the S4 pocket similar to the P4 alanine side chain (Figure 4.44A). The moieties, which form the β-sheet like hydrogen bonding pattern include (i) amides, (ii) imides and (iii) acylureas (Figure 4.44B).

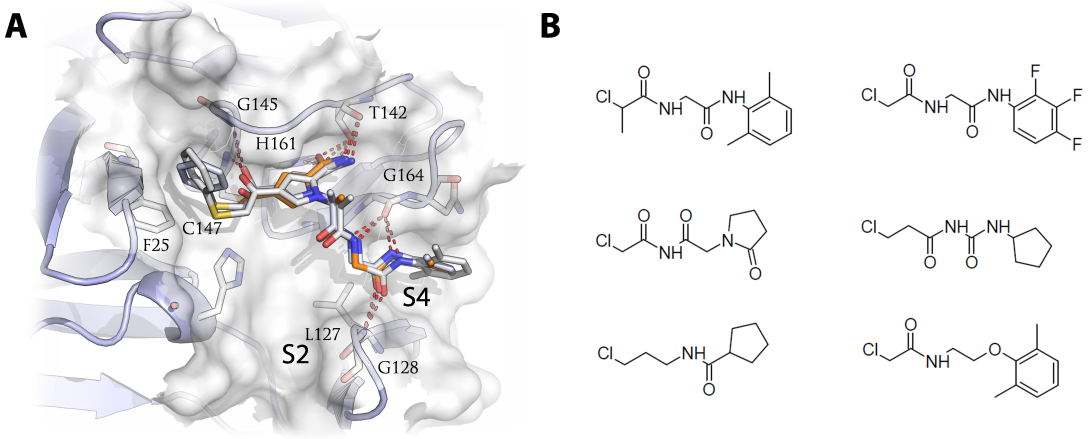

Figure 4.44: Fragment growing toward $3C^{pro}$ S4 pocket. (A) Docking poses of a representative fragment growing analogue of **C5** occupying the P3,4 binding area of the $3C^{pro}$ active site. All three states are depicted: initial complex (white stick coloring), covalent thiohemiketal intermediate (light blue stick coloring) and final thioether (orange coloring). Hydrogen bonds are indicated by red dashed lines. (B) Chemical structures of the BB utilized in the *de novo* design of the **C5** fragment growing analogues in the P3,4 subset.

The virtual growing of fragment **C5** enabled the identification of analogues, which can occupy additional pockets in the $3C^{pro}$ active site, while maintaining the initial binding mode of **C5**. These served as guidance for the subsequent chemical synthesis, which was facilitated by the utilized *de novo* design workflow as each analogues of **C5** has been generated through feasible chemical reaction schemes and from tangible chemical reagents.

4.6.3 Fragment growing: Synthesis & biological evaluation

The synthesis of the virtually generated **C5** analogues was realized in collaboration with the group of Prof. Rademann. The synthesis was conducted by Amira Atef Suliman within her master thesis entitled "Establishing a synthesis route for the growing of fragment-based viral C3 protease inhibitors". The synthesis of the S2 addressing analogues was pursued at first as the utilized building blocks are more common reagents in comparison to those used for the S4 occupying ligands, which makes them not only cheaper but also more convenient to work with. In addition, the *de novo* design suggested a nucleophilic substitution for the alkylation of the pyrrole, which is a bad nucleophile in general as the nitrogen lone pair takes part in the aromatic π-system. Therefore, the more reactive alkyl bromides, which can only be found in the selection for the S2 pocket, were preferred in the creation of a synthesis route (Figure 4.45).

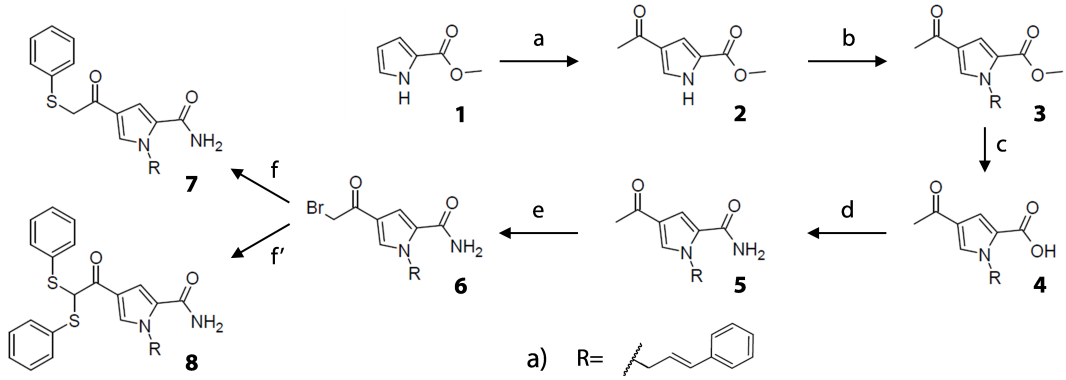

Figure 4.45: Scheme for the synthesis of the C5 analogues. Reagents and conditions: (a) $AlCl_3$, CH_3COCl, DCE, rt; (b) K_2CO_3, alkylhalide, DMF; (c) KOH, MeOH, THF; (d) $(COCl)_2$, NH_3, THF; (e) PTTB, THF; (f) K_2CO_3, DMF, C_6H_5SH.

The envisioned synthesis starts from the methyl 2-pyrrole-carboxylate **1**, from which the analogues methyl ester **2** of the initially intended precursor (4-acetyl-pyrrole-2-carboxylic acid) could be obtained through Friedel-Crafts acylation with acetyl chloride in dichlorethane (DCE) and catalytic amounts of aluminum chloride. The ester building block was not only cheaper but also putatively more inert to side reactions of the carboxylic acid in the subsequent alkylation. This was done through deprotonation of the pyrrole nitrogen with potassium carbonate in N,N-dimethylformamide (DMF) and reaction with the respective alkyl halide to give the alkylated pyrrole **3**. Next, the ester was converted into the amide. Attempts for direct conversion through aminolysis with liquid ammonia did not

work, so that the ester had to be hydrolyzed under basic conditions. The resulting acid **4** was activated with oxalyl chloride and reacted with ammonia in tetrahydrofurane (THF) to give the amide **5**. The next step was the formation of the α-thiophenolmethyl ketone moiety. Therefore, the ketone **5** was brominated in α-position with phenyltrimethylammonium tribromide (PTTB) in THF to give the bromomethyl ketone **6**, which was further converted with thiophenol under basic conditions in DMF to the thiophenolmethyl ketone **7**.

As alkyl halide BB the 2-bromoacetyl-naphthalene was used at first but the bromination turned out to be challenging, probably due to the presence of a second ketone moiety. Therefore, the 1-bromo-3-phenyl-propene was used instead, for which the desired thiophenolmethyl ketone **7a** could be obtained. In addition, the double substituted dithiophenolmethyl ketone **8a** was also obtained through the use of an excess of thiophenol. This unintentionally synthesized analogue provided the opportunity to assess the specificity of the thiophenolmethyl ketone warhead, which was extended by including the synthesis intermediates **5a** and **6a** in the analysis as they only differ to **7a** in the comprised warheads, which are a ketone and a bromomethyl ketone, respectively. These four prepared analogues of **C5** were biologically evaluated in the enzyme kinetic assay. This revealed inhibitory activity for all four synthesized compounds (Table 4.6).

Table 4.6: Binding characteristics and inhibition values for synthesized C5 analogues. Kinetics denotes the reversibility of the covalent interaction with the CVB3 $3C^{pro}$.

ID	IC_{50}* [µM]	K_i [µM]	k_{inact}/K_I [$M^{-1}s^{-1}$]	kinetics
C5	127 ± 13	-	16.8 ± 5	irreversible
5a	360 ± 27	197 ± 8	-	reversible
6a	0.4 ± 0.05	-	$1,707 \pm 33$	irreversible
7a	80 ± 10	-	54.4 ± 4	irreversible
8a	302 ± 60	272 ± 11	-	reversible

The ketone **5a** shows activity in the low three-digit micromolar range with a time-dependent mode of action (Figure 4.46A,B,D). The plot of k_{obs} as a function of the **5a** concentration shows a reversible mode of action with a K_i of 197 ± 8 µM (Figure 4.46E). This is in accordance with the expected mechanism of the ketone warhead. The activity of **5a** is lower than that of the fragment hit **C5**. However, the other pyrrole-based fragments, which do not comprise the thiophenolmethyl ketone warhead, did not show activity so that the incorporation of the phenylpropenyl moiety provides an apparent boost in activity. The disubstituted

derivative **8a** shows also a reversible kinetics in a similar activity range as **5a** (Figure 4.46A,C,F). But due to the double substitution, **8a** is very lipophilic and thus poorly soluble, which hampered the biological evaluation. The two substituents, not only have a negative effect on the physicochemical properties but also on the interaction with the $3C^{pro}$ as it is only as active as the significantly smaller **5a**. The dithiophenolmethyl ketone **8a** is probably too bulky, which hampers the covalent interaction with the catalytic C147. The reversible covalent binding of **8a** can be seen in analogy to per-halomethyl ketones. These are reversible inhibitors in contrast to their irreversibly binding mono halogenated analogues [126].

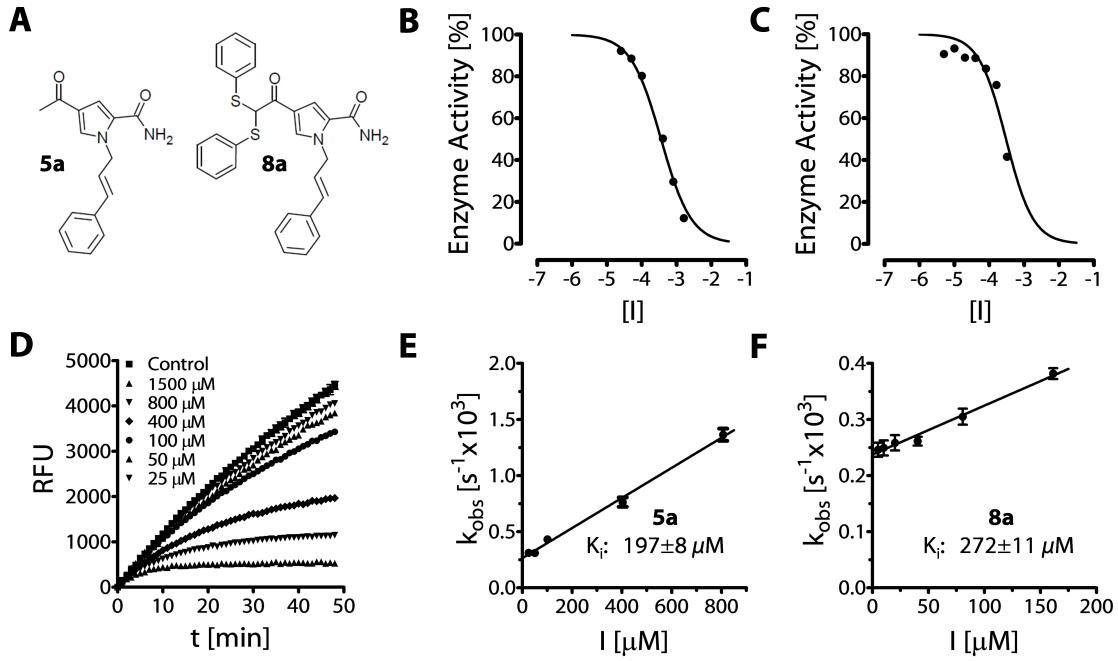

Figure 4.46: Biological characterization of 5a and 8a. (A) Chemical structures of fragment growing analogues with reversible binding kinetics. Dose-response curve for **5a** (B) and **8a** (C). (D) Enzyme reaction progress curves in the presence of different concentrations of **5a**. Plot of k_{obs} as a function of [I] for **5a** (E) and **8a** (F).

In contrast, **6a** and **7a** show irreversible binding kinetics (Figure 4.47A). The reactive synthesis intermediate **6a** is active in the three-digit nanomolar range (Figure 4.47B). The plot of k_{obs} as a function of the concentration of **6a** reveals a linear graph, which passes through the origin. The slope gives a k_{inact}/K_I value of $1,707 \pm 33\ M^{-1}s^{-1}$ (Figure 4.47E). The two-step mechanism of the α-halomethyl ketones is well documented in literature [129]. Therefore, one would expect a hyperbola for the k_{obs} vs [I] plot as described in subsection 1.2.4. But due to the property of the bromine to be a very good leaving groups, the reaction kinetics are primarily dominated by the rate constant for the elimination of the bromine (k_3),

which makes it linear. The desired fragment growing analogue **7a** shows the same binding characteristic as **C5** with a clear time-dependent activity (Figure 4.47D) and an irreversible mode of action, which follows a two-step mechanism as indicated by the hyperbolic curvature of the plot of k_{obs} as a function of the concentration of **7a** (Figure 4.47F). The inhibitory activity is in the low two digit micromolar range with a k_{inact}/K_I value of $54.4 \pm 4\ M^{-1}s^{-1}$.

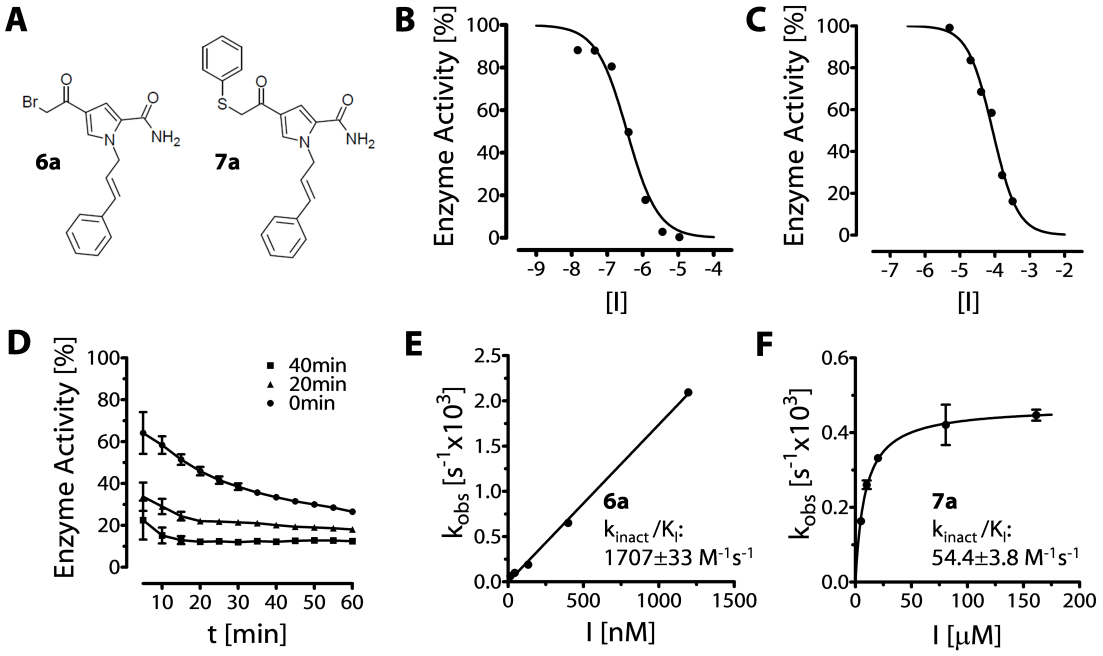

Figure 4.47: Biological characterization of 6a and 7a. (A) Chemical structures of fragment growing analogues with irreversible binding kinetics. Dose-response curve for **6a** (B) and **7a** (C). (D) Relative enzymatic activity of the $3C^{pro}$ in the progress of the enzymatic reaction after incubation with **7a** at 50 µM for 40 min (square), 20 min (triangle) or 0 min (circle). Plot of k_{obs} as a function of [I] for **6a** (E) and **7a** (F).

This biological evaluation shows a threefold increase in activity for **7a** in comparison to its point of origin, fragment **C5**. This activity boost is less than expected in comparison to the increase in activity determined for the simple ketone **5a**. The question at hand was, why does the addition of the pyrrole substituent for the occupation of the S2 pocket has not the same effect for **7a** as for **5a**? Therefore, the structural model was revisited with a special focus on the transition from the reversible thiohemiketal occurring in the reaction of both ligands to the irreversible thioether, only formed by **7a**. The apparent difference in both reaction products is the position at which the thiol moiety of the catalytic C147 is bound to the ligand and thus the distance to the amide moiety bound in the S1 pocket. This distance, initially identified as key descriptor for $3C^{pro}$ inhibition, increases by one single

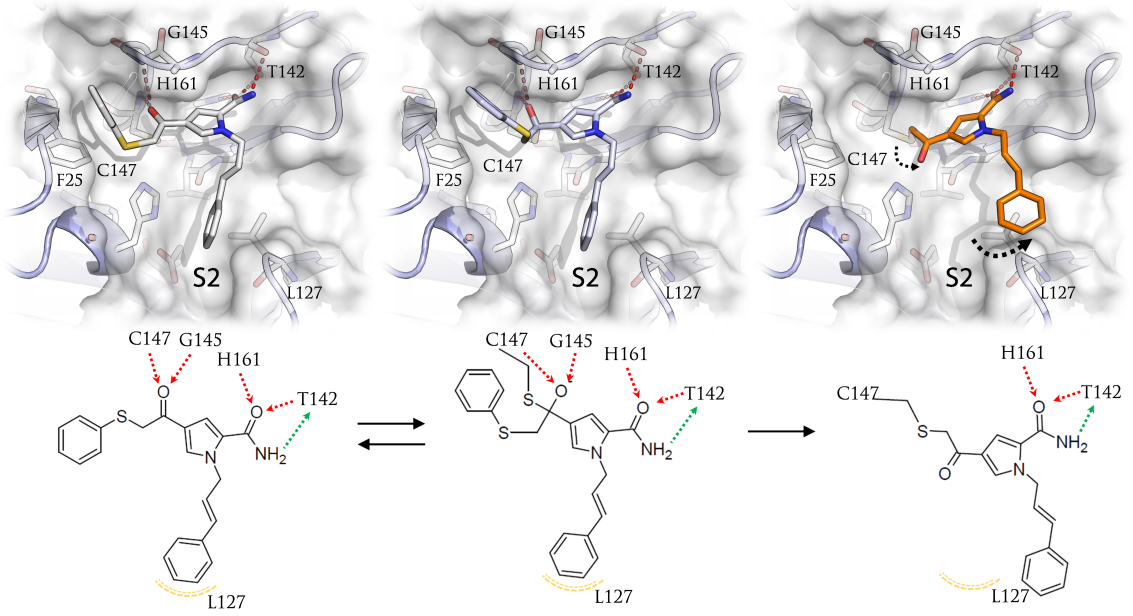

Figure 4.48: 3C^{pro} binding mode of 7a. 3D (top) and 2D (bottom) depiction of the three steps of the 3C^{pro} binding by **7a**. Dashed arrow indicates the repositioning of the phenylpropenyl substituent during transition from the thiohemiketal to the thioether. Hydrogen bonds are indicated as red dashed lines in the 3D depiction. Interactions in the 2D depiction are colored as follows: Yellow lines indicate lipophilic contacts, green and red arrows indicate hydrogen bond donor or acceptor functionalities.

bond during the transition from the thiohemiketal to the thioether. Thereby, the pyrrole system has to turn slightly in order to maintain the crucial hydrogen bonds in the S1 pocket while accommodating the ketone carbonyl moiety. This affects the pyrrole N-substituent, which is of little relevance for **C5** as its methyl substituent does not engage in interactions with the 3C^{pro}. However, for **7a** the slight movement of the pyrrole putatively disturbs the binding of the phenylpropenyl substituent in the S2 pocket (Figure 4.48). Therefore, the elimination of the thiophenol leaving group of **7a** is hindered, which reduces the benefit of the warhead for the 3C^{pro} inhibition in **7a** in comparison to **C5**.

In addition, we were able to evaluate the activity of the developed inhibitors against the EV D68 3C^{pro} and thus their potential as general 3C^{pro} inhibitors. The high structural similarities as well as docking analysis suggested that the binding mode of the inhibitors is conserved for EV D68 3C^{pro} (Figure 4.49A). This was confirmed by the obtained activity values listed in Table 4.7, which are similar to those obtained for the CVB3 3C^{pro}. The ketone **5a** is slightly less active with a K_i of 422 ± 8 µM and a reversible mode of action (Figure 4.49C). The fragment **C5** and the optimized analogue **7a** are slightly more active with k_{inact}/K_I values of 21.4 ± 8 $M^{-1}s^{-1}$ and 160 ± 26 $M^{-1}s^{-1}$ against EV D68 3C^{pro} (Figure 4.49C). In

Table 4.7: Binding characteristics and inhibition values for EV D68 3C^{pro}. Kinetics denotes the reversibility of the covalent interaction with the 3C^{pro}

ID	IC_{50}* [μM]	K_i [μM]	k_{inact}/K_I [$M^{-1}s^{-1}$]	kinetics
C5	116 ± 15	-	21.4 ± 8	irreversible
5a	376 ± 11	422 ± 8	-	reversible
6a	0.8 ± 0.02	-	$1,214 \pm 12$	irreversible
7a	25 ± 3	-	160 ± 26	irreversible

contrast, the rate of inactivation by the bromomethyl ketone **6a** is reduced by one-third to $1,214 \pm 12\ M^{-1}s^{-1}$ in comparison to the CVB3 3C^{pro} and it shows a steep curvature of the dose-response plot with a hill slope of 1.8 according to equation 1.16 (Figure 4.49B,C). This could result from unspecific enzyme inhibition or multiple **6a** binding. Since this was not observed for the CVB3 3C^{pro}, we surmised that it results from a specific characteristic of the EV D68 3C^{pro}. Structural analysis of the EV D68 3C^{pro} revealed that it comprises a second cysteine residue (C60) in addition to the catalytic C147, which is the only cysteine in the CVB3 3C^{pro}. Additional covalent binding of **6a** to the non-catalytic cysteine of the EV D68 3C^{pro} would explain the high steep dose-response curve. Protein MS of the irreversible 3C^{pro} inhibitors with the EV D68 3C^{pro} was used to verify this hypothesis and confirm the binding of the evaluated inhibitors (Figure 4.49D). For the apo EV D68 3C^{pro} a single peak of 21261.26 Da was found. After reaction with **6a**, two main peaks can be found with 21527.6 Da and 21793.67 Da, respectively. These correspond to mass shifts of 266 Da and 532 Da, respectively, which is twice as much as the former. The mass shift of 266 Da equals the mass of 6a except the bromine. Thus, the peaks reflect single and double binding of **6a** through unspecific reaction with either cysteine residue. In contrast, for **7a** only the adduct with 21527.05 was found, which corresponds to the addition of **7a** except the thiophenol. This indicates that **7a** binds to the catalytic cysteine exclusively. The small abundance of the peak is in accordance to the restricted irreversible binding described above. The specific binding of the thiophenolmethyl ketone warhead is further confirmed by **C5**, for which also only a single adduct peak of 21426.13 Da could be detected. The mass shift of 164 Da the same as detected for the CVB3 3C^{pro} and described in subsection 4.4.3, which further confirms the proposed binding mechanism.

The specificity of the discovered α-thiophenolmethyl ketone warhead was further investigated through its reaction with the nucleophile glutathione (GSH) (Figure 4.50B). This biologically ubiquitous tripeptide contains a cysteine thiol moiety

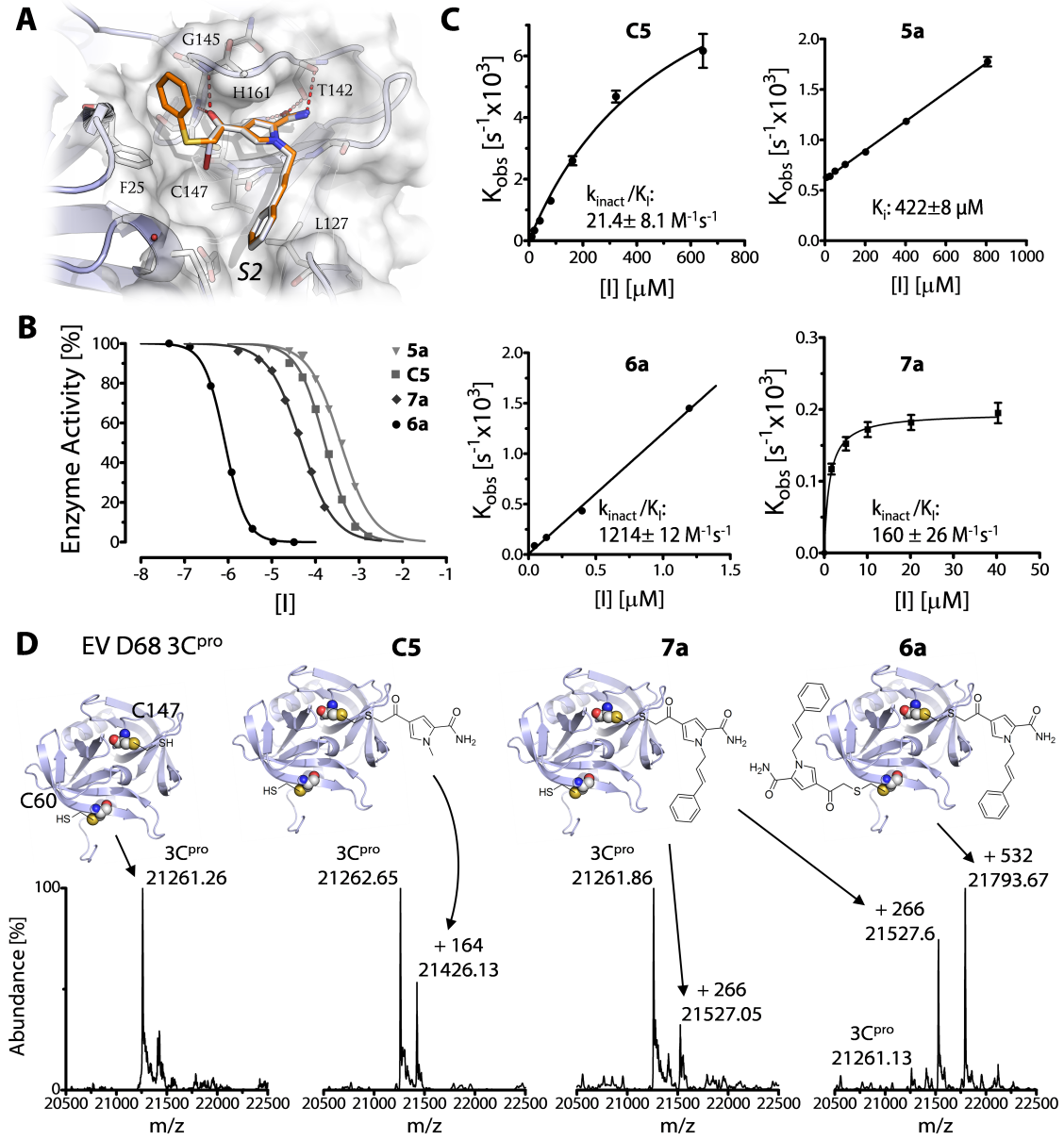

Figure 4.49: EV D68 3C^{pro} binding and inhibition by pyrrole-based inhibitors. (A) Superimposed putative non-covalent binding conformations of **5a**, **6a** and **7a** in EV D68 3C^{pro} active site. Hydrogen bonds are depicted as red dotted lines. (B) Dose-response curve for **5a** (triangle), **C5** (square), **7a** (diamond) and **6a** (circle). (C) Plot of k_{obs} as a function of [I] for **C5**, **5a**, **6a** and **7a**. Activity is given as K_i for reversible and as k_{inact}/K_I for irreversible inhibitors. (D) Deconvoluted ESI/QTOF mass spectra of EV D68 3C^{pro} apo structure and after incubation with **C5**, **7a** and **6a** (from left to right). Main peaks are labeled with m/z. Schematic depictions of EVD68 3C^{pro} or 3C^{pro}-inhibitor adducts are shown to indicate binding to catalytic C147 or non-catalytic C60, respectively. Note that the single adduct peak of **6a** results from binding to either cysteine residue.

and can thus react in analogy to the catalytic cysteine of the 3C^{pro}. In addition to the active ligands **C5** and **7a**, tangible thiophenolmethyl ketone containing fragments were included in the analysis. The indole **C8** was obtained from the in-

house small molecule collection, whereas the acetophenon-based thiophenolmethyl ketone **C9** was kindly prepared through a single-step synthesis from bromo-acetophenon and thiophenol by Dr. Stefan Wagner from the group of Prof. Rademann. Both compounds do not show activity against the CVB3 $3C^{pro}$ in the micromolar range. Furthermore, the synthesized dithiophenolmethyl ketone **8a** and the bromomethyl ketone **6a** were analyzed (Figure 4.50A). Each compound was incubated with GSH at equimolar concentrations. The reaction was monitored via LC/MS for 24 h (Figure 4.50C). The formation of new entities or a reduction of the initial compound could only be detected for **6a** within the analyzed time period

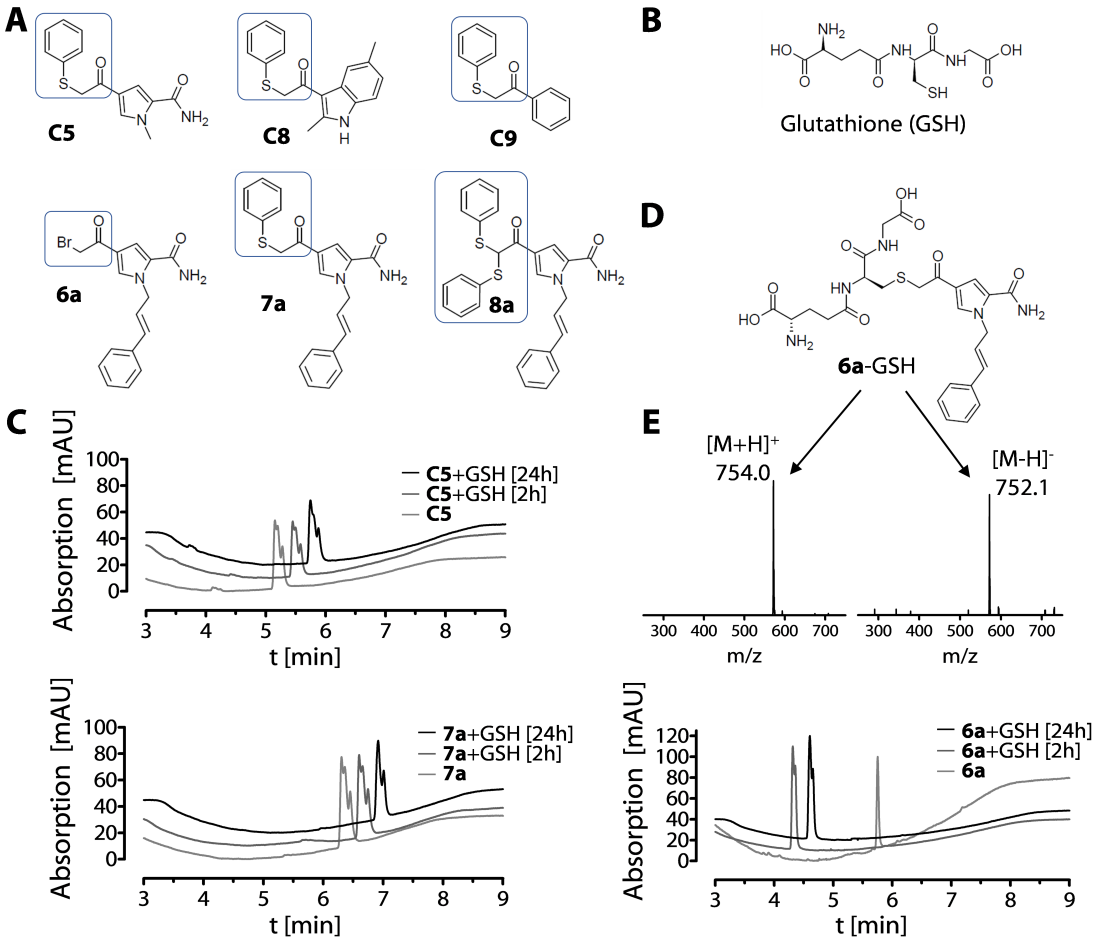

Figure 4.50: Glutathione assay for warhead reaction specificity. (A) Assembled set of α-thiophenolmethyl ketones and synthesized analogues of **7a**. Warhead substructure is highlighted by blue frame. (B) Chemical structure of the nucleophile glutathione (GSH). (C) Superimposed LC chromatograms of **C5** (top) and **7a** without (gray) and after incubation with GSH for 2 h (dark gray) and 24 h (black). (D) Chemical structure of the putative reaction product **6a**-GSH in neutral state. (E) (Top) Mass spectra of the reaction product **6a**-GSH in the positive (right) and negative mode (left) with relevant ions and m/z. (Bottom) Superimposed LC chromatograms of **6a** without (gray) and after incubation with GSH for 2 h (dark gray) and 24 h (black).

(Figure 4.50E). For **6a**, a second peak with a smaller retention time appears, whereas the peak of the original compound vanishes. This indicates the formation of reaction product between **6a** and GSH. This is supported by the $\mathrm{m/z}$ of the ions in the MS spectra of the new peak, which fit the calculated $\mathrm{m/z}$ of a reaction product formed through a nucleophilic attack of the thiol moiety at the α-bromomethyl ketone group and elimination of the bromine (Figure 4.50D). This reactivity of **6a** is known and was used for the synthesis of the desired compound **7a**. It also served as positive control for this analysis as the same GSH-adduct with the same $\mathrm{m/z}$ and retention time would have been formed by **7a**. This indicates a general stability of the thiophenolmethyl ketone warhead against non-catalytic nucleophiles and is in accordance to the MS data for the specific binding to the catalytic C147.

In order to further optimize the pyrrole-based $3C^{pro}$ inhibitors, a variation of the warhead moiety was pursued. The apparent suboptimal binding affinity of **7a** putatively results from the transition from the reversible thiohemiketal to the irreversible thioether and the occurring rearrangement. Therefore, warhead structures were envisioned, which do not show transitions during covalent binding. This includes trifluormethyl ketones and α-ketoamides as activated ketone derivatives and Michael acceptors as common warheads for $3C^{pro}$ inhibition (Figure 4.51A). The synthesis of the Michael acceptor derivative of **7a** was achieved by Je-Ryun Lee as part of her internship in the group of Prof. Rademann under the supervision of Carolin Tauber. The steps in the pursued synthesis are similar to the scheme utilized in the first synthesis (Figure 4.51B). In contrast, the 2-formyl substituted derivate **9** of the initial building block was used as starting point to enable the subsequent formation of the Michael acceptor moiety. As first step, the phenylpropenyl side chain was introduced through nucleophilic substitution in acetone under basic conditions to obtain the substituted pyrrole **10**. Once more, the direct conversion of the ester to the primary amide did not work but also through the activation of the acid **11** via an acyl chloride the desired product could not be obtained. Therefore, dicyclocarbodiimde (DCC) and dimethylaminopyridine (DMAP) were used in dichlormethane (DCM) to activate the acid **11**, from which the primary amide **12** was generated after reaction with ammonia. Subsequently, the Michael acceptor **13** was synthesized by using the Horner-Wadsworth-Emmons reaction with NaH in DMF.

The biological evaluation only revealed modest activity for **13a** with an apparent $\mathrm{IC_{50}}$* around 500 µM (Figure 4.52C). The weak activity was surprising as Michael acceptors are well established as warhead for $3C^{pro}$ inhibition. The

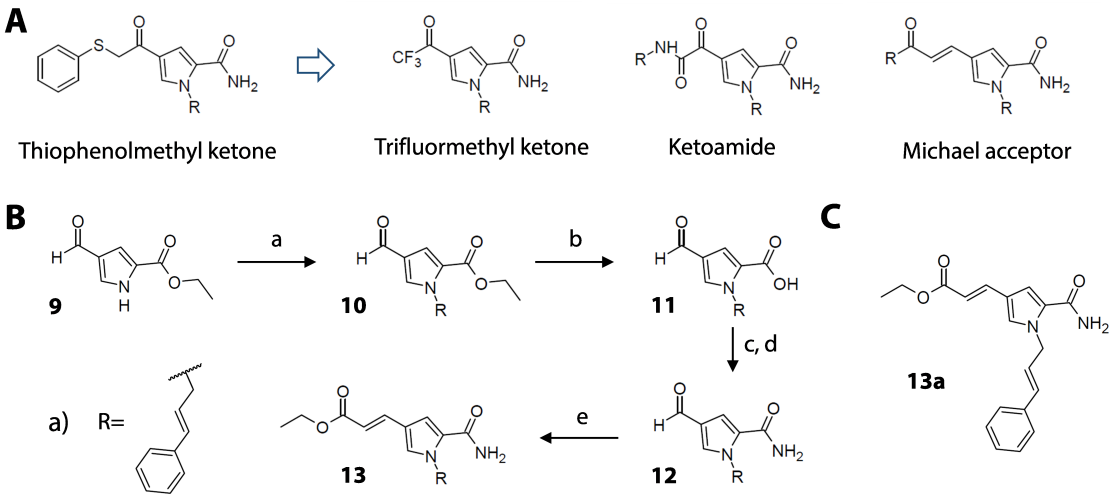

Figure 4.51: Fragment growing warhead optimization. (A) Proposed structures of pyrrole-based $3C^{pro}$ inhibitor scaffold with putative optimized warheads. (B) Steps, reagents and conditions for the synthesis of the Michael acceptor containing analogues: (a) K_2CO_3, TBAI, alkylhalide, acetone; (b) NaOH, EtOH; (c) DCC, DMAP, DCM; (d) NH_3, THF; (e) $(EtO)_2POCH_2CO_2Et$, NaH, DMF. (C) Chemical structure of the synthesized Michael acceptor containing compound **13a**.

progress curve shows no time-dependent activity indicating that **13a** does not form a covalent bond with the $3C^{pro}$. However, structural modeling analysis, which guided the selection of suitable warheads, showed that **13a** can bind in the $3C^{pro}$ active site in accordance to the developed binding model by forming the crucial interactions in the S1 and the S2 pockets in both binding states. In order to rationalize the experimental data, a reduction of the reactivity of the Michel acceptor moiety of **13a** was hypothesized. This could result from an instability of the ester group in the aqueous assay buffer. The free acid (**13a***) would be less reactive due to the diminished electron withdrawing effect of the ionized carboxyl group on the β-carbon (Figure 4.52A). Therefore, stability experiments were conducted, which did not show instability for **13a**. This analysis was further extended by investigating the reaction of **13a** with GSH as performed for the thiophenolmethyl ketones. In theory, the thiol moiety of GSH should react with **13a** through a Michael addition reaction leading to the putative adduct **13a**-GSH depicted in Figure 4.52D. However, even after 24 h of incubation, no reaction product could be detected. This also indicates a diminished reactivity of the Michael acceptor moiety in **13a**. Another putative reason for this could be the vinylogous arrangement of the Michael acceptor moiety to the pyrrole-nitrogen (Figure 4.52A). This enables delocalization of the nitrogen lone pair to the α,β-unsaturated ester, which putatively reduces its electrophilicity and thus reactivity for Michael addition and ester hydrolysis ester.

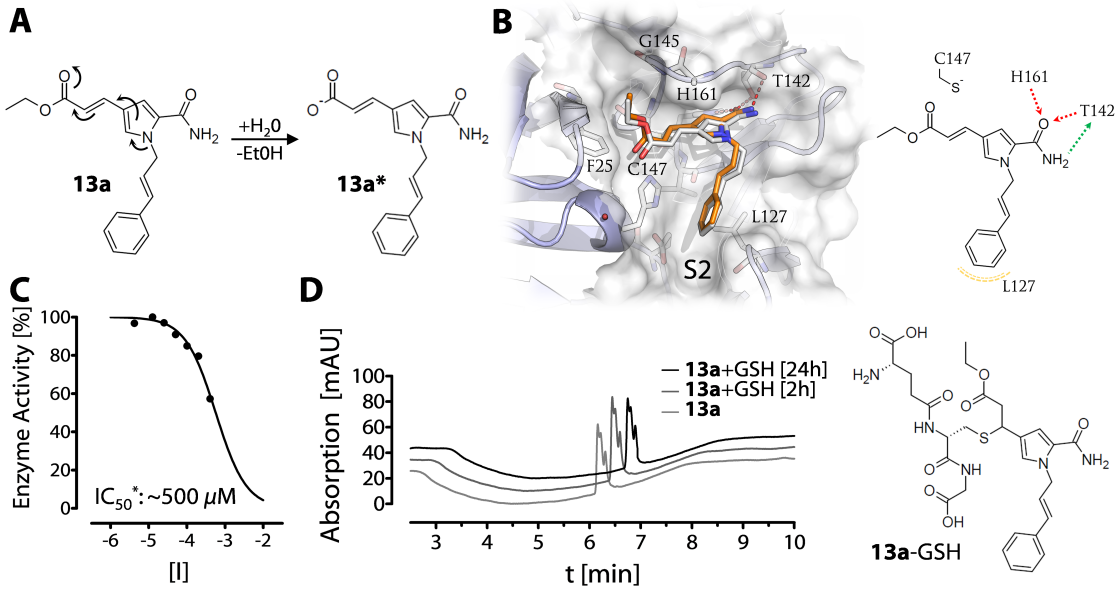

Figure 4.52: Binding and stability data of 13a. (A) Putative resonance between the Michael acceptor and pyrrole-nitrogen of **13a**. Chemical structure of a putative hydrolysis product **13a***. (B) Proposed binding mode of **13a** in the $3C^{pro}$ active site in 3D (left) and 2D (right). Covalent (orange) and non-covalent complex (white) are superimposed. Hydrogen bonds are depicted as red dashed lines in the 3D depiction. Interactions in the 2D depiction are colored as follows: Yellow lines indicate lipophilic contacts, green and red arrows indicate hydrogen bond donor or acceptor functionalities. (C) Dose-response curve of **13a** with apparent IC_{50}* value. (D) Superimposed LC chromatograms of **13a** without (gray) and after incubation with GSH for 2 h (dark gray) and 24 h (black). Chemical structure of putative Michael addition product of **13a** and GSH.

The developed *de novo* design workflow was used for the optimization of fragment hit **C5**. Based on virtually generated fragment growing analogues of **C5**, able to occupy the S2 pocket of the $3C^{pro}$, more potent $3C^{pro}$ inhibitors were obtained through chemical synthesis. The biological evaluation revealed general activity against enteroviral $3C^{pro}$ and further enabled the verification of the specific binding mode of the thiophenolmethyl ketone warhead.

Chapter 5

Results Part II: Allosteric Modulators

5.1 Biological evaluation

The scaffold hopping based on the fragment hit **F1**, as described in section 4.4, additionally led to fragment **C1**, whose biological activity as an allosteric $3C^{pro}$ inhibitor is distinct from the previously described covalently binding fragments and will thus be individually described in this chapter of the thesis. The following section will describe the biological characterization and validation of **C1** as well as its structural exploration through the biological evaluation of a series of fragment analogues.

5.1.1 Allosteric fragment hit identification and validation

The 4-aminoisoxazole fragment **C1** comprises the same chemical functionalities necessary for binding to the S1 pocket as the other fragments identified through the developed structural model used for scaffold hopping (Figure 5.1A). Therefore, **C1** was selected for biological evaluation. Dose-response analyzes revealed a reasonable IC_{50} value of ~ 450 µM considering its small size with only twelve heavy atoms, which results in a LE of 0.39 (Figure 5.1B). This was even more impressive as the progress curve of the enzymatic reaction do not show a sign of time-dependent activity and thus no indication for covalent $3C^{pro}$ inhibition (Figure 5.1C). However, this made **C1** less attractive for the initially pursued approach due to the apparent advantages of covalent binding as described in section 1.3. But the temptingly high LE in combination with no signs of covalent inhibition triggered further investigations. In order to elucidate the binding characteristics of **C1**, the inhibition modality was determined. Dixon and Cornish-Bowden plots indicate a noncompetitive mode of action with a K_i of 435 ± 5.5 µM and thus no binding to the active site, which would lead to competition with the substrate (Figure 5.1D,E). This surprising result renders fragment **C1** even more interesting as allosteric inhibition of the CVB3 $3C^{pro}$ had not been described to that date.

The high concentrations used in fragment screenings make these approaches prone to false-positives. Therefore, further verification of the activity of **C1** was performed. Due to the non-covalent mode of action of **C1**, mass spectrometry, employed for validation of the other fragment-based inhibitors developed in this thesis, could not be use. However, we exploited the possibility to covalently modify the $3C^{pro}$ by using a fluorescein-tagged covalent label synthesized by Franziska Gottschalk in the group of Prof. Rademann (Figure 5.1F). This peptide is composed of the $3C^{pro}$ substrate recognition sequence $EALFQ$ and contains a diazo-

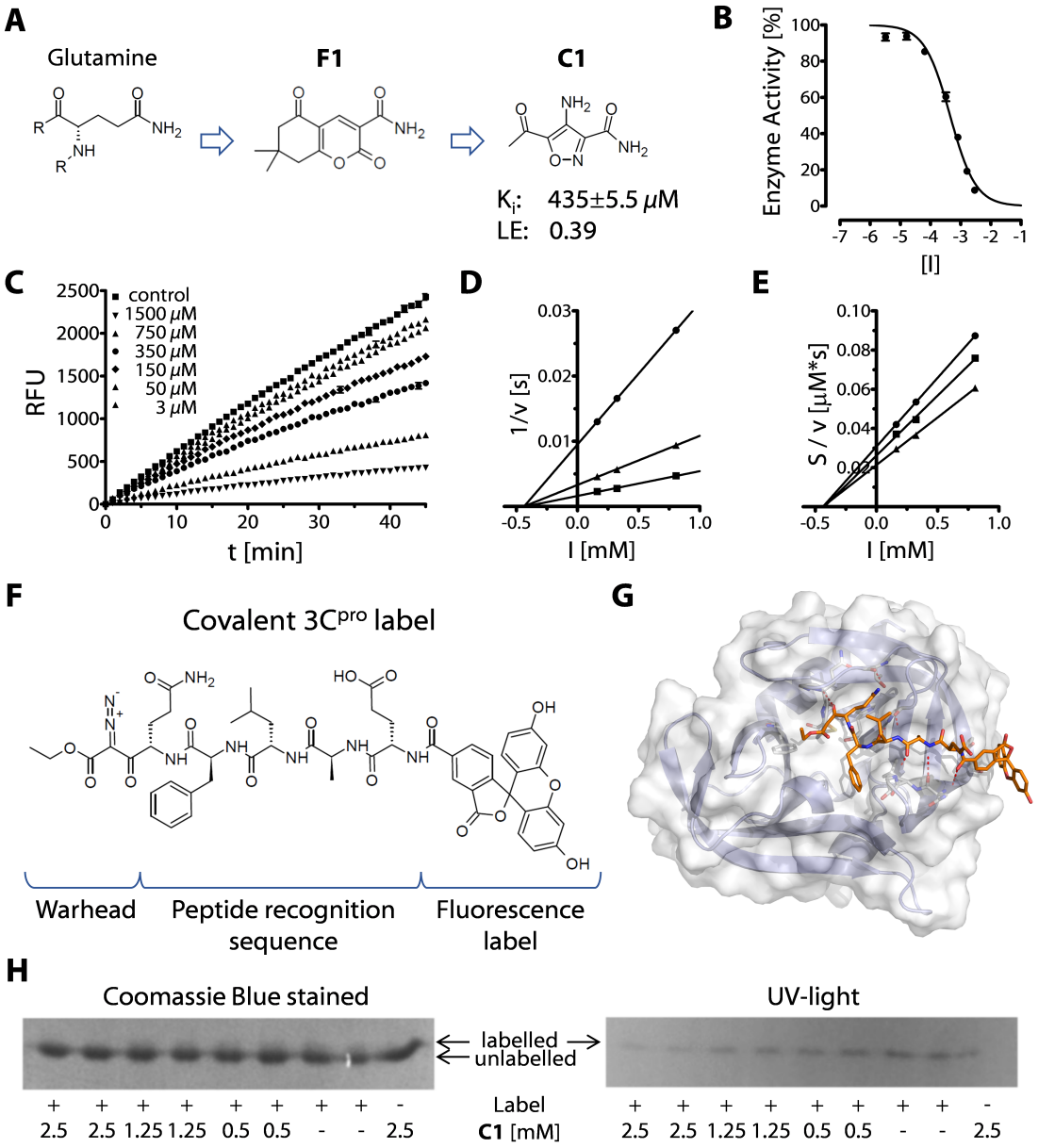

Figure 5.1: Biological characterization of C1. (A) Identification of **C1** as glutamine mimetic through scaffold hopping from **F1** and K_i and LE values. (B) Dose-response curve of C1. (C) Enzyme reaction progress curve of the FRET-substrate 3C^{pro} assay with different concentrations of **C1**. (D,E) Dixon and Cornish-Bowden plots for **C1**. (F) Chemical structure and composition of the fluorescein-tagged covalent 3C^{pro} label. (G) Model of the covalent-label bound in 3C^{pro} active site. (H) SDS-polyacrylamide gel of CVB3 3C^{pro} after incubation with the probe with or without **C1** detected by Coomassie Blue staining (left) or UV-light (right).

methylketone warhead. This can react irreversible with the catalytic C147 of the 3C^{pro} through the same two-step mechanism as other α-substituted methyl ketones such as the thiophenolmethyl ketones developed in this thesis (Figure 5.1G). To assess the effect of **C1** on the covalent modification, which can be detected

through the incorporated fluorescein tag, the $3C^{pro}$ was incubated with the covalent label in the presence of different concentrations of **C1** and evaluated in a SDS-polyacrylamide gel electrophoresis. Detection was performed using UV light for the labeled $3C^{pro}$ and Coomassie Blue staining for the modified and apo $3C^{pro}$, respectively (Figure 5.1H). The intensity of the fluorescence band of the labeled $3C^{pro}$ increased with decreasing concentrations of **C1**. Likewise, the band of Coomassie Blue stained labeled-$3C^{pro}$, visible as a small band above the unmodified $3C^{pro}$, increased with decreasing concentrations of **C1**. This concentration-dependent negative effect of **C1** on the covalent modification of the $3C^{pro}$ by the covalent label is in analogy to the inhibition of the catalytic activity measured in the FRET-substrate assay. This suggests a specific mode of action of **C1** independent from the utilized evaluation assay and justifies further structural exploration of the **C1** scaffold.

5.1.2 Structure exploration: 4-aminoisoxazole SAR

In order to further explore **C1** and to obtain initial information about key structural characteristic necessary for allosteric binding, readily available fragments with the same scaffold were collected from the in-house small molecule and commercial fragment collections for biological evaluation (Table 5.1). The structural variations represented by the selected fragments include the following: (i) different acyl substituents, (ii) amide substitution or replacement with a nitrile moiety and (iii) acetylation of the primary amine (Figure 5.2). The benzoyl substituted

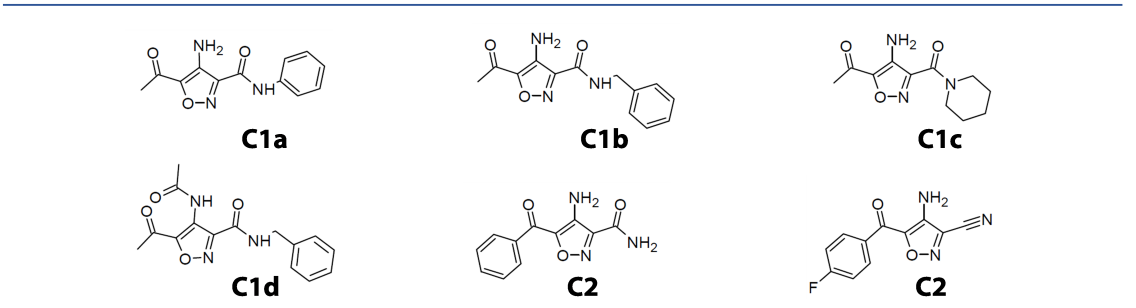

Figure 5.2: 4-Aminoisoxazole analogues of C1.

fragments (**C2**, **C2a**) could not be evaluated in the FRET assay due to their intrinsic fluorescence. From the other fragments, three showed concentration-dependent $3C^{pro}$ inhibition with noncompetitive binding kinetics (Figure 5.3A-C). Fragment **C1c**, whose amide is composed of a piperidine ring, is less active than **C1** with a K_i of 704 ± 14 µM. The benzyl substituted fragments **C1b** and **C1d**, which is in

Table 5.1: Binding characteristics of 4-aminoisoxazole analogues of C1. Inhibition indicates reduction of $3C^{pro}$ activity at a fragment concentration of 0.8 mM. Fluorescent indicates intrinsic fluorescence of the respective fragments and thus no biological evaluation. Nd stands for 'not determined' and indicates that the binding kinetics were not determined due to limited activity or intrinsic fluorescence. HA indicates the number of heavy atoms and LE the ligand efficiency.

ID	Inhibition or K_i [μM]	HA	LE	Kinetics
C1	435 ± 6	12	0.39	noncompetitive
C1a	10 %	18	-	nd
C1b	380 ± 6	19	0.25	noncompetitive
C1c	704 ± 14	18	0.27	noncompetitive
C1d	250 ± 10	22	0.23	noncompetitive
C2	fluorescent	17	-	nd
C2a	fluorescent	18	-	nd

addition acetylated at the primary amine, are more active than **C1** with K_i values of 380 ± 5.8 μM and 250 ± 10 μM, respectively. The phenyl substituted analogue **C1a** only displayed 10 % inhibition at 0.8 mM and was therefore deemed inactive without determination of its inhibition modality.

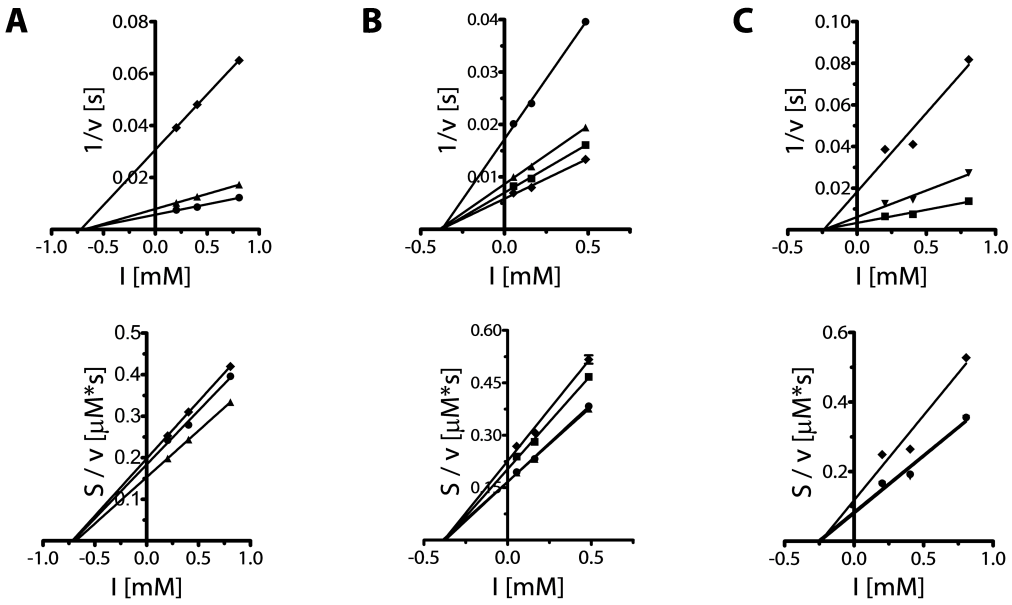

Figure 5.3: Biological characterization of C1 analogues. Dixon- (top) and Cornish-Bowden (bottom) plots of **C1c** (A), **C1b** (B), **C1d** (C).

The biological evaluation of the **C1** analogues enabled the identification of additional allosteric $3C^{pro}$ inhibitors. The obtained data indicate that the allosteric modulation results from the shared scaffold but in a structure-dependent manner as the substituents affect the activity in a reasonable fashion. However, none

of the structural variations improved the activity significantly, which is indicated by the reduction in LE (Table 5.1). The missing biological data for the fluorescent analogues limited the information gained about the relevance of the different moieties, especially the ketone part of the scaffold, for allosteric binding to the $3C^{pro}$.

5.1.3 Structure exploration: Extended analogue SAR

To enable further structural exploration of **C1**, the assembly of an additional set of analogues was pursued. Hereby, more general similarity criteria were used, which represent the shape- and electrostatic complementary essential for a ligand binding. Electrostatic complementary was described by pharmacophore features through a ligand-based 3D-pharmacophore in LigandScout [164]. To identify fragments with similar shape the software tool ROCS was utilized [186]. This program overlays molecules based on their three-dimensional shape and assess the similarity based on the overlap of the respective conformation of the molecules. Both approaches were used to screen the in-house molecule collection and commercial fragment libraries to identify shape- and feature-based analogues of **C1** (Figure 5.4A). The final selection comprised 18 fragments (Figure 5.5).

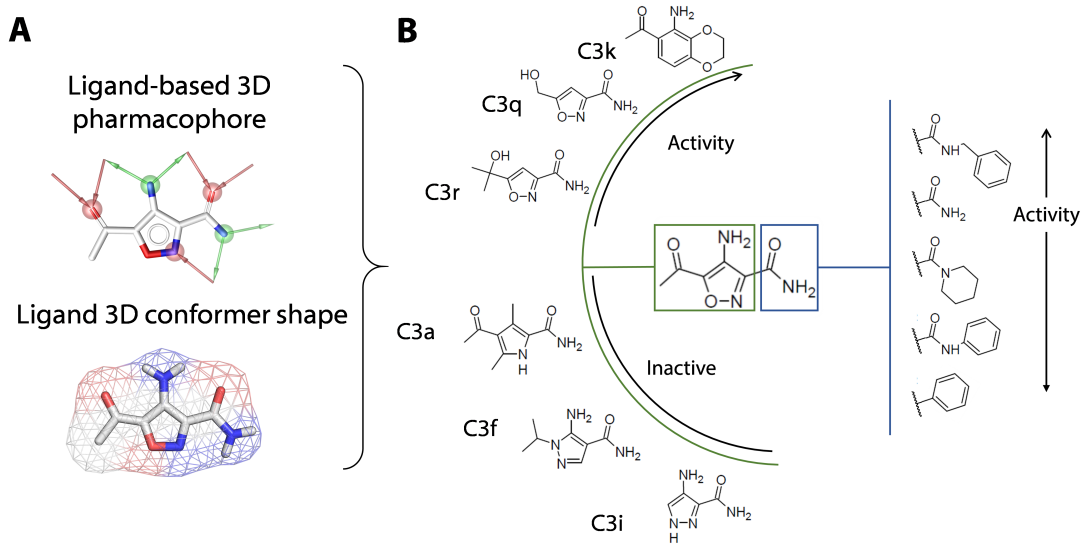

Figure 5.4: C1 analogue search and SAR analysis. (A) **C1** analogue search using ligand-based 3D pharmacophore (top) and shape-based queries (bottom). (B) Key findings of SAR for **C1** with representative fragments and substructures shown.

In the biological evaluation, eleven fragments were measured inactive, three could not be evaluated due to intrinsic fluorescence (**C3c**, **C3n**, **C3o**) and the remaining four showed biological activity (Table 5.2). This includes **C3p**, which

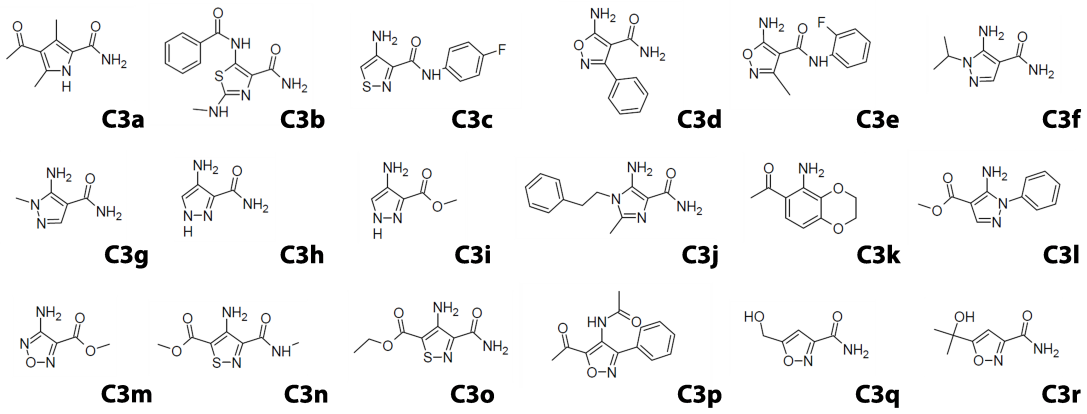

Figure 5.5: Biologically evaluated shape- & feature-based analogues of C1.

Table 5.2: Binding characteristics of extended SAR analogues of C1. Inhibition indicates reduction of $3C^{pro}$ activity at a fragment concentration of 0.8 mM. Fluorescent indicates intrinsic fluorescence of the respective fragments and thus no biological evaluation. nd stands for 'not determined' and indicates that the binding kinetics were not determined due to limited activity or intrinsic fluorescence of the fragment.

ID	Inhibition or K_i	Kinetics	ID	Inhibition or K_i	Kinetics
C3a	≤ 10 %	nd	**C3j**	≤ 10 %	nd
C3b	≤ 10 %	nd	**C3k**	0.8 mM	noncompetitive
C3c	fluorescent	nd	**C3l**	≤ 10 %	nd
C3d	≤ 10 %	nd	**C3m**	≤ 10 %	nd
C3e	≤ 10 %	nd	**C3n**	fluorescent	nd
C3f	≤ 10 %	nd	**C3o**	fluorescent	nd
C3g	≤ 10 %	nd	**C3p**	~ 3 mM	competitive
C3h	≤ 10 %	nd	**C3q**	2.1 mM	noncompetitive
C3i	≤ 10 %	nd	**C3r**	15 %	nd

is an analogue of **C1d** except for the substitution of the amide to a phenyl moiety. It shows a K_i around ~ 3 mM with competitive binding kinetics and thus no allosteric mode of action. **C3q** and **C3r** contain a hydroxymethyl and a hydroxyisopropyl group, respectively, instead of the acetyl moiety of **C1** and no primary amine group. The smaller fragment **C3q** is more active than its homologue **C3r** with a K_i of 2.1 mM and noncompetitive binding kinetics. Due to the limited activity of **C3r**, which inhibits the $3C^{pro}$ by only 15 % at 0.8 mM, its inhibition modality was not determined. The o-aminoacetophenon derivative **C3k** shows a K_i of 0.8 mM and noncompetitive binding kinetics.

The obtained biological activity data indicate the following structural characteristics to be essential for allosteric inhibition of the $3C^{pro}$ (Figure 5.4B). (i) The

amide moiety is probably important due to its hydrogen bond donor/acceptor functionality as double substitution of the amide reduces activity (**C1c**) and elimination of the amide abolishes activity (**C3p**). (ii) Lipophilic substituents of the amide do not improve activity significantly indicating solvent exposure (**C1b**, **C1d**). (iii) The primary amine moiety is important as analogues devoid of it are less (**C3q**) or not active (**C3a**). (iv) The ketone moiety putatively acts as hydrogen bond acceptor since isosteric replacements lead to inactive fragments (**C3f**), whereas exchange to a hydroxyl moiety maintains activity (**C3q**). (v) The binding location of the ketone moiety is probably space limited as big substituents are not tolerated at this position and the hydroxymethyl substituent in **C3q** is more active than the hydroxyisopropyl substituent in **C3r**. The o-aminoacetophenon derivative **C3k** shows some structural divergence in comparison the other allosterically active fragments, yet it shares the same substitution pattern of an acetyl moiety in ortho-position to an aromatic amine. This seems to be the key motif as the fragment **C3q**, which is missing this substructure, is one order of magnitude less active.

The key structural characteristics for allosteric $3C^{pro}$ inhibition were elucidated through the conducted SAR studies of the fragment analogues of **C1**. Yet, structural understanding of the allosteric binding to the $3C^{pro}$ was still in the dark.

At the same time Kim *et al.* [102] reported that the anti-parkinsonian drug benserazide is an allosteric inhibitor of the CVB3 $3C^{pro}$ with a K_i of 4.7 μM (Figure 5.6A). They performed a small SAR with benserazide analogues comprising different substitution patterns of the hydroxylated benzene ring of which only the 3,4-dihydroxy analogue **9e** was active with a IC_{50} of 50 μM (Figure 5.6B). The proposed binding conformations obtained from protein-ligand docking neither explain the SAR of the benserazide analogues nor the allosteric effect on the catalytic activity of the $3C^{pro}$. Furthermore, the structural models, which were made available in the supplementary data of the publication, show an inappropriate ionization state of the ligand, precisely a neutral state of the aliphatic amine. This renders the model less reliable as the proposed allosteric binding area involves multiple basic residues (Figure 5.6C).

Nonetheless, the reported biological data of the benserazide SAR provided additional information that were combined with the fragment SAR in order to generate structural models, which are able to rationalize allosteric modulation of the CVB3 $3C^{pro}$.

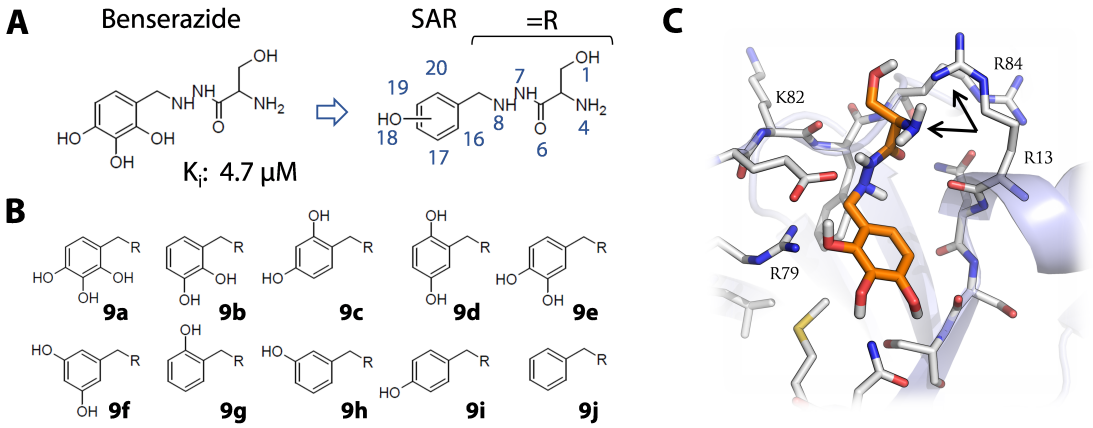

Figure 5.6: Benserazide SAR and proposed binding mode reported by Kim et al. [102]. (A) Chemical structure, biological activity and atom numbering of benserazide. (B) Structures of benserazide analogues utilized for SAR analysis. (C) Proposed binding mode of benserazide reconstructed from published model. Arrows indicate the spatial proximity of the primary amine of benserazide and the guanidine moiety of R13.

5.2 Computational studies

The identification of allosteric $3C^{pro}$ modulators, described in the previous section, provides novel insights and opportunities to address the CVB3 $3C^{pro}$. In order to improve the significance of these findings and to exploit it for the design of potent allosteric $3C^{pro}$ inhibitors a rationalization through structural models was necessary. The following section describes the application of computational methods to address the following questions: (i) Where are putative allosteric binding locations, (ii) what are the binding modes of the allosteric inhibitors and (iii) how does binding to an allosteric pocket affect the active site and thus the catalytic activity of the $3C^{pro}$?

5.2.1 Allosteric pocket identification

At first, for the identification of putative allosteric binding sites, the structure of the CVB3 $3C^{pro}$ was investigated using the FTMap web server application [230]. Hereby, small molecule probes, which represent different possibilities to interact with a protein, are docked all-over the protein surface. The obtained poses are energy-ranked and clusters of different probe molecules are considered as binding hot spots. As input, apo and co-crystal structures of the $3C^{pro}$ were used as the noncompetitive nature of the allosteric ligands suggests equal affinity to both states. Binding hot spots proposed by FTMap were almost identical for both protein structures, which is due to their high structural similarity indicated by an Cα RMSD $\sim$ 0.5 Å. High ranked probe clusters are located in the active site pockets S1, S2 and S1' as well as several non-orthosteric sites (Figure 5.7A). The top ranked allosteric cluster is located in an area that has been shown to be involved in viral replication through the interaction with stem loop structures in the viral RNA as described in subsection 1.2.3. Through investigations of the RNA-$3C^{pro}$ interaction of different enteroviruses, the involved binding site and residues were identified [231, 232]. In addition, it was shown that RNA binding and proteolytic activity mutually effect one another [77, 233]. This observation and the general function of the area suggested as putative binding hot spot would provide biological implications for allosteric ligand binding. Therefore, we surmised that the identified allosteric $3C^{pro}$ inhibitors bind to this site.

In order to identify putative binding conformations of the allosteric ligands at this hot spot protein-ligand docking was used. However, the obtained poses did not show an enrichment of a distinct binding mode, neither for the fragments nor

for the benserazide analogues, which would explain the SAR findings. Therefore, MD simulations were conducted to (i) enable higher conformational sampling of the $3C^{pro}$ than in the rigid docking as allosteric modulation typically involves binding to different protein conformations and (ii) to obtain statistical information of the distribution of the ligand binding modes generated through protein-ligand docking. Representative docking poses were used as starting geometries for the

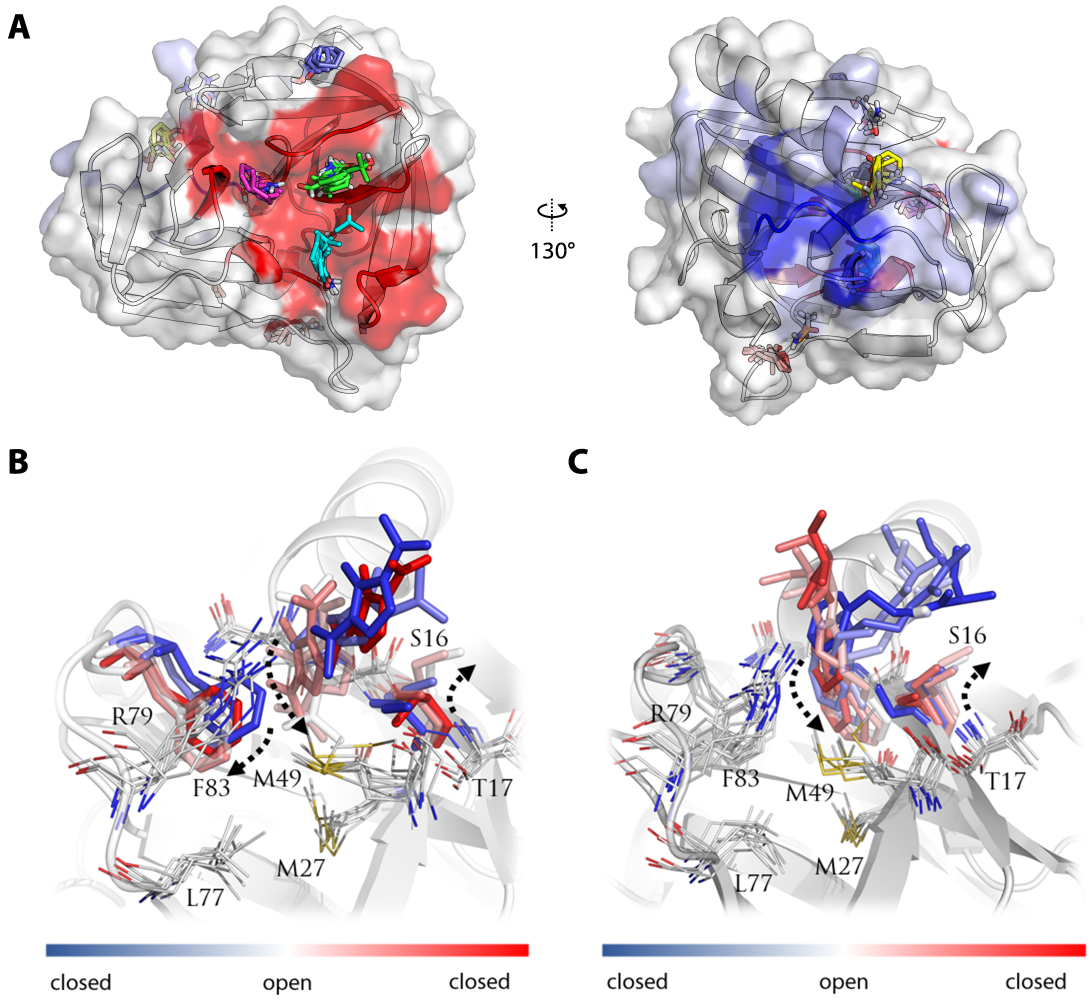

Figure 5.7: $3C^{pro}$ binding hot spots and transient allosteric pocket identification. (A) Top ten binding hot spots of $3C^{pro}$ determined by FTMap web server [230]. Clusters bound to active site pockets S1, S2 and S1' are colored green, cyan and purple, respectively. Cluster at RNA binding site is colored yellow. Surface coloring indicates active site in red, residues involved in RNA binding in light blue and the conserved KFRDI motif in blue. The binding of **C1** (B) and the benserazide analogue **9j** (C) to the transient allosteric pocket from superimposed snapshots of MD simulations.

MD simulations. These did not lead to a prioritization of a distinct binding mode as they showed little stability but revealed a transient allosteric pocket at the RNA binding site, which was induced by **C1** as well as the benserazide analogue **9j** (Fig-

ure 5.7B,C). The transient pocket is located between the β-sheets at the top of the β-barrel within the RNA binding site and formed by the side chains of the residues S16, M27, M49, R79, F83 and the backbone of the residues N14-T17. Their slight movement facilitates an rearrangement of the side chain of S16, which opens up the transient pocket. Additionally, the phenyl moiety of F83, which belongs to the conserved KFRDI motif, slightly turns.

5.2.2 Binding mode rationalization

This transient pocket provides a rational for the binding of small molecules, such as the identified allosteric $3C^{pro}$ modulators, in the RNA binding area. Therefore, $3C^{pro}$ conformations with the identified allosteric pocket in the open state were derived from the respective MD trajectories and serve as starting points for a second protein-ligand docking study.

For the benserazide series of ligands, the obtained docking poses show a distinct binding mode. The phenol moiety is placed in the transient pocket by forming lipophilic contacts with M27, M49 and F83, whereas the serine substructure of the R-enantiomers can be well accommodated on top of pocket by forming hydrogen bonds to R13, R79, E81, R84 and a coulomb interaction to E81 via the positively charged primary amine (Figure 5.8A,B). The different hydroxylation patterns of the benzene moiety of the ligands lead to slight variations in the docking poses. On the one hand, this is due to the variability in the formation of hydrogen bonds but on the other hand differences also result from the spatial requirements of the hydroxyl groups. Meta- and para-substitutions as in the benserazide analogues **9c**, **9d** and **9f**, are less well accommodated in the allosteric pocket. However, the SAR shows that not only these analogues are inactive but also all others except the 3,4-dihydroxy derivative **9e**.

MD simulations were performed to elucidate the effect of the hydroxylation pattern on the stability of the identified binding pose in the allosteric pocket. The above-described binding conformation of the benserazide analogues within the allosteric pocket served as starting geometry. The interactions formed by the ligands were analyzed using dynamic 3D pharmacophores (dynophores). This method was developed by Dominique Sydow in our group and extends the classical concept of describing protein ligand interactions through pharmacophore features by introducing statistical significance from MD simulations [234]. Thereby, interaction frequencies or distribution of binding modes characterized by specific interaction-patterns can be identified. 100 ns MD simulations with benserazide

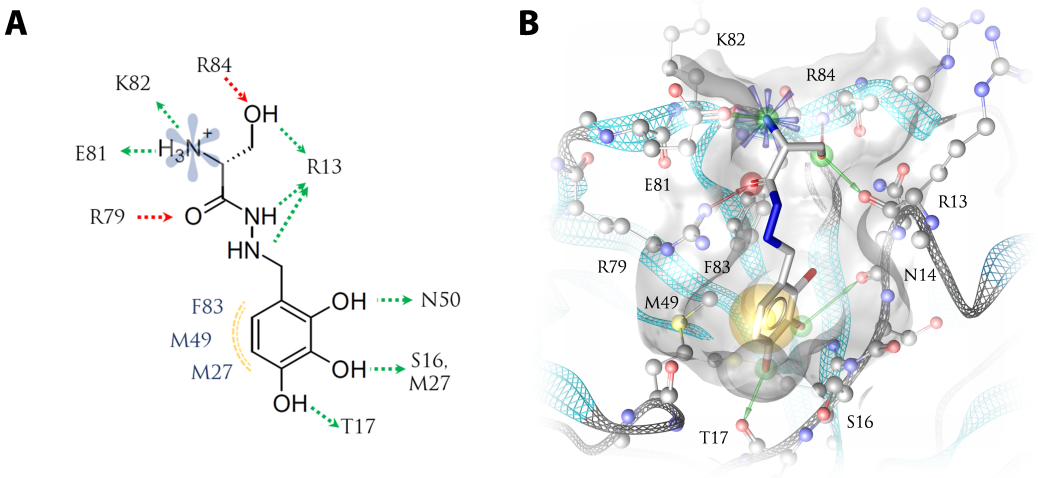

Figure 5.8: Static allosteric binding model of benserazide. Docking-based binding conformation of benserazide in the allosteric pocket of $3C^{pro}$ depicted in 2D (A) and 3D (B). Interactions are described as pharmacophore features. Yellow spheres indicate lipophilic contacts, blue star indicates coulomb interactions of a positively ionized group, green and red arrows indicate hydrogen bond donor or acceptor functionalities, respectively.

showed pose stability throughout the simulation runs (Figure 5.9A). This is reflected by the generated dynophore, which shows high frequencies for the interaction also present in the docking-based start geometry (Figure 5.9B). From the 2,3,4-trihydroxy benzene moiety, the 4- and 3-hydroxy groups frequently form hydrogen bonds, whereas the 2-hydroxy group, missing in the active analogue **9e**, interacts only in $\sim$ 10 % of the trajectory frames with the protein. In addition, less frequent hydrogen bonds are also formed by the hydrazide nitrogen atoms. Dynophores were also generated for the trajectories of the MD simulations of the benserazide analogues and the interaction patterns and frequencies were compared (Table 5.3). Surprisingly, only the analogue **9e**, which is also active, shows binding pose stability and interaction frequencies similar to benserazide (Figure 5.9A). The other analogues were either not stable in the proposed binding mode, indicated by overall small interaction frequencies (analogues **9b**, **9c**, **9d**) or less stable than the active ligands (see Appendix B.1). This resulted in high variations between the frequencies of the different interactions due to alteration of the binding mode, whereas a certain trend is apparent. The mono-hydroxylated analogues **9h** and **9i** as well as the non-hydroxylated derivative **9j**, which all represent substructures of the active ligands **9a** and **9e**, show greater stability than the other inactive compounds. However, Kim *et al.* did not report an activity cut off for the biological testing. Therefore, it remains unclear if differences exist between the analogues flagged as inactive and thus the biological relevance of the identified

Table 5.3: Benserazide analogue interaction frequencies from dynophore analysis. Interaction frequencies determined through dynophore analysis of MD simulations for benserazide and its analogues bound in the allosteric pocket of $3C^{pro}$. Interactions are described as pharmacophore features in which HBA denotes hydrogen bond acceptor, HBD stands for hydrogen bond donor, PI denotes coulomb interactions of a positive ionizable group and lipophilic denotes lipophilic contacts to hydrophobic moieties. The numbering of the benserazide ligand series atoms (left) corresponds to figure 5.6.

ID	**9a**	**9b**	**9c**	**9d**	**9e**	**9f**	**9g**	**9h**	**9i**	**9j**
Interaction	**Interaction frequency [%]**									
O1-R84 HBD	66	0	0	1	87	82	66	56	38	81
O6-R79 HBA	95	0	0	13	98	97	56	75	43	68
O1-R13 HBA	88	0	0	10	90	71	58	61	33	72
N4-K82 HBA	93	0	3	0	85	51	15	47	39	27
N4-E81 HBA	86	20	0	1	93	48	24	49	47	47
N4-E82 PI	93	25	4	3	89	65	43	64	39	63
Lipophilic	99	2	0	100	99	0	100	92	60	98
O16 HBD	12	15	0	33	-	-	37	-	-	-
O17 HBD	55	15	-	-	40	-	-	82	-	-
O18 HBD	77	-	0	-	76	88	-	-	43	-
O19 HBD	-	-	-	71	-	88	-	-	-	-
O20 HBD	-	-	-	-	-	-	-	-	-	-

trend is uncertain.

The $3C^{pro}$ conformations derived from the MD simulations with benserazide bound in the allosteric pocket were compared to the conformation of the crystal structure for further analysis and verification. This revealed that the allosteric site hosts several well resolved crystal waters. Their binding locations and interactions are picked up by benserazide in the proposed binding hypothesis (Figure 5.10A). One of these water molecules bridges the backbones of R13 and R84 by acting as a hydrogen bond donor and acceptor at the same time. Benserazide replaces the water molecule with its hydroxyl group, which has the same functionalities. For the elucidation of the binding mode of the allosteric fragments, the water molecule plays a key role as will be described below. In the crystal structures the transient pocket is closed, yet a water molecule is bound next to the side chain of S16, whose outward movement leads to the opening of the pocket. Superpositioning of the benserazide conformation bound in the open transient pocket reveals that the hydroxyl groups at the 3 and 4 position of the benzene ring superimpose with the oxygens of the water and the S16 side chain, respectively. This serves as an explanation for the relevance of this hydroxylation pattern, which is only present

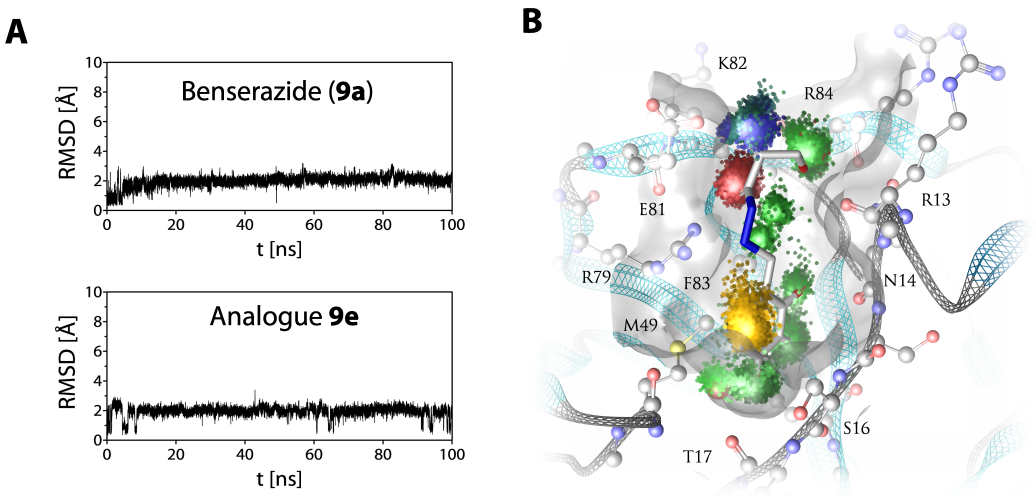

Figure 5.9: Dynamic allosteric binding model of benserazide. (A) RMSD plots of the MD simulations of benserazide (top) and the analogue **9e** (bottom). (B) MD trajectory derived dynophore mapped on representative benserazide binding conformation in allosteric pocket. Each point represents a specific interaction within a frame of the trajectory according to pharmacophore feature color coding. Point clouds represent the ensemble of interactions from the trajectory and indicate statistical significant interactions.

in the active compounds of this series and provides further confirmation for the relevance of the proposed binding hypothesis.

The MD simulations of the fragments show less stable binding conformations than benserazide, which reflects the difference in biological activity. Despite this, dynophore analysis of the trajectories indicate specific interaction patterns, which occur more frequently and thus were used to generate binding mode hypothesis for the fragment-based allosteric $3C^{pro}$ inhibitors (Figure 5.11A). For the prioritization of the putative binding conformations, the analysis of the crystal water binding spots described above was used in addition to integration of the SAR data. Two main states for the crystal water spot in between R13 and R84 could be differentiated from the MD simulations, in which either the respective fragment occupies this spot or water molecules. These showed stable occupation in general (Figure 5.10B,C). The fragments displaced the water through their amide moiety, but only the oxygen forms a hydrogen bond to R84, as the amide group itself is too big to fit as a water molecule or the hydroxyl group of benserazide. Therefore, we reasoned that this spot is more likely to be occupied by a water molecule than a fragment. This led to the final binding mode hypothesis depicted in Figure 5.11B,C. Hereby, the water can interact with the fragments in a conformation, which includes frequent interactions indicated by the generated dynophores and which are in accordance to the key SAR findings. In this model, the fragments

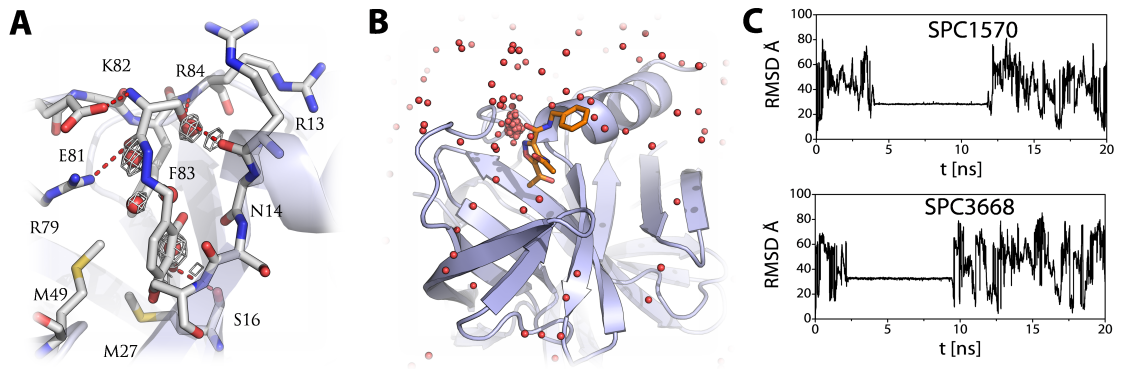

Figure 5.10: 3C^{pro} Crystal water in MD simulations. (A) Benserazide allosteric binding conformation superposed to the allosteric site of the crystal structure of 3ZYD with closed transient pocket. Crystal water molecules are depicted as red balls and their electron density map as gray mesh. (B) Superpositioning of stable water molecules of every 50th frame of a representative MD trajectory. (C) RMSD plots of stable water molecules from MD simulations.

bind to the transient pocket in a semi-opened state as the S16 side chain forms hydrogen bonds to both, the primary amine as well as the ketone moiety, whose methyl group fills the transient pocket. This explains the relevance of this substructure for the biological activity and its reduction for the less active hydroxymethyl derivative **C3q** and the almost inactive **C3r**. The former can form the hydrogen bond to S16 but the missing methyl group leaves the pocket insufficiently filled, whereas the hydroxyisopropyl moiety of the latter is too bulky to bind properly. The amide moiety forms hydrogen bonds with the side chain of R79, similar to the hydrazide carbonyl oxygen in benserazide, and with the backbone carbonyl of R13. The latter is lost in **C1c** due to the double nitrogen substitution. The benzyl substituents in **C1b** and **C1d** do not form any specific interaction as indicated by the respective dynophores. This is in accordance to the determined small increase in activity that this additional group provides. The phenyl substituent in the nearly inactive **C1a** probably induces steric clashes with the side chain of E81 due to its rigidity and thus abrogates binding. The included water molecule can form a hydrogen bond to the nitrogen in the isoxazole ring. This is missing in the active fragment **C3k**, which can otherwise bind similarly to the S16 in the transient pocket. The ether oxygens in the dioxane ring may act as hydrogen bond acceptors for the interactions with R79 and the water.

So far, two of the three initially raised questions concerning the allosteric modulation of the 3C^{pro} were addressed. The transient pocket within the RNA binding site provides a good possibility for the binding of small molecules. Despite the apparent differences, consistent binding models could be generated for both al-

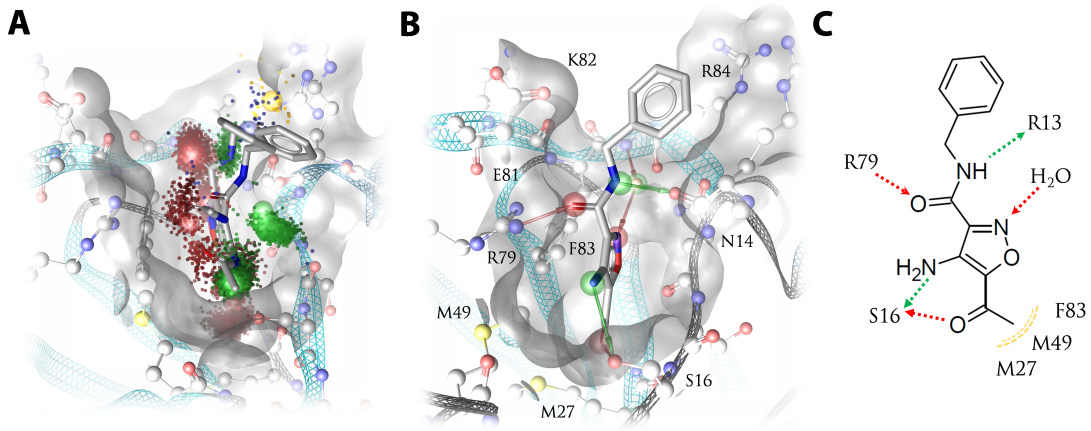

Figure 5.11: Allosteric binding model for C1-based fragments. (A) Dynophore generated from MD simulation of C1d with two representative binding conformations in the allosteric pocket of $3C^{pro}$. Each point represents a specific interaction within a frame of the trajectory according to pharmacophore feature color coding. Point clouds represent the ensemble of interactions from the whole trajectory and indicate statistical significant interactions. (B) Fragment binding mode in allosteric pocket illustrated by **C1b** with interactions represented by pharmacophore features. Green and red arrows indicate hydrogen bond donor or acceptor functionalities. (C) Interactions of C1b in allosteric pocket as 2D depiction.

losteric inhibitor series, which are able to rationalize allosteric binding.

However, inhibition of the catalytic activity through small molecule binding to the allosteric site has not been explained so far. Due to the noncompetitive mode of action the allosteric ligands bind to the apo structure of the enzyme as well as the enzyme substrate complex and subsequently block the catalytic cleavage of the substrate. Therefore, the allosteric binding does not affect the binding of the substrate but residues and parts of the enzyme, which are relevant for the catalytic reaction. Structural changes of the $3C^{pro}$ observed in the performed MD simulations could not be correlated with the binding of the allosteric ligands nor the opening of the transient pocket. The major structural alteration observed involved a rearrangement of the oxyanion hole (Figure 5.12A,B). A crystal water molecule is bound in close proximity to the oxyanion hole, in the groove between the two β-barrels. During some of the MD simulations, the water molecule leaves this position and the backbone carbonyl oxygen of C147 steps in, to form the same hydrogen bonds to T26. This is facilitated through a flip of the backbone of C147 and a rearrangement of the oxyanion hole forming loop comprised of the residues T142-C147. As a result, the backbone nitrogen atoms of C147 and G145 can no longer stabilize an oxyanion formed as transition state during the catalytic reaction. Thus, this conformation represents a less active state of the $3C^{pro}$, in which the formation of the enzyme substrate complex is unchanged. Crystal structures

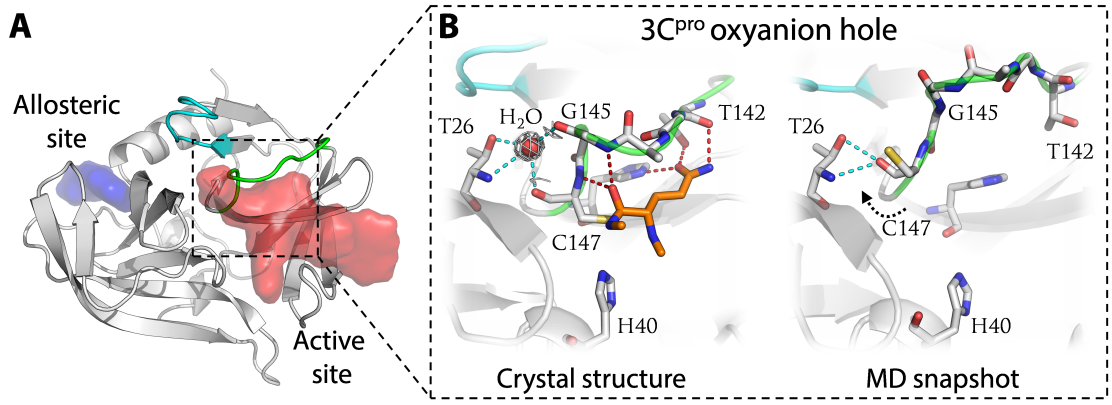

Figure 5.12: MD-based conformational changes of the 3C^{pro} oxyanion hole. (A) Structure of 3C^{pro} with the active site volume depicted in red, the allosteric site volume depicted in blue, the oxyanion hole forming loop colored in green and the adjacent loop formed by the residues N104-N111 is colored in cyan. (B) Close up on the oxyanion hole in the crystal structure (left) of the 3C^{pro} and in a representative snapshot from MD simulations (right). The crystal water molecule is depicted as red ball with its electron density as gray mesh. The oxyanion formed by the P1 of the substrate is depicted in orange. Hydrogen bonds in the back of the oxyanion hole are represented by dashed lines colored in cyan. Movement of C147 backbone atoms and the oxyanion hole forming loop (colored in green) in the MD simulations is indicated by a dashed arrow.

with improperly formed oxyanion holes have been reported for other viral proteases such as the West Nile virus NS3/2B protease [235]. A second loop, formed by the residues N105-N111, is located adjacent to the oxyanion hole as well as to the allosteric binding site. The flexibility of this loop allows it to work as a lid to control the leaving of the water molecule at the backside of the oxyanion hole. The direct interactions with the transient pocket forming β-sheet residues would provide a rationale for an effect of the binding of allosteric ligands on the stability of the oxyanion hole through this loop. However, the missing correlation of the structural changes in the MD simulations to the binding of the allosteric ligands renders this as a hypothesis.

5.2.3 Fragment merging

The generated binding models for the allosteric inhibitors provide a basis for their rational optimization, despite the missing structural explanation of the allosteric modulation of the 3C^{pro}. The structural models suggest differences in the interaction patterns between the fragments and benserazide, respectively its analogue **9e**. Interactions of the former are especially formed in the transient pocket with S16 and through the amide moiety. Additional substituents do not form favorable interactions on top of the transient pocket, which is in contrast to benserazide.

Its serine substructure forms stable interaction at this site. Therefore, both series could be effectively merged by combining the serine substructure with the core scaffold of the allosteric fragments.

In order to suggest putative structures for merged fragments, the carbonyl moieties of the serine substructure and the allosteric fragments served as an anchor point as both interact with the side chain of R79. However, they do overlap in a way that different structural variations are possible without forfeiting this interaction. For merging the serine substructure with the 4-aminoisoxazole scaffold different combinations were envisioned leading to ten virtual compounds (see Appendix B.2). Through protein-ligand docking their ability to bind in the allosteric pocket of the $3C^{pro}$ was assessed. For four of the virtual compounds (**M3**, **M4**, **M6**, **M9**), binding poses were obtained, which comprise the interactions of both inhibitors series (Figure 5.13A-C).

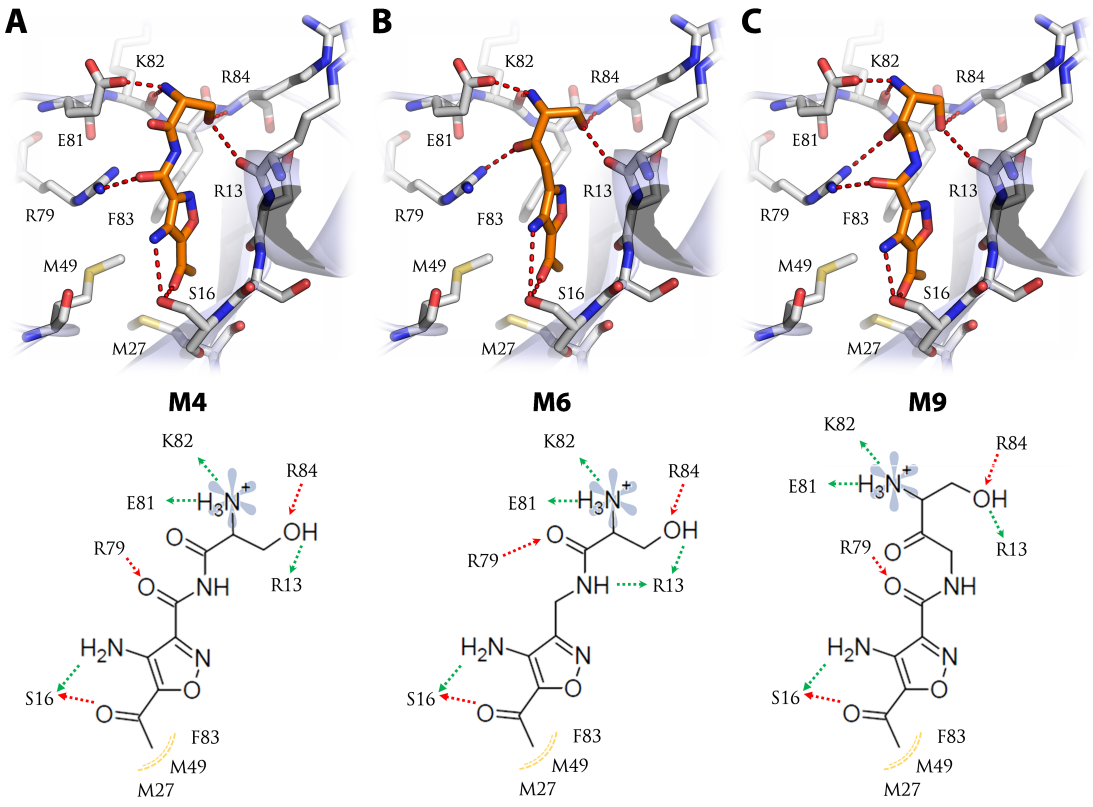

Figure 5.13: Fragment merging ligands binding modes. Binding modes of virtual fragment merging ligands **M4** (A), **M6** (B), **M9** (C) in the allosteric pocket of the $3C^{pro}$ depicted in 3D (top) and 2D (bottom). Interactions are described as follows: Yellow spheres indicate lipophilic contacts, blue star indicates coulomb interactions of a positively ionized group, green and red arrows indicate hydrogen bond donor or acceptor functionalities.

These four proposed ligands show favorable binding characteristic and thus

are promising structures for the design of allosteric $3C^{pro}$ inhibitors. Yet, these structures are only virtual and in contrast to the fragment growing approach, pursued for the covalently binding fragments, synthetic traceability was not directly included in the design. However, synthesis would not only be essential for the subsequent design of allosteric $3C^{pro}$ inhibitors but also for a validation of the generated structural model for allosteric binding to the CVB3 $3C^{pro}$.

Chapter 6

Discussion

Discussion

The aim of the thesis was the design of non-peptidic 3C protease inhibitors, which serve as basis for the development of antiviral agents as treatment for enterovirus infections. This was pursued through a fragment-based approach in collaboration with the group of Prof. Rademann and close integration of computational methods to drive the individual steps (i) fragment screening, (ii) fragment hit exploration and (iii) fragment hit optimization.

In the following sections, the results of each design step will be discussed (Figure 6.1). As first step, a structural model for covalent $3C^{pro}$ binding was generated (section 4.1, 4.2) and subsequently used for fragment screening (section 4.3). In the second part, scaffold hopping for the structural exploration of the initial fragment hit was performed, which led to the identification of covalently binding fragments (section 4.4) as well as a fragment with an allosteric binding mode (section 5.1 and 5.2). In the final step, one of the fragment hits was optimized. For this purpose, a *de novo* design workflow was developed (section 4.5) and used to sample putative fragment analogues, which were synthesized by the collaborating group and biologically evaluated (section 4.6).

6.1 Fragment screening

6.1.1 Covalent $3C^{pro}$ binding model

At first, a structural model for the $3C^{pro}$ was developed to enable the rational design of covalent inhibitors. For this purpose, 3C proteases of the different human pathogenic enteroviruses were analyzed. The identified similarity confirmed the pursued design of pan-enterovirus inhibitors by addressing this target.

From crystal structures of peptidic ligands, a 3D pharmacophore model was developed, which comprises the key interactions for the binding to the CVB3 $3C^{pro}$. Through incorporation of the newly developed covalent bond feature, the covalent interaction was successfully reflected [169]. The model was subsequently validated by using non-peptidic ligands. This was mandatory to ensure that the model, which was derived from co-crystallized peptidic ligands, is able to retrieve bioactive, non-peptidic $3C^{pro}$ inhibitors. However, the only reported non-peptidic $3C^{pro}$ inhibitors with a verified binding mode have only been tested against the HRV 14 $3C^{pro}$. The high sequence and structural similarity of the $3C^{pro}$ prior elu-

cidated, especially in the active site, suggests binding mode conservation across the 3C proteases of the different enteroviruses. Through validation, good model characteristics with a sensitivity of 0.87 and a very high specificity of 0.99 were achieved. These also result from the specific arrangement of features, especially the covalent bond feature in close proximity to the hydrogen bond donor/acceptor features, which represent the interactions within the crucial S1 pocket.

The validation confirmed that the developed model comprises the features essential for binding to the $3C^{pro}$ active site and thus its applicability for the aspired virtual screening. The model was further enhanced by the conducted docking study for the covalent $3C^{pro}$ inhibitors developed by Daniel Becker via fragment ligations and the subsequently performed extensive analysis of covalent crystal structure, which revealed the relevance and utility of the initial complex for the rational design of covalently binding ligands. The analysis of the assembled set of covalent complexes showed that non-covalent binding conformations for over 75 % of the ligands could be obtained, which are similar to the covalent complex. The conservation of the binding interactions during the different steps of covalent binding confirms the utility of interaction patterns, such as 3D pharmacophores, derived from covalent crystal structures for virtual screening and covalent docking. Through the incorporation of the initial complex into the docking study for the dynamic ligation-based $3C^{pro}$ inhibitors, the structure of the putative ligation product and its binding mode could be elucidated. The crucial hydrogen bonds in the S1 pocket can only be formed in the non-covalent as well as the covalent bound state for the secondary alcohol. This ligation product was experimentally confirmed by Daniel Becker through the synthesis of the desoxy-analogue.

The value of incorporating the initial complex as descriptor for covalent binding was therefore used for the rational design of covalently binding ligands in this thesis. As stated earlier, the experimental verification of the true initial complex is sparsely done and extremely challenging. Therefore, the use of computational methods is of great value and often the only source of information about the initial complex.

6.1.2 Initial fragment screening

The developed model was subsequently used to screen the in-house small molecule collection, which was pruned prior to the VS to ensure fragment-like characteristics and potential covalent binding through the incorporation of a warhead. From the remaining 397 fragments, 12 were selected due to their ability to bind to the

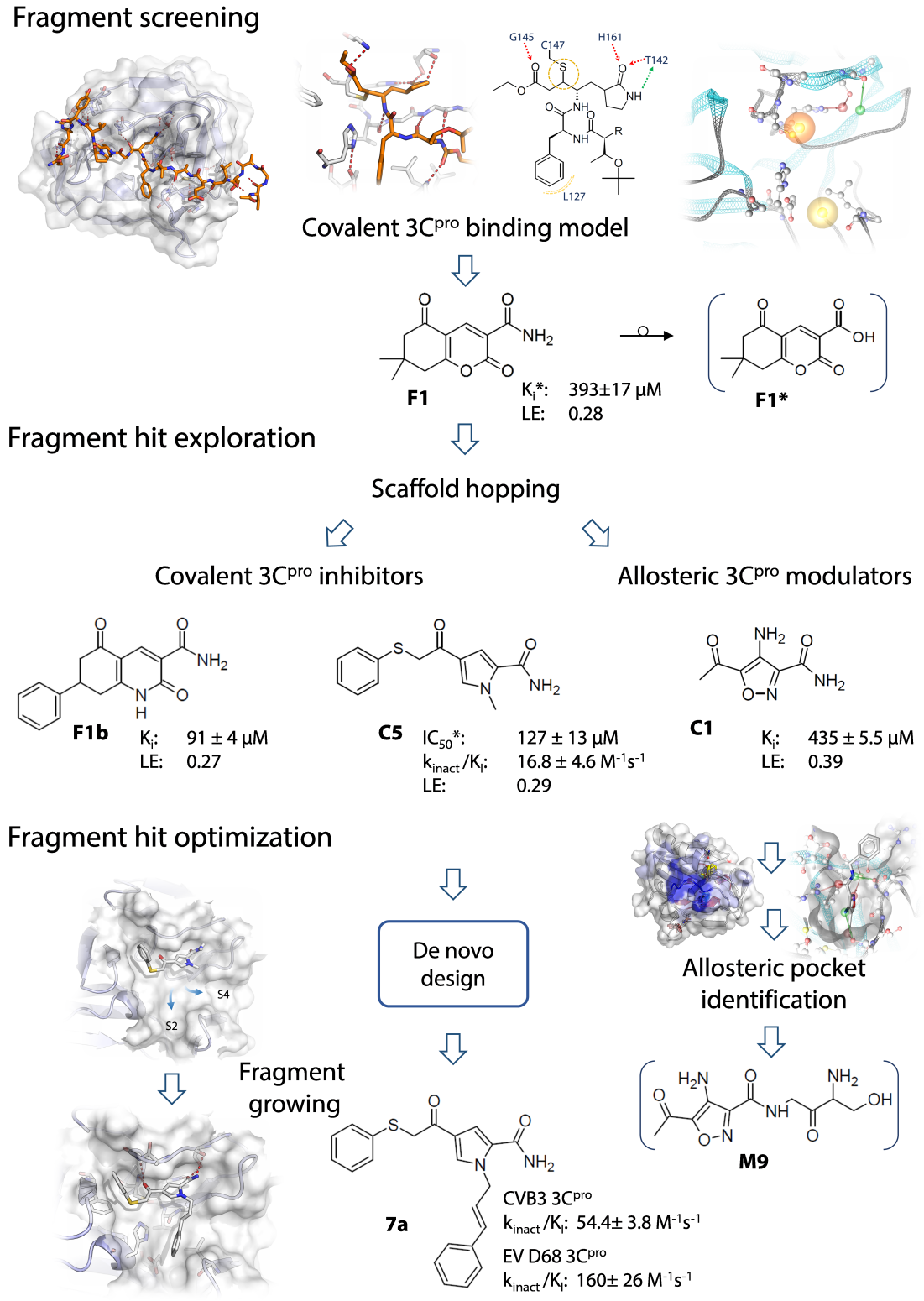

Figure 6.1: Thesis results overview. Key 3C^{pro} inhibitors identified in this thesis are highlighted and given with their appropriate activity value. Hypothetical or virtual ligand structures are given in blue brackets.

$3C^{pro}$ active site by forming the necessary hydrogen bonds in the S1 pocket in the covalently as well as in the non-covalently bound state.

In the biological evaluation six fragments showed apparent activity, but only the fragment **F1** could be verified with reasonable activity in the mid three-digit micromolar range. Furthermore, the time-dependent mode of action, displayed by **F1**, indicates the desired covalent interaction with the $3C^{pro}$ through the formation of the reversible covalent thiohemiketal. This provided further evidence for the proposed binding mode to be in accordance to the developed model for covalent $3C^{pro}$ inhibition, which was further supported by the detected chemical instability of **F1**. The conversion into the instability reaction product **F1*** correlates with a decrease in activity. The MS data obtained for the conversion product was used to elucidate the occurring reaction. This is proposed to proceed via an opening of the hydrolytically labile lactone moiety of **F1** and subsequent recyclization through the elimination of ammonia, which leads to the formation of the corresponding acid **F1***. This product is in accordance to the detected ions in the high-resolution mass spectra and the reported synthesis scheme for **F1** [221]. Furthermore, it confirms the proposed binding mode as the exchange of the primary amide group to an acid reflects the reported specificity of the S1 pocket and thereby the diminished activity of **F1*** [82, 99]. Further structural investigations of **F1*** that would confirm the proposed chemical structure, could not be undertaken due to the small amount of **F1** available in the molecule collection, which is not sufficient for NMR experiments.

From the 12 VS hits, one could be verified as $3C^{pro}$ inhibitor, which corresponds to a hit rate of 8 % and exemplifies the enrichment power of VS. However, for the set of 397 covalent fragments, which were included in the VS, this corresponds to hit rate of 0.25 % and only 0.03 % for the overall set of fragments, respectively. This is well below typical hit rates of fragment screening, which are in average around 2-3 % [236]. This indicates that potential actives might have been missed, especially by excluding all fragments devoid of a warhead from the VS. However, in fragment-based approaches only those hits are further pursued for which the binding mode could be elucidated via X-ray crystallography or NMR spectroscopy. Therefore, any putative active that has not been selected for the biological evaluation would have little value due to the uncertain binding mode as it would not bind according to the developed model. Therefore, tremendous efforts would have to be undertaken to elucidate the binding mode, which is exemplified by analysis conducted for the allosterically binding fragment **C1** discussed below.

The small hit rate also results from the specific requirements for binding in the $3C^{pro}$ active site as illustrated by the high specificity of the 3D pharmacophore model but also from the little desolvation possibilities within the flat and solvent exposed- active site.

This exemplifies the challenge of identifying small molecule inhibitors for the $3C^{pro}$, which was successfully done with the fragment **F1**. This comprises the essential features for $3C^{pro}$ inhibition, a ketone warhead and a primary amide for S1 binding, within a novel scaffold, namely a tetrahydrobenzopyran-2,5-dione. The incorporation of the aforementioned elements in a bicyclic system within the spatial distance of five single bonds was identified as essential characteristic for activity. This was used for the subsequent structural exploration of **F1** as it was not applicable for optimization due to its chemical instability.

6.2 Fragment hit exploration

6.2.1 Scaffold hopping

Substructure-based search in commercial fragment libraries led to the identification of two promising scaffolds comprising three series of bioactive fragments.

Fragments of the first series share a tetrahydroquinoline-2,5-dione scaffold, which is highly similar to the tetrahydrobenzopyran-2,5-dione of **F1**. Therefore, we surmised similar biological activity but in addition chemically stability due to the exchange of the lactone to a lactam moiety. Both could be verified, but the biological activity is significantly lower as fragment **F1a**, the aza-analogue of **F1**, shows only 25 % inhibition at 350 µM, which corresponds to the IC_{50}* of **F1**. A possible reason is a reduced reactivity of the ketone moiety due to a vinylogy with the incorporated nitrogen. Despite this, more potent tetrahydroquinoline-2,5-diones were identified. Especially substitution at the C-7 provides a boost in affinity through the occupation of the S1' pocket. The respective fragments **F1b** and **F1e** show K_i values around 100 µM.

All in all, the tetrahydroquinoline-2,5-dione scaffold provides the same attractive $3C^{pro}$ binding characteristics as rigid P1 glutamine mimetic as **F1**, but with the additional chemical stability and a putative synthetic handle for optimization through the incorporation of the lactam nitrogen. This would allow to grow the scaffold to adjacent pockets in the $3C^{pro}$ active site as has been done for the pyrrole-based fragments. Yet, the reduced activity of this scaffold diminished its

attractiveness, especially due to the limited room for optimization of the ketone-based warhead as it is incorporated within a fixed scaffold.

The second series of fragments comprise substituted pyrroles. These fragments were scarcely active in the upper micromolar range, which is probably due to the reduced rigidity as the ketone warhead is not part of the cyclic system in contrast to the bicyclic fragments. Only the pyrrole **C5** comprising a thiophenolmethyl ketone moiety showed activity in the low three-digit micromolar range with a time-dependent, irreversible mode of action. The apparent activity cliff indicating unspecific binding was elucidated through protein MS and pinpointed to result from the ketone α-substituent, which was shown to act as leaving group. This was verified through the formation of a disulfide adduct with the $3C^{pro}$ in addition to the thioether as primary covalent protein-ligand adduct. However, either covalent adduct could result from impurities since thiophenol as well as an α-halogenated methyl ketone have putatively been used as precursors for the synthesis of **C5**. But neither of them was found as impurity for **C5** and the concurrent occurrence of both protein-ligand adducts supports the proposed reaction mechanism.

Therefore, further investigations of the thiophenolmethyl ketone moiety were conducted as part of the optimization for fragment **C5**. Despite the reduced activity of the pyrrole scaffold in comparison to the bicyclic fragments, the potential incorporation of different warheads in contrast to the quinoline-based fragments and the attractiveness of the pyrrole nitrogen as synthetic handle for optimization through the incorporation of novel groups rendered **C5** attractive as starting point for fragment optimization. The irreversible binding characteristic, which enabled the subsequent verification of the specific binding mode in accordance to the developed structural model provided more evidence than the putative reversible covalent binding of the quinoline-based fragments.

6.2.2 Allosteric $3C^{pro}$ inhibitors

The third series of fragments comprises substituted 4-aminoisoxazoles. The initial hit fragment **C1** was identified as potential glutamine mimetic during the scaffold hopping. Despite the apparent similarity no binding to the active site could be determined, probably due to steric clashes resulting from the 4-amino substituent, which circumvents occupation of the essential S1 pocket. Instead, **C1** shows non-competitive $3C^{pro}$ inhibition with a K_i of 435 ± 5.5 µM and a LE of 0.39. This values render **C1** more attractive than the identified covalently binding fragments. How-

ever, due the initially missing binding mode and the questionable specificity of action associated with the high-concentration fragment screening, the covalently binding fragments were prioritized for optimization. Nonetheless, additional experiments and literature-derived information confirm the specificity of allosteric binding and provide structural information that enabled the generation of a model for allosteric $3C^{pro}$ binding.

C1 specificity was determined through its negative modulation of the $3C^{pro}$ labeling by a covalent, peptidic, fluorescent probe in a SDS-PAGE experiment. The noncompetitive kinetics of **C1** in the enzymatic assay indicate no competition with the substrate in the active site and thus no competition with the probe, which binds accordingly. Therefore, the detected effect for **C1** results either from an allosteric modulation or from assay interference such as the readout. This is fluorescence-based in both cases, but at different wavelengths. Secondly, the fragment is not present anymore in the SDS-PAGE as it is rapidly washed from the gel due to its small size. The inhibition of the $3C^{pro}$ activity in two different assay systems suggests a specific mode of action for **C1**. Yet, both assays work indirectly so that the detection of direct $3C^{pro}$ binding would provide further evidence.

Initial insight about the structural activity relationship for allosteric binding was obtained from the biological evaluation of **C1** analogues. Thereby, additional active 4-aminoisoxazole-based fragments were identified. But none provided an efficient improvement of binding affinity as LE decreases. This was later explained through the binding model as the incorporated substituents do not provide favorable contacts or interactions with the $3C^{pro}$. Further evidence for the specific allosteric $3C^{pro}$ modulation was drawn from the reported noncompetitive CVB3 $3C^{pro}$ inhibition by benserazide, which was successfully integrated into our own data for the generation of binding hypothesis.

Computational simulations for pocket finding indicate the RNA binding site to host an allosteric binding pocket, which is underlined by the biological relevance of this area in the enterovirus replication and the conservation of involved $3C^{pro}$ residues, especially the KFRDI motif. Identification of a transient pocket through MD simulations of the allosteric inhibitors provide further evidence for small molecule binding at this site. In addition to mutation- and NMR-based data, a structural model for RNA binding to the $3C^{pro}$ precursor 3CD has been reported. This was obtained from small angle X-ray scattering and shows binding of the RNA to the allosteric binding site but the inherent low resolution prevents any detailed structural information about the transient pocket [76, 231, 232].

The transient allosteric pocket enabled the development of binding hypothesis for both series of allosteric $3C^{pro}$ inhibitors through MD simulations, which reflect the experimentally determined structure activity relationships. Despite the structural dissimilarity of both scaffolds, a common binding mode was identified. The formation of additional favorable interactions of the serine substructure of benserazide was utilized to propose and virtually evaluate potential analogues of **C1** by merging its scaffold with the serine substructure. Several analogues can bind in the allosteric site by forming the interactions of both scaffold and thus serve as starting point for synthetic optimization. This is necessary to further explore and exploit the identified potential for allosteric $3C^{pro}$ modulation.

Further investigations are necessary despite the apparent consistency between the experimentally verified RNA binding site and the developed allosteric binding model. Especially the missing link between the allosteric binding of small molecules and its effect on the active site within the MD simulations requires further experimental and computational studies. This may result from insufficient sampling time in the MD simulations as the detection of conformational rearrangements, which transmit allosteric effects to the active site of the protein, require long simulations times. Furthermore, investigation on the allosteric effect on $3C^{pro}$ from other enteroviruses are necessary to elucidate if this is a common characteristic or specific for CVB3 $3C^{pro}$.

6.3 Fragment hit optimization

6.3.1 De novo design

To enable the rational optimization of the identified fragment hits, a *de novo* design workflow was developed. To ensure chemical rationality of the virtually generated molecules, synthesis tailored transformation rules were applied in combination with tangible building blocks. The utilization of readily available chemical reagents mimics common synthesis procedures and enables a direct and rapid transition from the virtually procedure to the synthesis bench. The description of chemical reactions by self-generated SMIRKS patterns allows for maximum control of the results and to take biological data into account without limitations. Despite containing only eight chemical transformations, the encoded set of reactions reflects a majority of the practically applied synthesis steps in drug design [226].

The use of a small set of chemical reactions and molecular BB limits the accessed chemical space. This was apparent in the virtual growing of **C5** toward the P3,4 area, for which no hit was obtained by using the Sigma Aldrich BB set. On the other hand, this prevents combinatorial explosion of the virtual molecule generation. Furthermore, the *de novo* design is based on biologically active fragments, which limits the necessity for exhaustive chemical space sampling. The essential aspect of the developed *de novo* design workflow is the attractiveness of the virtually generated ligands, which results from the synthetic feasibility and the biological relevance due to their origin from fragment hits. Therefore, a suitable starting point has to be found, which not only provides biological activity but also a synthetic accessible core structure, which can be used for the virtual fragment optimization.

The developed *de novo* design workflow was utilized at multiple instances in this thesis, which exemplifies its utility for the virtual generation or transformation of molecules.

6.3.2 Fragment optimization

Virtual growing of **C5** was performed according to the experimentally verified binding mode through the *de novo* design workflow by alkylation of the pyrrole nitrogen and subsequent VS for the identification of analogues able to occupy S2 or S4 pocket, respectively. In collaboration with the group of Prof. Rademann, synthesis of the S2 pocket addressing compounds was realized in accordance to the virtual design scheme, which confirms its chemical rationality.

Synthesized analogues of **C5** comprise a phenylpropene moiety for occupation of the S2 pocket, which provides a boost in activity. This is especially apparent for the ketone **5a**, which is active in the low three-digit micromolar range, whereas the increase in activity is less for the thiophenolmethyl ketone **7a** with activity in the two-digit micromolar range. This putatively results from the specific binding mechanism of the α-substituted methyl ketone, in which a rearrangement of the covalently protein bound atom occurs. This disturbs the favorable distance to the S1 occupying primary amide. Therefore, the conversion from the thiohemiketal, formed as primary covalent complex to the thioether product is less favorable than for the initial fragment hit **C5** and thus the gain in activity is limited. A similar effect has been observed for the activity of diazomethyl ketone-based peptides synthesized by Franziska Gottschalk in the group of Prof. Rademann, which were utilized to investigate the specificity of the allosteric fragment **C1** in this the-

sis (subsection 5.1). The diazomethyl ketone containing pentapeptide is active in a three-digit micromolar range against the CVB3 $3C^{pro}$, whereas a tetrapeptide with the same warhead designed for the caspase-3 is active in the nanomolar range. Structural analysis indicates that the transition from the initially formed thiohemiketal is also unfavorable for the $3C^{pro}$ but not for the caspase-3, which provides a rationale for the difference in biological activity.

Reactivity of the thiophenolmethyl ketone was investigated by analyzing its reaction with glutathione (GSH). This revealed that the warhead is stable against such non-catalytic nucleophiles. The bromomethyl ketone **6a**, obtained as intermediate in the synthesis of **7a**, was included in the analysis and served as positive control since it reacts with nucleophiles such as GSH through nucleophilic substitution and the resulting product could be detected. This would have been also formed by putative impurities of the α-thiophenolmethyl ketones, which were not detected. Mass spectrometry further confirmed the specific binding of the thiophenolmethyl ketones to the catalytic cysteine of the $3C^{pro}$ as only single adducts were found in contrast to **6a**. The MS data further support the reduced irreversible binding characteristics of **7a**. Furthermore, the α-thiophenolmethyl ketones included in the analysis, devoid of the necessary characteristics for $3C^{pro}$ active site binding, also display no protein binding. This implies that the covalent binding results from the proper spatial positioning of the warhead in the initial complex of the bioactive thiophenolmethyl ketones **C5** and **7a** comprising the essential chemical features, especially the primary amide moiety for S1 pocket binding. Therefore, the α-thiophenolmethyl ketone warhead can be classified as a quiescent affinity label. However, additional investigations for the reaction with other proteases are necessary to further characterize and evaluate the potential of the discovered warhead.

Despite the specific binding of the thiophenolmethyl ketone warhead, it is of limited use for targeting the $3C^{pro}$. Therefore, optimization of **7a** through the introduction of different warheads could provide better binding characteristics. However, the pursued incorporation of a α,β-unsaturated ester was not successful as the synthesized compound **13a** did not inhibit the $3C^{pro}$ in the expected activity range. Further optimization could not be realized in the time frame of the thesis but would be necessary to improve the potency of the identified and developed inhibitors. Potential optimization includes the warhead, for which the not synthesized ketone derivatives are promising. Furthermore, occupation of the P3,4 area in the $3C^{pro}$ active site could provide additional interactions.

6.4 Final remarks

Taken together, the aspired fragment-based approach was successfully pursued in an academic environment through the close integration of computational and biological methods. The developed structural models enable an enrichment for the selection of fragments for the biological evaluation and in combination with the experimentally verified covalent mode of action the generation of valid binding mode hypothesis. These provide the foundation for rational fragment hit optimization, for which the developed *de novo* design workflow was used to propose suitable but also synthetically feasible analogues.

The relevance of the developed structural models is exemplified by the continuously optimized biological activity of the identified $3C^{pro}$ inhibitors. This was subsequently increased from the initial fragment hit **F1** with a K_i* around 400 µM to the *de novo* designed **C5** analogue **7a** with activity in the two-digit micromolar range and a k_{inact}/K_I of $160 \pm 26\ M^{-1}s^{-1}$ for the EV D68 $3C^{pro}$. However, it is also clear that further chemical optimizations are necessary in order to achieve sufficient activity for use as anti-enteroantivirus agent.

Chapter 7

Conclusion & Outlook

Conclusion & Outlook

Enterovirus infectious have emerged as significant threat for the human health with no specific treatment available to date. The 3C protease is essential for the maturation of all enteroviruses by processing the viral polyprotein and cellular proteins, which suppresses host immune responses. Therefore, the $3C^{pro}$ has been pointed out as promising target for the development of anti-enterovirus agents. The primary aim of this thesis was the design of non-peptidic $3C^{pro}$ inhibitors.

This was pursued through a computer-driven fragment-based approach in collaboration with the group of Prof. Rademann. At first, a structural model for the covalent binding of small molecules to the $3C^{pro}$ active site was generated. This enabled the virtual screening-based identification of fragment hits and binding-mode rationalization through integration of the biological data. The initial fragment hit **F1** shows favorable $3C^{pro}$ binding characteristics but also limited chemical stability. Therefore, a scaffold hopping was performed through an analogue search in commercial fragment libraries, which led to the identification of novel, non-peptidic $3C^{pro}$ inhibitor scaffolds. For the rational fragment optimization, a *de novo* design workflow was developed, in which tangible building blocks and synthesis reaction rules are used to virtually generate molecules. This enabled the structure-guided sampling of the synthetically relevant fragment growing space of the irreversibly binding fragment hit **C5**. Chemical synthesis of a promising fragment growing analogue of **C5** was conducted by the collaborating group and furnished the micromolar inhibitor **7a**. Its structure serves as basis for further optimization through additional occupation of the $3C^{pro}$ active site for the development of potent enterovirus inhibitors.

Furthermore, the α-thiophenolmethyl ketone moiety of **C5** was identified by using protein mass spectrometry as novel type of warhead, which can covalently bind to catalytic cysteine residues through the elimination of thiophenol. Further investigations of the thiophenolmethyl ketone moiety are required to explore its reaction with a wider range of protein targets and nucleophiles in order to estimate its utility as covalent binding substructure.

The developed *de novo* design workflow represents a general approach for the rational fragment optimization through specific chemical transformation patterns and prepared building block collections. The combination of tangible reagents and synthesis tailored reaction rules with biologically verified fragment hits as starting points renders the *de novo* generated molecules synthetically feasible and

practically relevant.

In addition, biological evaluation of **C1** and its analogues led to the identification of a putative allosteric binding pocket for small molecules in the RNA-binding site of the CVB3 $3C^{pro}$ relevant for enterovirus genome replication. Through computational simulations, a binding mode hypothesis for small molecule binding in the allosteric pocket was generated and structural variations for optimization of the allosteric fragment hits were proposed. These can be used to optimize the biological activity of the identified fragments but also to verify the proposed structural model. In addition, further investigation for the possibility of allosteric modulation of the 3C protease from other enteroviruses have to be done to explore this opportunity as general strategy for targeting the enteroviral $3C^{pro}$. This is highly appealing as combinations of antiviral agents, addressing the same target, are favorable due to a delayed emergence of drug resistant virus strains.

In conclusion, the close integration of biological data and computational models enabled the identification and design of structurally novel inhibitors for the $3C^{pro}$ through a fragment-based approach. In addition to covalently binding inhibitors addressing the active site, fragments were identified, which modulate the $3C^{pro}$ through binding to an allosteric pocket. These findings provide not only the structural basis for the development of enterovirus agents but also enable a novel way for addressing the $3C^{pro}$.

Chapter 8

Experimental Procedures

Computational Protocols

Structure preparation

Protein structure preparation: Preparation of protein structures was performed in Molecular Operating Environment (MOE) version 2014.09. Crystal structures were obtained from the Protein Data Bank (PDB) [211] and prepared using the Structure Preparation Application in MOE. This involves protonation, assignment of partial charges, removal of crystal water, buffer components and additional proteins from the unit cell. The protein-ligand interactions were visually inspected and manually curated if necessary to set proper ionization or tautomeric states.

Ligand 3D conformation generation: CORINA 3.4 [237] was used for the generation of small molecule 3D conformations for molecular dynamics simulations or protein-ligand docking. The generation of multi-conformer ligand libraries for shape- or 3D pharmacophore-based virtual screening was performed using OMEGA v2.4.6. [238].

3D-Pharmacophore modeling

3C^{pro} 3D pharmacophore model generation: After preparation in MOE version 2014.09, the crystal structures 3ZZ6, 3ZZ7, 3ZZ8, 3ZZ9, 3ZZA, 3ZZB, 3ZZC and 3ZZD were load into the structure-based panel of LigandScout. For each structure a feature pattern representing the interactions of the protein-ligand complex was generated automatically and manually curated to ensure chemical reasonability. A shared feature pharmacophore was generated from the obtained feature patterns in the alignment panel of LigandScout.

Pharmacophore feature pattern of the covalent complex: The prepared crystal structures were loaded into the structure-based panel of LigandScout. Feature patterns representing the interactions in the protein-ligand complex were generated automatically and manually curated to ensure chemical reasonability. Feature representing interactions formed by the warhead were omitted as the conservation of the interactions during covalent bond formation depends on the nature of the respective warhead.

Dynophore generation: Dynophores were obtained from molecular dynamics simulation trajectories by using the in-house generated tool DynophoreApp [234]

Docking pose scoring: Docking poses were inserted into the crystal structure 3ZZA of the CVB3 $3C^{pro}$ in the structure-based panel of LigandScout. The pharmacophore fit S_{FCR} is calculated according to equation 3.5 and 3.6 based on the generated 3D pharmacophore model.

Relative feature fit calculation: The docking poses were inserted into the respective crystal structure in the structure-based panel of LigandScout. The feature pattern derived from the covalent complex is used to calculate the pharmacophore fit S_{FCR} according to equation 3.5 and 3.6. For comparison, the pharmacophore fit is transformed by using equation 8.1 to give the relative feature fit (Fit), which is a normalization over the feature pairs within the pharmacophore.

$$Fit = \frac{S_{FCR}}{c \cdot (N+1)} \tag{8.1}$$

Protein-ligand docking

Docking into the active site of the $3C^{pro}$: Protein-ligand docking was performed using GOLD v 5.0.2 by the Cambridge Crystallographic Data Center (Cambridge, UK, Europe). PDB entry 3ZZA for CVB3 and 3ZVG for EV D68 $3C^{pro}$ were used after preparation in MOE version 2014.09. 3D conformations of the ligands were generated with CORINA. The reacted state of covalently binding ligands was generated through the reactor application by applying the appropriate SMIRKS patterns in respect to the warhead of the ligand as described in section 4.5. If the stereochemistry was known or a new stereo center was introduced through the covalent binding, all possible stereoisomers were generated. The binding site was defined using a radius of 10 Å around the sulfur atom of C147. Docking was carried out using the PLP scoring function [239] with 25 GA runs per ligand in the most accurate mode without water molecules and default parameter settings. Covalent docking was performed by constraining the sulfur atom in the ligand and C147 to have the same position through the use of methanethiol as common substructure. Docking poses were analyzed and prioritized according the interactions represented in the 3D pharmacophore models.

Docking into the allosteric site of the $3C^{pro}$: PDB entries 3ZZA and 3ZYD were either directly used or taken from a representative frame of the MD simulations with the open transient allosteric pocket. The preparation was performed in MOE version 2014.09. 3D conformations of the ligands were generated with CORINA.

The binding-site was defined using a radius of 6 Å around a representative ligand manually placed in the allosteric pocket or derived from a MD simulation. The docking was carried out using the PLP scoring function with 25 GA runs per ligand in the most accurate mode without water molecules and default parameter settings. Docking poses were analyzed in LigandScout.

Docking study for comparison of initial and covalent complex: Crystal structures were taken from the PDB and prepared in MOE version 2014.09. Ligand structures were either derived from the crystal structure for the reacted state or built according the initial structure described in the respective publication. Water molecules were removed, except those involved in ligand building. The binding site was defined as a 4 Å radius around the co-crystallized ligand. For each ligand 50 diversity-based poses were generated with a cluster size of two and a cluster RMSD of 1.5 Å in triplicate. For covalent redocking the covalent bond was defined based on the atom and warhead type indicated by the crystal structure. The docking was carried out using the PLP scoring function in the most accurate mode with otherwise default parameter settings. Docking poses were analyzed through scoring with the respective pharmacophore feature pattern and calculation of the core RMSD between crystal structure and the docking pose and the relative feature fit.

Shape-based screening

Shaped-based virtual screening was performed: using ROCS 3.2.0.4. Query conformations were generated with CORINA. The screening was performed with default settings and the top 500 scored hits per query were retrieved and analyzed.

Molecular dynamics simulations

Molecular dynamics simulations were performed: using Desmond 3.2 [240] using the all-atom force field OPSL-AA. The protein structures of the CVB3 $3C^{pro}$ 3ZYD and 3ZZA taken from the PDB were used for the MD simulations. These were prepared in MOE version 2014.09. 3D conformations of the ligands were generated with CORINA. Both were loaded into MAESTRO (SCHRÖDINGER release 2014-2, version 9.8) to set up the MD simulation system. The protein was placed in an orthorhombic box filled with SPC water to ensure periodic boundary conditions. NaCl was added up to a concentration of 0.15 M to neutralize the

system. The NPT ensemble with Berendsen thermostat and barostat and otherwise default parameter settings were used for all simulations. Each simulation was initially minimized using the Desmond relaxation protocol for NPT conditions. This was followed by the actual simulation, which were run for either 25 ns or 100 ns on the Soroban computer cluster at the Freie Universität Berlin. The MD simulation trajectories were analyzed using the software VMD [241] and the dynophoreApp [234].

SMARTS & SMIRKS pattern generation

SMARTS: SMARTS patterns for substructure searches were manually generated in the text editor gedit of GNOME v. 3.8.4. The patterns were validated using Daylight web page test application [196]. SMARTS pattern-based substructure searches were performed using workflows generated in KNIME 2.11.0 (KNIME GmbH, Konstanz, Germany) as described below.

SMIRKS: SMIRKS patterns for chemical transformation were manually created in the text editor gedit of GNOME v. 3.8.4. The patterns were validated using Daylight web page test application [196]. SMIRKS patterns based chemical transformations were performed using in-house reactor application. Preparation of input structures and processing of product molecules were done using workflows generated in KNIME 2.11.0 (KNIME GmbH, Konstanz, Germany) as described below.

KNIME workflow generation

KNIME workflows: KNIME 2.11.0 (KNIME GmbH, Konstanz, Germany) was used in addition with nodes from the following community contributions [242]: CDK [243], Erl Wood Cheminformatics [244], Indigo [245] and RDKit [246].

Experimental Protocols

General information about used reagents

Reagents sources: Synthesis chemicals were purchased from ABCR GmbH & Co. KG (Karlsruhe, Germany), Carl Roth GmbH & Co. KG (Karlsruhe, Germany), Enamine (Ukraine), Fluka Chemie GmbH (Buchs, Switzerland), Merck KGaA (Darmstadt, Germany) and Sigma Aldrich (Steinheim, Germany) with the highest purity and used without further purification. Solvents were purchased in p.a. quality from Carl Roth GmbH & Co. KG (Karlsruhe, Germany), Sigma Aldrich (Steinheim, Germany) and VWR 71 International GmbH (Darmstadt, Germany). A Milli-Q station from Millipore (Eschborn, Germany) was used as source for purified water.

Vendor compounds: Commercially available screening compounds were purchased from Enamine (Ukraine) and Vitas-M Ltd. (Hong Kong). Purity (> 95) % and identity of the compounds was determined by using an Agilent (Santa Clara, USA) 1100 LC/MS system equipped with a Luna-C18 RP-column (3 µm) by Phenomenex (Torrance, USA). Samples were preapred by mixing 1 µL of the inhibitor DMSO stock solution with 99 µL of a $1:1$ mixture of water and acetonitrile. As solvents water and acetonitrile were used containing each 0.1 % (v/v) of formic acid. The ratio of water to acetonitrile was changed from $95:5$ to $1:99$ during the chromatographic run with a constant flowrate of 1 ml/min.

In-house small molecule collection: The small molecule collection, that served as a source of fragments in this thesis, was assembled using a maximum common substructure concept based on bioactive molecules listed in the World Drug Index [247] as reported by Lisurek *et al.* [215]. The compounds were purchased from ChemDiv Inc. (San Diego, USA). Identity and purity checks of the fragments selected for biological evaluation were performed via LC/MS as described above.

Chemical synthesis: Identity and purity of synthesized compounds were determined by LC/MS as described above, ^{1}H- and ^{13}C-NMR spectroscopy, which was performed with either a Joel ECP500 or a Bruker Avance III 700 using deuterated solvents obtained from Eurisotop (Saint-Aubin Cedex, France).

Biochemical assays

Enzyme kinetic assay: Biochemical evaluation for CVB3 and EV D68 $3C^{pro}$ inhibition was performed using the FRET substrate N-Dabcyl-KTLEALFQGPPVYE-(Edans)-NH_2 purchased from Biosynthan (Berlin, Germany). In 384 well plates (Corning) 10 µL buffer consisting of 100 mM HEPES-NaOH, 1 mM EDTA and 0.01 % TritonX (v/v) at a pH of 7.5 were mixed with 10 µL of $3C^{pro}$ (40 µM), 10 µL of substrate diluted in the assay buffer to a concentration of 20 µM and either 1 µL of the inhibitor diluted in DMSO or only DMSO for the positive control. Inhibitors were either directly obtained from chemical vendors or through chemical synthesis as described above. The fluorescence signal was measured at a wavelength of 538 nm after irradiation at 355 nm with a Safire II Fluorimeter (Tecan) over 30 min. Each concentration was tested in triplicates and the obtained data was analyzed using GraphPad Prism. IC_{50} values were calculated by non-linear regression according to equation 1.16. Observed first-order rate constant k_{obs} values were calculated for reversible inhibition using equation 1.20. K_i values were obtained from the ratio k_{off}/k_{on} according to equation 1.22. For irreversible inhibition the equation 1.24 was used. The second order-rate constants k_{inact}/K_I were derived from the slopes of the linear part of the graph of k_{obs} as a function of the inhibitor concentration based on equation 1.27.

Enzyme inhibitor incubation assay: The incubation assay was performed by using the same reagents and quantities as described above. The protein solution was mixed with inhibitor or DMSO for the control in the assay buffer and covered with a lid to prevent evaporation. After 60, 40, 20 or zero minutes the enzymatic reaction was initiated by addition of the substrate and monitored for 60 min. The incubation was performed in triplicates for each time frame of incubation and the respective control. The obtained data was analyzed using GraphPad Prism. The relative enzyme activity for every five minute time step of the enzymatic reaction was calculated for every distinct time frame of incubation. The obtained relative enzyme activity was plotted as a function of the enzyme reaction time.

Reversibility of enzyme inhibition: The reversibility of enzyme inhibition was determined by using the following procedure: 1 µL of inhibitor dissolved in DMSO at a concentration which results in 85 % enzyme inhibition according to equation 1.16 was mixed with 9 µL of $3C^{pro}$ (400 µM) and 10 µL buffer. As control, 1 µL DMSO was used. The mixture was incubated for 30 min and subsequently di-

luted through addition of 460 µL buffer. 6 µL of the substrate solution were added to 25 µL aliqouts of the diluted mixture to initiate the enzymatic reaction, which was monitored for 30 min and evaluated as described above.

Inhibition modality determination: The inhibition modality of the inhibitors was determined by using the same assay described above. Three different inhibitor dilutions were used, which result in 80 %, 50 % and 20 % enzyme inhibition, respectively, according to equation 1.16. The enzymatic reaction was monitored for 30 min for each inhibitor dilution with the following four substrate stock solution concentrations 120 µM, 90 µM, 60 µM and 30 µM. The data was analyzed using GraphPad Prism through non-linear regression according to equation 1.9, 1.12 and 1.14. The graphical evaluation was done through Dixon- and Cornish-Bowden plots. K_i values for noncompetitive inhibitors were obtained from the intersection of the lines in the respective Dixon- and Cornish-Bowden plot.

Glutathione reaction assay: At 2.5 mM the inhibitor, dissolved in DMSO, was mixed with an equimolar solution of glutathione dissolved in the assay buffer to a volume of 50 µL. The mixture was incubated at room temperature for 2 h and 24 h, respectively. After each incubation period an aliquot was analyzed via LC/MS as described above. Controls for each inhibitor were prepared and treated analogous, except the omission of glutathione.

SDS-polyacrylamide gel electrophoresis: In 10 µL assay buffer, 10 µL of VCB3 $3C^{pro}$ stock were incubated for 60 min with 1 µL covalent probe (250 µM) and 1 µL **C1** (2.5 mM, 1.25 mM, 0.5 mM), DMSO or vice versa for the positive and negative control, respectively. The mixture was run on a 16 % polyacrylamide SDS gel and analyzed under UV-light (355nm) and subsequently stained with Coomassie Blue.

Mass spectrometry

CVB3 $3C^{pro}$ MS: For protein MS the same reagents were used as described above. For CVB3 $3C^{pro}$ MS 10 µL of the stock solution were mixed with 20 µL buffer and 1 µL of the inhibitor dissolved in DMSO, which was added at a concentration, which leads to 50 % reduction of $3C^{pro}$ activity after 30 min in the kinetic assay. The samples were incubated for 60 min and then subjected to LC/MS analysis using the following procedure: For chromatographic separation a Phenomenex Jupiter C-18 RP-column (5 µm) was used. As solvents water and acetonitrile were utilized

containing 0.1 % (v/v) of formic acid each. The ratio of water to acetonitrile was changed from $95:5$ within the first 35 min of the run to $1:99$, which was kept constant for 5 min and then turned back to the initial ratio within the last 4 minutes of the chromatographic run. The flow rate was set to 0.5 mL/min. For ionization the following parameters were used: Capillary voltage 3.3 kV, sampling cone 40 V, source offset 60 V, source temperature 100 °C, desolvation temperature 250 °C, desolvation gas flow 600 L/h. MaxEnt1 method of the MassLynx software (Waters Co., Milford, UK) was used for deconvolution of the obtained protein mass spectra.

High resolution & EVD68 3C^{pro} mass spectrometry: For high resolution masses and the mass spectrometry of the EV D68 3C^{pro} the following procedures were used. Small molecule samples were prepared as described above. For protein MS, 200 µL of the EV D68 3C^{pro} stock solution were mixed with 50 µL buffer and 10 µL of the inhibitor dissolved in DMSO, which was added at a concentration, which leads to 50 % reduction of 3C^{pro} activity after 30 min in the kinetic assay. The samples were incubated for 60 min and subsequently concentrated and desalted through ultrafiltration using Roti®Spin MINI-10 devices from Carl Roth GmbH & Co. KG. For sample purification 1 mL of purified water was used. The sample was concentrated to a volume of 20-30 µL. Sample measurements were performed using an Agilent HPLC/QTOF-MS instrument. The Infinity 1290 UHPLC equipped with a Phenomenex Jupiter C-18 RP-column (5 µm) was used for chromatographic separation. 10 µL of injected samples were eluted with a flow rate of 0.5 mL/min using a gradient of water and acetonitrile containing 0.1 % (v/v) of formic acid each, which was changed from $95:5$ to $1:99$ within 3 min of the chromatographic run. This was held constant for the next 6 min and then changed back within the next minute. The 6550 iFunnel QTOF was utilized for mass analysis with the following parameters of the electrospray ionization-source: Gas temperature 200 °C, gas flow 14 L/min, nebulizer gas at 35 psig, sheath gas temperature 350 °C, sheath gas flow 11 L/min, capillary voltage 5 kV, nozzle voltage 1 kV, fragmentor voltage 420 V. The analysis was run in the positive mode with the reference masses of 121.050873 m/z and 922.009798 m/z and in the negative mode with the reference mass of 112.985587 m/z and 1033.988109 m/z for ion correction. Data acquisition was in the rage of 100-1700 m/z with a scan rate of 1 s^{-1}. Data analysis was performed with MassHunter software by Agilent Technologies. High resolution masses for **F1** and **F1*** were determined using the "Find by formula algorithm" with $C_{12}H_{13}NO_4$ and $C_{12}H_{12}O_5$ allowing for $[M+H]^+$ and

$[M+Na]^+$ ions in the positive mode and $[M-H]^-$ in the negative mode within 10 ppm of the calculated m/z. Protein mass spectra deconvolution was performed using the Maximum entropy algorithm with a mass range of 5-30 kDa, a mass step of 1 Da, within a m/z range of 500-5,0000 a baseline factor of 8 and an isotope width of 25 Da.

Appendix A

Results Part I

A.1 3C^{pro} inhibitors for model validation

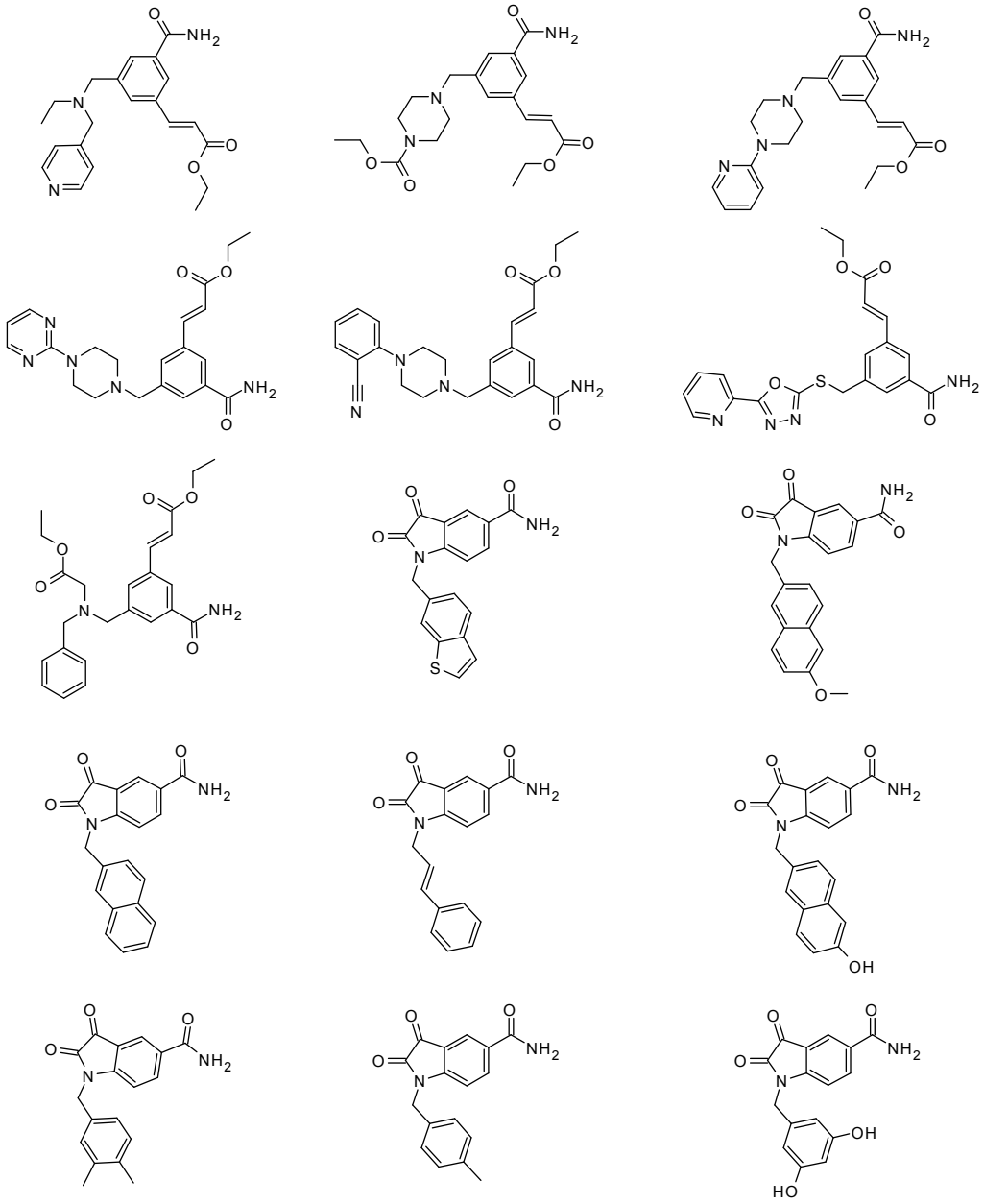

Figure A.1: 3C^{pro} inhibitors used for 3D pharmacophore model validation set 1.

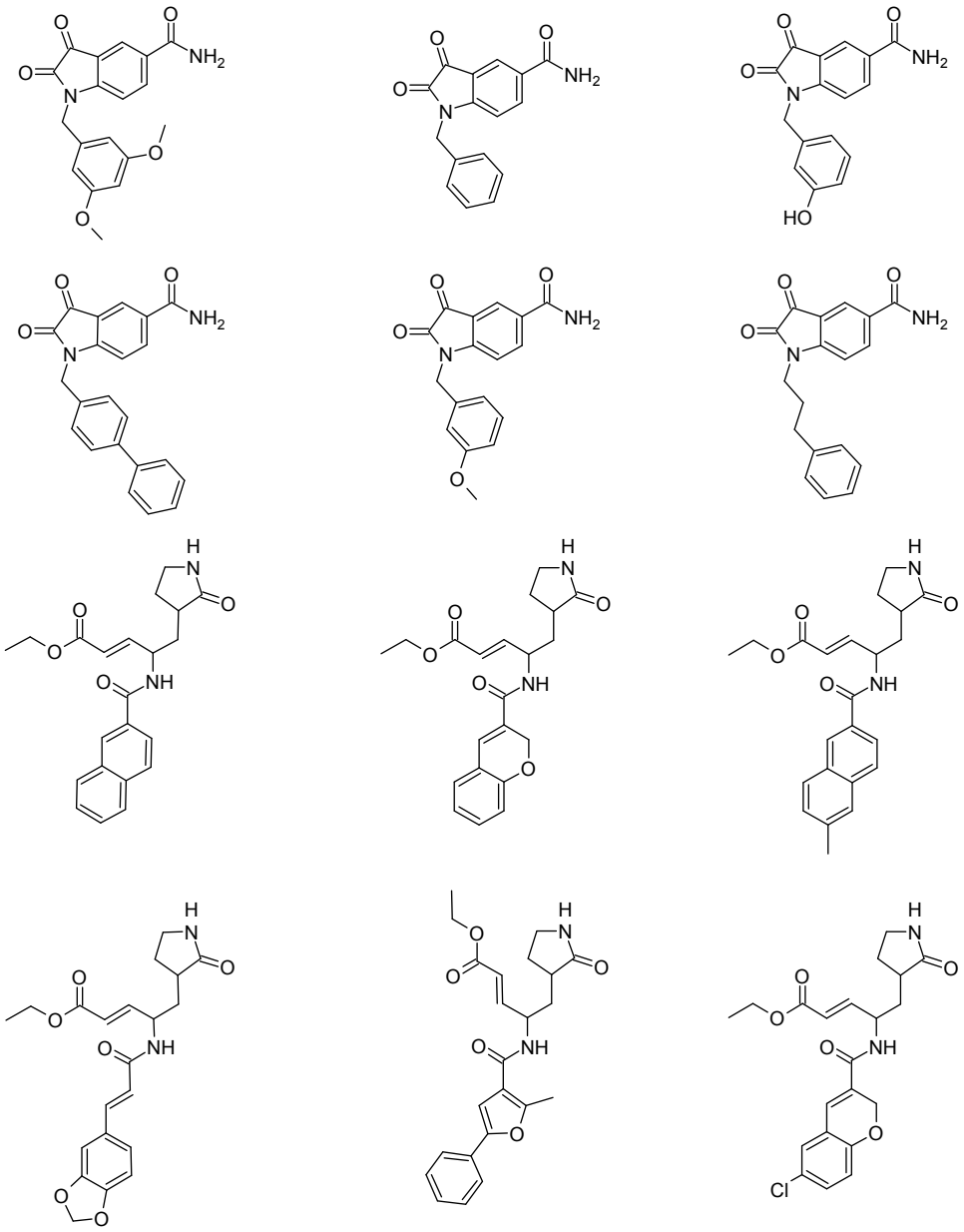

Figure A.2: 3C^{pro} inhibitors used for 3D pharmacophore model validation set 2.

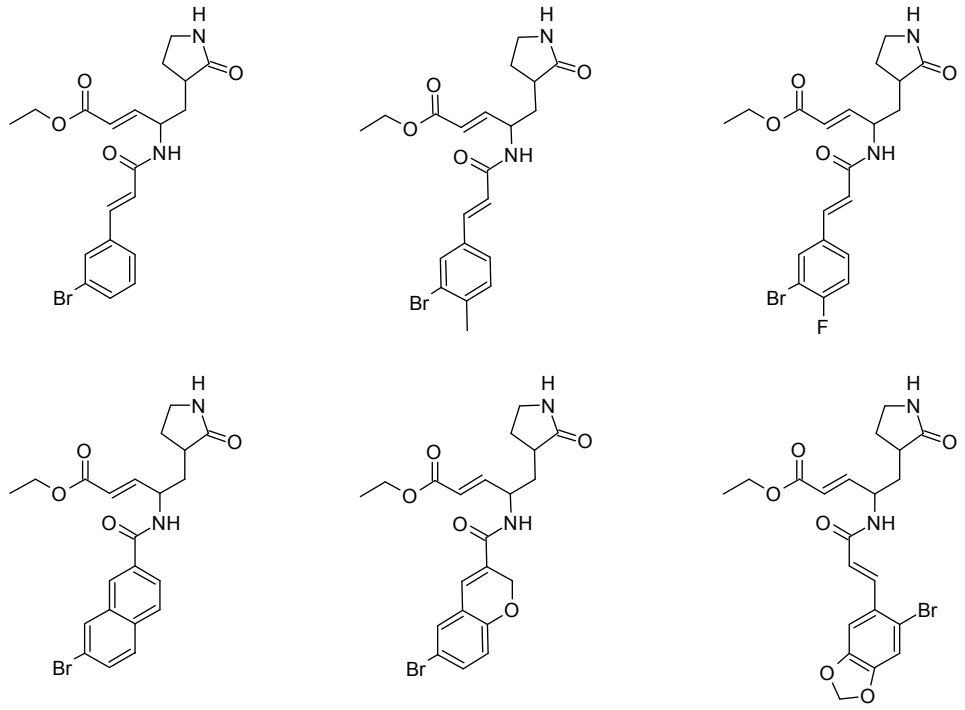

Figure A.3: 3C^{pro} inhibitors used for 3D pharmacophore model validation set 3.

A.2 Covalent ligands for initial complex generation

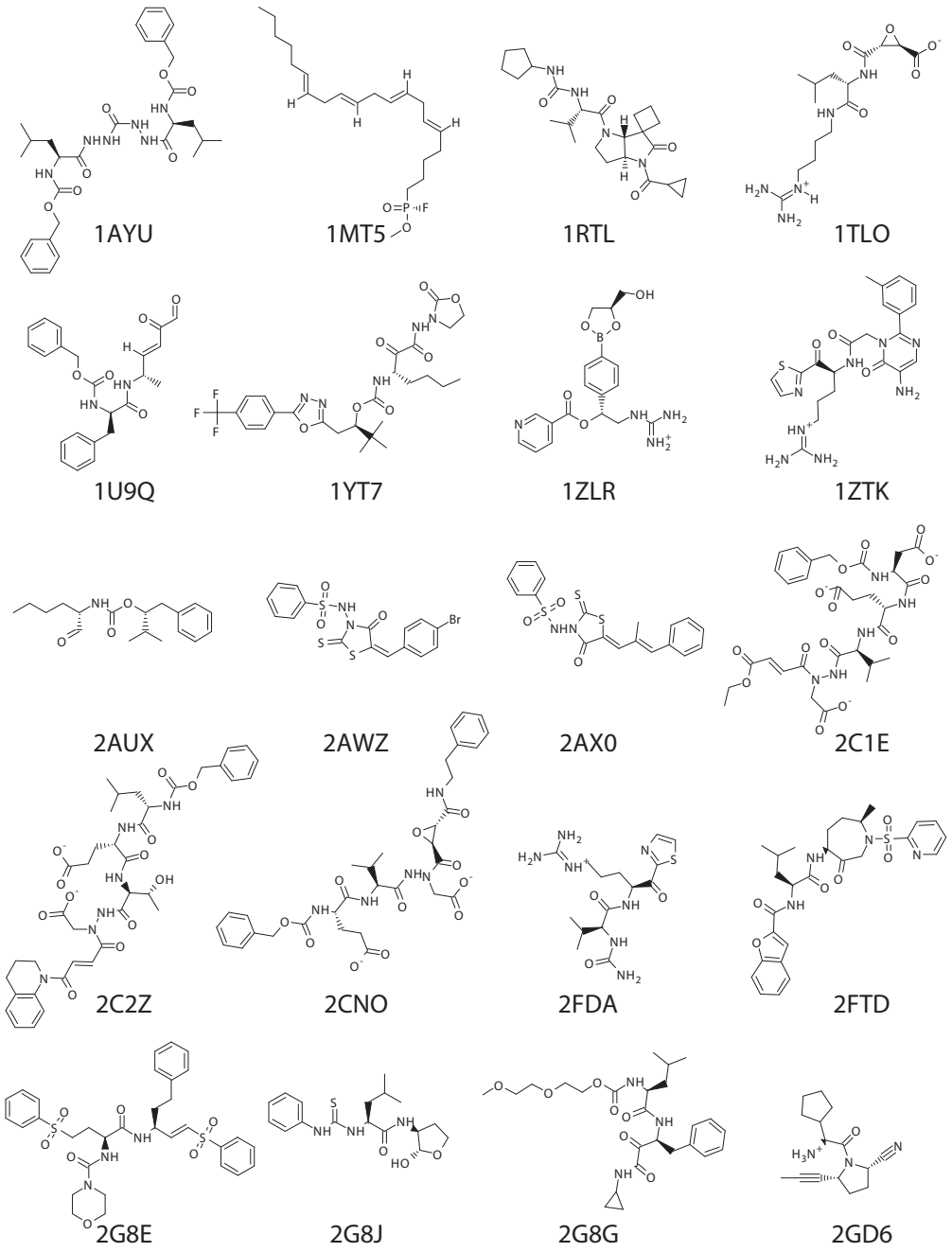

Figure A.4: Chemical structure and PDB ID for covalent ligands used for initial complex generation, set 1.

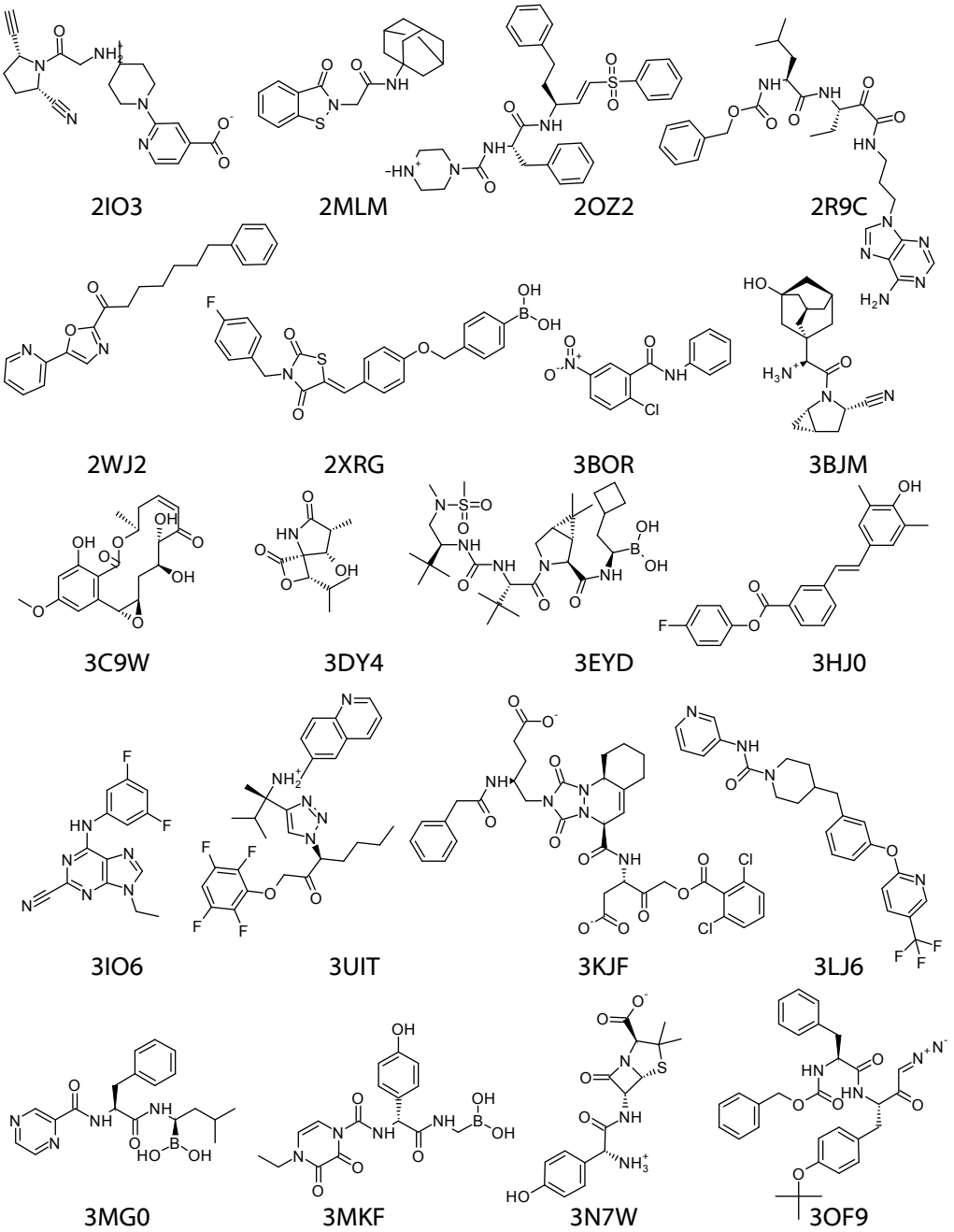

Figure A.5: Chemical structure and PDB ID for covalent ligands used for initial complex generation, set 2.

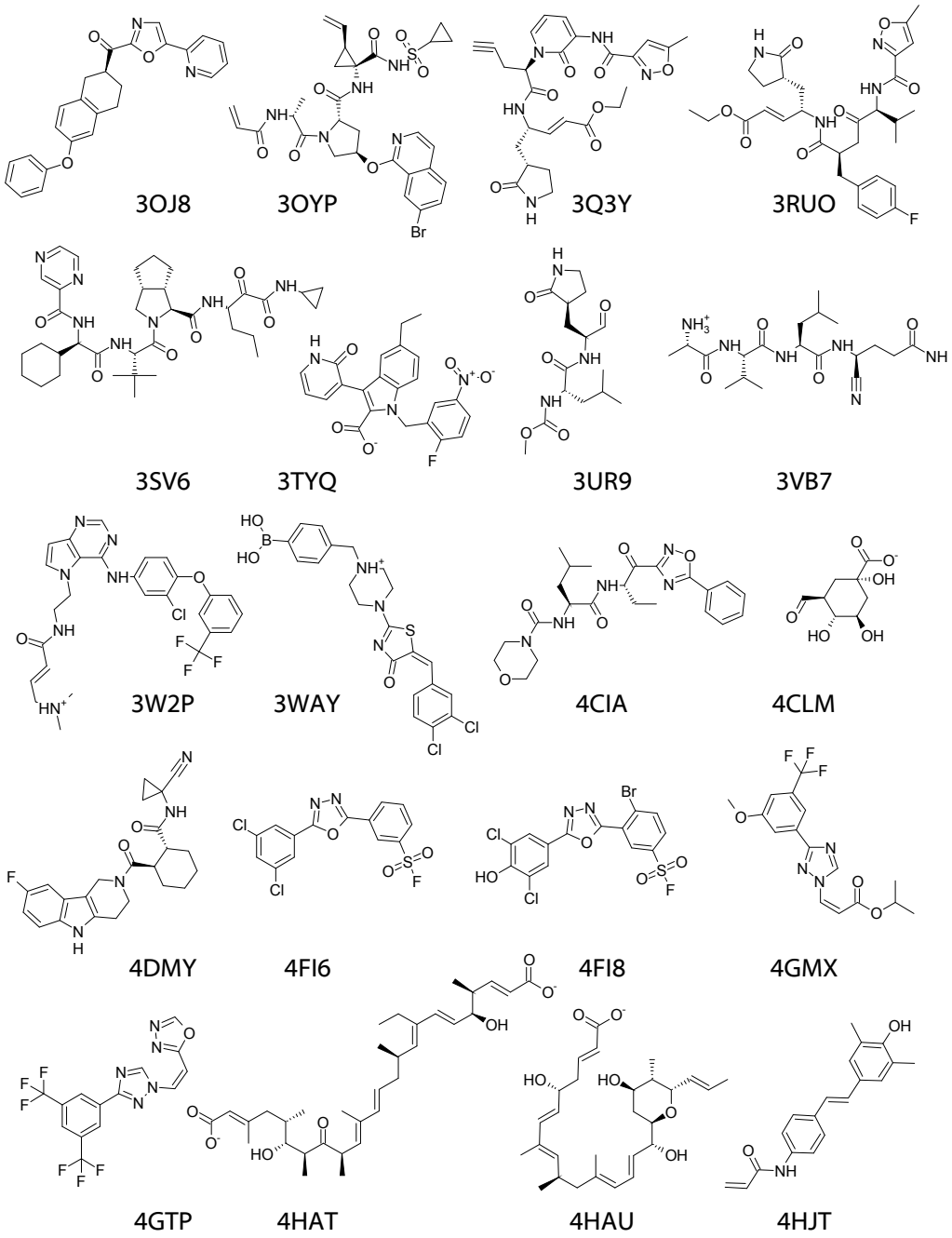

Figure A.6: Chemical structure and PDB ID for covalent ligands used for initial complex generation, set 3.

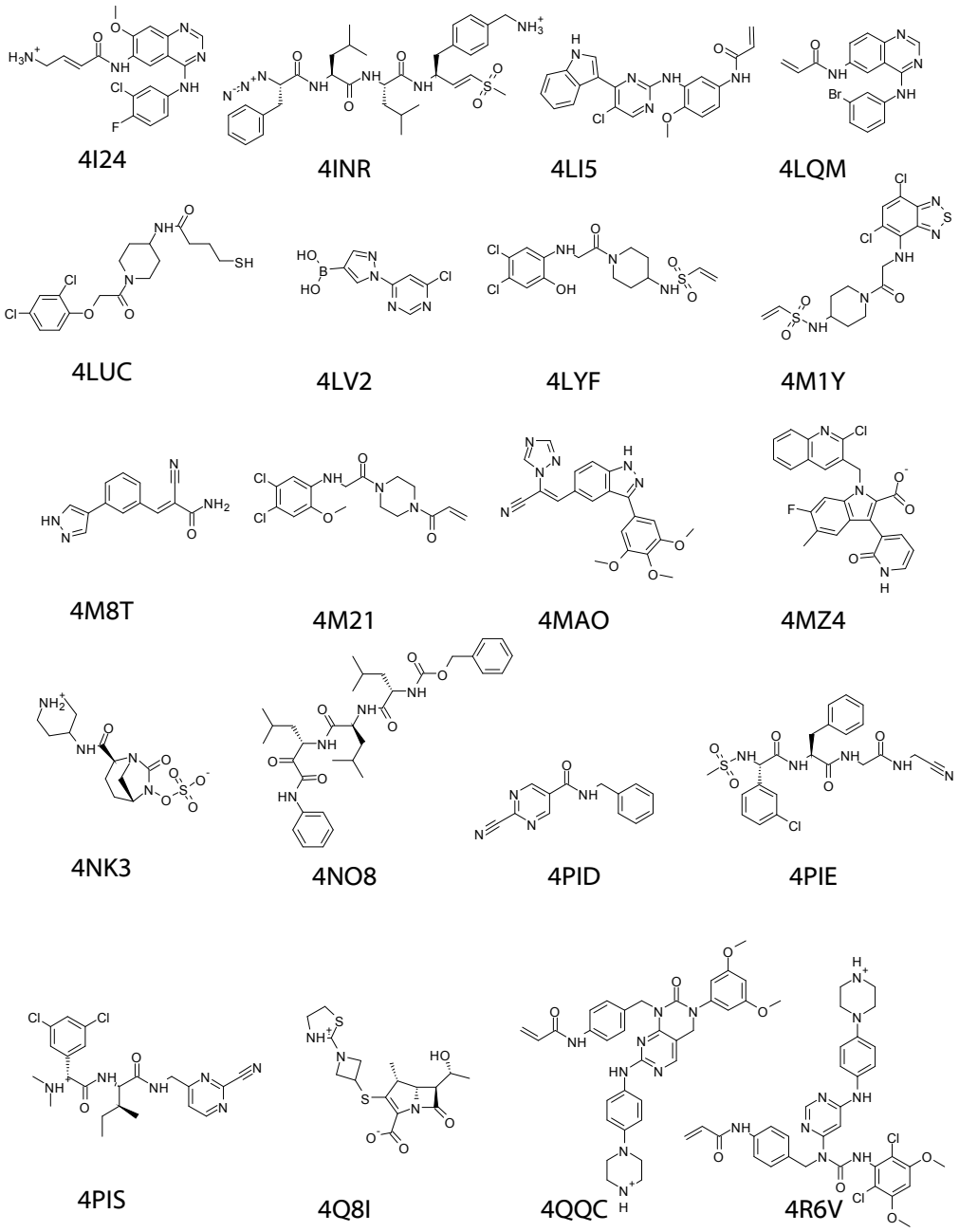

Figure A.7: Chemical structure and PDB ID for covalent ligands used for initial complex generation, set 4.

A.3 Putative dynamic fragment ligation products

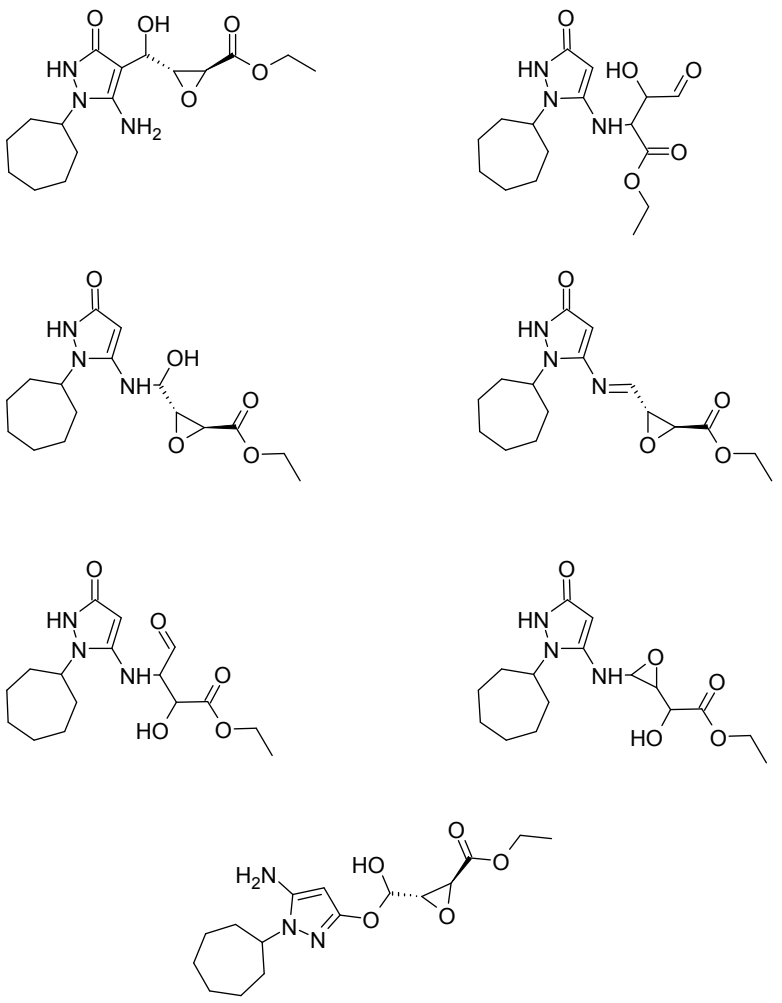

Figure A.8: Chemical structures of putative dynamic fragment ligation products utilized for binding mode analysis.

A.4 SMIRKS patterns

Reductive amination of aldehydes	[H][NX3;!$([NH0]);!$(N~[#7,#8,#15,#16]);!$(NC[#7,#8,#15,#16]);!$(NC=[#7,#8,#16]);!$(NC=CC(=O));!$(N#*);!$(N*#*);!$(N=*):1].[#6:3][CH1;!$(C-#7,#8,#16]);$(CC(=[O,C,N,S]));!$(C*[Cl,Br,I]);!$(CC(O)O):2](=O)>>[C:2]([#6:3])[N:1]
Reductive amination of ketones	[#6:4][C;$(C([#6])[#6]);!$(CC=[#7,#8,#16]);!$(CCC=[#7,#16]);!$(C(C=C)C=C);!$(CC=C[#7,#8]):3](=O).[H][NX3;!$([NH0]);!$(N~[#7,#8,#15,#16]);!$(NC[#7,#8,#15,#16]);!$(NC=[#7,#8,#16]);!$(NC=CC(=O));!$(N#*);!$(N*#*);!$(N=*):1]>>[C:3]([#6:4])[N:1]
Nucleophilic substitution	[H:6][C;$([CX4]);!$(C=O);!$(C([Cl,Br,I])[Cl,Br,I]):3]([H:5])[*;Br,Cl,I:4].[H][*;$([O,#7]);!$([H0]);!$(*=*);!$(*~[#5,#7,#8,#15,#16]);!$(*C[#7,#8,#15,#16]);!$(*C=[#7,#8,#16]);!$(NC=CC(=O));!$(N#*);!$(N*#*):1][#6:2]>>[C:2][*:1][C:3]([*:5])[*:6].[*:4]
Buchwald-Hartwig amination	[c:3][*;Cl,Br,I:4].[H][*;N,O;!$([H0]);!$(*C=CC(=O));!$(Nc);!$(N=[#6]);!$(*[#7,S,O,B]);!$(*C=[O,S,N]):1][#6:2]>>[#6:2][N:1][*:3].[*:4]
Chan-Lam coupling	[c:3]B(O)O.[H][N,O;!$([H0]);!$(OC);!$(Nc);!$(N=C);!$(*[N,SX2,SX3,#8,B]);!$(*C=N);!$([#7][#6](=O)[#7]);!$([#7][SX4][#7]):1]>>[*:1][*:3]
N-Acylation	[H][#7;!$([H0]);!$(N=[#6]);!$(*[O,N,S,B,P]);!$(*[#6]=[O,S,N]);!$(N(c)c):1][#6:2].[C:3](=O)[*;Br,Cl,I:4]>>[*:2][N:1][C:3](=O).[*:4]
N-Sulfonylation	[H][#7;!$([H0]);!$(N=[#6]);!$(*[O,N,S,B,P]);!$(*[#6]=[O,S,N]);!$(N(c)c):1][#6:2].[#6:4][S:3](=O)(=O)[Cl:5]>>[#6:2][N:1][S:3](=O)(=O)[*:4].[*:5]
Suzuki coupling	[*;#6X3:2][B]([O])[O].[*;#6X3;!$([#6]=O):1][*;Cl,Br,I:3]>>[*:1][:2].[*:3]
Sonogashira coupling	[H][C:4]#[#6:5][#6:6].[#6:7][N:1]=[N;$([+1]):2]=[N;$([-1]):3]>>[#6:7][#7:1]1[#7:2]=[#7:3][#6:5]([#6:6])=[#6:4]1

Figure A.9: SMIRKS patterns for chemical transformation.

α,β-unsaturated carbonyl compounds

[*;#1,#6:99][C;!$(C[#8,#7,#16]):1]([*;#1,#6:98])=[C:2]([*;#1,#6,#7,#8:97])-C(=O)-[*$([#1,#7,#6,#8]);!$([-1,OH1]):3]>>[*:99][C@;!$(C[#8,#7,#16]):1]([*:98])(S)-[C@:2]([*:97])([H])-C(=O)-[*$([#1,#7,#6,#8]);!$([-1,OH1]):3]
[*;#1,#6:99][C;!$(C[#8,#7,#16]):1]([*;#1,#6:98])=[C:2]([*;#1,#6,#7,#8:97])-C(=O)-[*$([#1,#7,#6,#8]);!$([-1,OH1]):3]>>[*:99][C@;!$(C[#8,#7,#16]):1]([*:98])(S)-[C@@:2]([*:97])([H])-C(=O)-[*$([#1,#7,#6,#8]);!$([-1,OH1]):3]
[*;#1,#6:99][C;!$(C[#8,#7,#16]):1]([*;#1,#6:98])=[C:2]([*;#1,#6,#7,#8:97])-C(=O)-[*$([#1,#7,#6,#8]);!$([-1,OH1]):3]>>[*:99][C@@;!$(C[#8,#7,#16]):1]([*:98])(S)-[C@:2]([*:97])([H])-C(=O)-[*$([#1,#7,#6,#8]);!$([-1,OH1]):3]
[*;#1,#6:99][C;!$(C[#8,#7,#16]):1]([*;#1,#6:98])=[C:2]([*;#1,#6,#7,#8:97])-C(=O)-[*$([#1,#7,#6,#8]);!$([-1,OH1]):3]>>[*:99][C@@;!$(C[#8,#7,#16]):1]([*:98])(S)-[C@@:2]([*:97])([H])-C(=O)-[*$([#1,#7,#6,#8]);!$([-1,OH1]):3]

vinyl sulfones

[*;#1,#6:99][C;!$(C[#8,#7,#16]):1]([*;#1,#6:98])=[C:2]([*;#1,#6,#7,#8:97])-C#N>>[*:99][C@;!$(C[#8,#7,#16]):1]([*:98])(S)-[C@:2]([*:97])([H])-C#N
[*;#1,#6:99][C;!$(C[#8,#7,#16]):1]([*;#1,#6:98])=[C:2]([*;#1,#6,#7,#8:97])-C#N>>[*:99][C@;!$(C[#8,#7,#16]):1]([*:98])(S)-[C@@:2]([*:97])([H])-C#N
[*;#1,#6:99][C;!$(C[#8,#7,#16]):1]([*;#1,#6:98])=[C:2]([*;#1,#6,#7,#8:97])-C#N>>[*:99][C@@;!$(C[#8,#7,#16]):1]([*:98])(S)-[C@:2]([*:97])([H])-C#N
[*;#1,#6:99][C;!$(C[#8,#7,#16]):1]([*;#1,#6:98])=[C:2]([*;#1,#6,#7,#8:97])-C#N>>[*:99][C@@;!$(C[#8,#7,#16]):1]([*:98])(S)-[C@@:2]([*:97])([H])-C#N

vinyl nitriles

[*;#1,#6:99][C;!$(C[#8,#7,#16]):1]([*;#1,#6:98])=[C:2]([*;#1,#6,#7,#8:97])-C#N>>[*:99][C@;!$(C[#8,#7,#16]):1]([*:98])(S)-[C@:2]([*:97])([H])-C#N
[*;#1,#6:99][C;!$(C[#8,#7,#16]):1]([*;#1,#6:98])=[C:2]([*;#1,#6,#7,#8:97])-C#N>>[*:99][C@;!$(C[#8,#7,#16]):1]([*:98])(S)-[C@@:2]([*:97])([H])-C#N
[*;#1,#6:99][C;!$(C[#8,#7,#16]):1]([*;#1,#6:98])=[C:2]([*;#1,#6,#7,#8:97])-C#N>>[*:99][C@@;!$(C[#8,#7,#16]):1]([*:98])(S)-[C@:2]([*:97])([H])-C#N
[*;#1,#6:99][C;!$(C[#8,#7,#16]):1]([*;#1,#6:98])=[C:2]([*;#1,#6,#7,#8:97])-C#N>>[*:99][C@@;!$(C[#8,#7,#16]):1]([*:98])(S)-[C@@:2]([*:97])([H])-C#N

propargyl carbonyl compounds

[*;#1,#6:99][C;!$(C[#8,#7,#16]):1]#[C:2]-C(=O)-[*$([#1,#7,#6,#8]);!$([-1,OH1]):3]>>[*:99]/[C;!$(C[#8,#7,#16]):1](/S)=[C:2]([H])-C(=O)-[*$([#1,#7,#6,#8]);!$([-1,OH1]):3]
[*;#1,#6:99][C;!$(C[#8,#7,#16]):1]#[C:2]-C(=O)-[*$([#1,#7,#6,#8]);!$([-1,OH1]):3]>>[*:99]/[C;!$(C[#8,#7,#16]):1](\S)=[C:2]([H])-C(=O)-[*$([#1,#7,#6,#8]);!$([-1,OH1]):3]

nitriles

[C;!$(CC=C):1]#[N:2]>>S[C;!$(CC=C):1]=[N:2]

aldehydes

[C:1](=O)(-[#1:2])-[#6:3]>>[C@:1](S)(O)(-[#1:2])-[#6:3]
[C:1](=O)(-[#1:2])-[#6:3]>>[C@@:1](S)(O)(-[#1:2])-[#6:3]

ketones

[c;$(*~[n,o,s]),$(*~c~[n,o,s]),$(*~c~c~[n,o,s]):1]-[C:2](=O)[*;#6:3]>>[*:1][C@:2](O)(S)[*:3]
[c;$(*~[n,o,s]),$(*~c~[n,o,s]),$(*~c~c~[n,o,s]):1]-[C:2](=O)[*;#6:3]>>[*:1][C@@:2](O)(S)[*:3]

dicarbonyl compounds

[#6:1][C:2](=O)C(=O)[*;H,N,O:3]>>[#6:1][C@:2](S)(O)C(=O)[*;H,N,O:3]
[#6:1][C:2](=O)C(=O)[*;H,N,O:3]>>[#6:1][C@@:2](S)(O)C(=O)[*;H,N,O:3]

trifluoromethyl ketones

[C:2](F)(F)(F)[C:1](=O)-[#6:3]>>[C@:1](S)(O)(-[*:2])-[#6:3]
[C:2](F)(F)(F)[C:1](=O)-[#6:3]>>[C@@:1](S)(O)(-[*:2])-[#6:3]

semicarbazones

[C:3]=[N:4]-[N:5]-[C:1](*:99)(=[*;O,S:2])>>[C:3]=[N:4]-[N:5]-[C@:1](*:99)(S)[*;O,S:2] semicarb1
[C:3]=[N:4]-[N:5]-[C:1](*:99)(=[*;O,S:2])>>[C:3]=[N:4]-[N:5]-[C@@:1](*:99)(S)[*;O,S:2] semicarb1

Figure A.10: SMIRKS patterns for warhead preparation for covalent docking.

Appendix B

Results Part II

B.1 RMSD Plots of molecular dynamics simulations

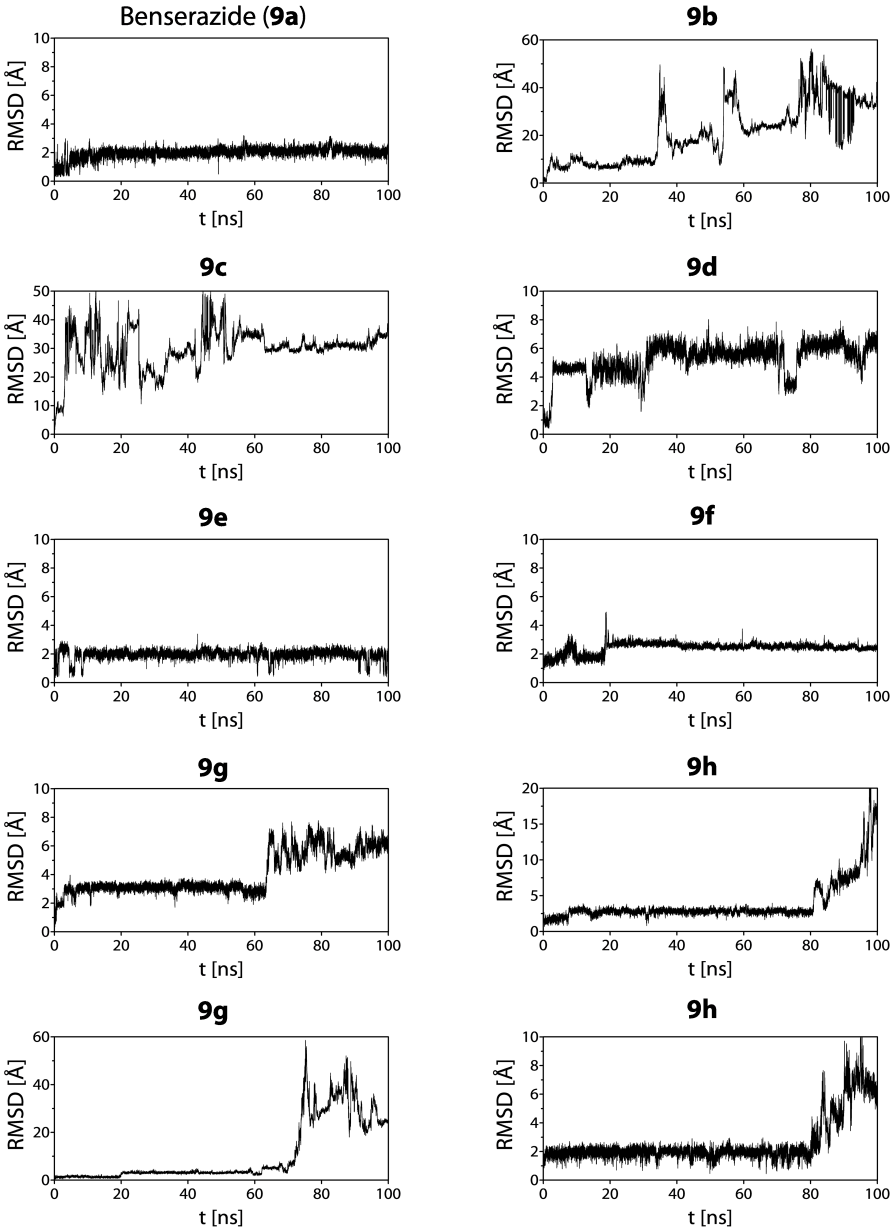

Figure B.1: RMSD plots of the MD simulation of benserazide and its analogues in the allosteric pocket of the $3C^{pro}$.

B.2 Fragment merging ligands

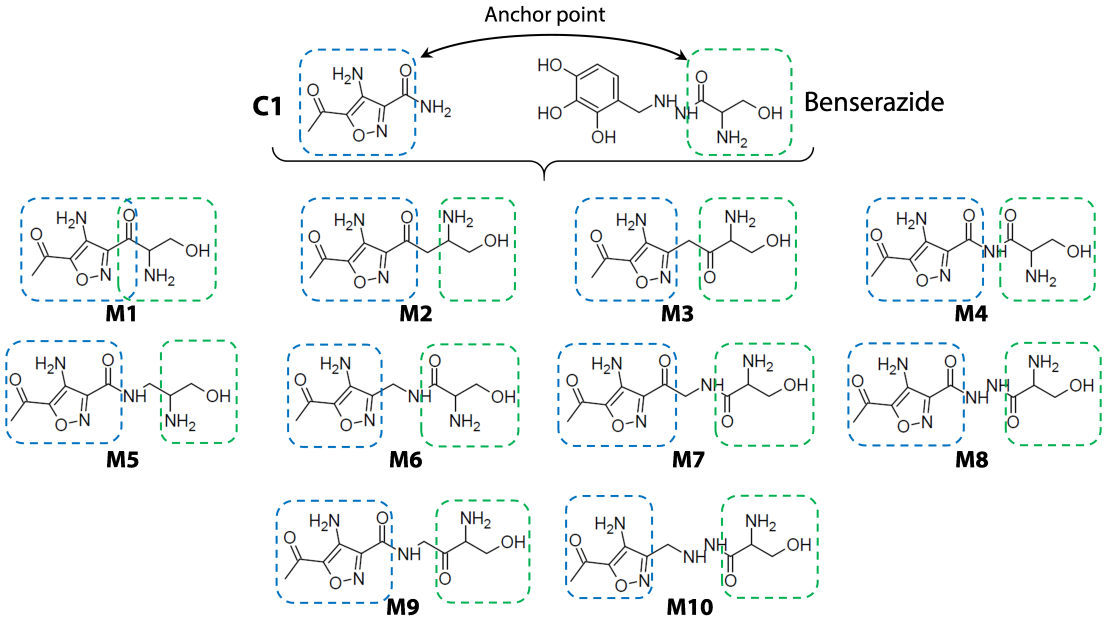

Figure B.2: Structures proposed for the fragment merging.

Bibliography

1. Hilgenfeld, R. & Peiris, M. From SARS to MERS: 10 years of research on highly pathogenic human coronaviruses. *Antivir Res* **100,** 286–295 (2013).
2. Burk, R. *et al.* Neglected filoviruses. *FEMS Microbiol Rev* **40,** 494–519 (2016).
3. Saiz, J. C. *et al.* Zika virus: the latest newcomer. *Front Microbiol* **7** (2016).
4. Chancey, C., Grinev, A., Volkova, E. & Rios, M. The global ecology and epidemiology of West Nile virus. *Biomed Research International* (2015).
5. Holm-Hansen, C. C., Midgley, S. E. & Fischer, T. K. Global emergence of enterovirus D68: a systematic review. *Lancet Infect Dis* **16,** e64–e75 (2016).
6. Bian, L. L. *et al.* Coxsackievirus A6: A new emerging pathogen causing hand, foot and mouth disease outbreaks worldwide. *Expert Rev Anti-Infe* **13,** 1061–1071 (2015).
7. Moreira, D. & Lopez-Garcia, P. Ten reasons to exclude viruses from the tree of life. *Nat Rev Microbiol* **7,** 306–311 (2009).
8. Baltimore, D. Expression of animal virus genomes. *Bacteriol Rev* **35,** 235–241 (1971).
9. Brennan, K. & Bowie, A. G. Activation of host pattern recognition receptors by viruses. *Curr Opin Microbiol* **13,** 503–507 (2010).
10. Biebricher, C. K. & Eigen, M. What is a quasispecies? *Curr Top Microbiol* **299,** 1–31 (2006).
11. Karesh, W. B. *et al.* Ecology of zoonoses: natural and unnatural histories. *Lancet* **380,** 1936–1945 (2012).
12. De Clercq, E. & Li, G. Approved antiviral drugs over the past 50 years. *Clin Microbiol Rev* **29,** 695–747 (2016).
13. Lieberman-Blum, S. S., Fung, H. B. & Bandres, J. C. Maraviroc: A CCR5-receptor antagonist for the treatment of HIV-1 infection. *Clin Ther* **30,** 1228–1250 (2008).

14. Wharton, S. A., Belshe, R. B., Skehel, J. J. & Hay, A. J. Role of virion M2 protein in influenza-virus uncoating - Specific reduction in the rate of membrane-fusion between virus and liposomes by amantadine. *J Gen Virol* **75,** 945–948 (1994).

15. Koch, H. Aciclovir - Antiviral drug for therapy of herpes infections. *Pharm Int* **5,** 191–192 (1984).

16. Frampton, J. E. & Perry, C. M. Emtricitabine - A review of its use in the management of HIV infection. *Drugs* **65,** 1427–1448 (2005).

17. De Clercq, E. Chemotherapy of the acquired immune deficiency syndrome (AIDS): Non-nucleoside inhibitors of the human immunodeficiency virus type 1 reverse transcriptase. *Int J Immunopharmacol* **13 Suppl 1,** 83–89 (1991).

18. Summa, V. *et al.* Discovery of raltegravir, a potent, selective orally bioavailable HIV-integrase inhibitor for the treatment of HIV-AIDS infection. *J Med Chem* **51,** 5843–5855 (2008).

19. Robertson, J. & Feinberg, J. Darunavir: a nonpeptidic protease inhibitor for antiretroviral-naive and treatment-experienced adults with HIV infection. *Expert Opin Pharmaco* **13,** 1363–1375 (2012).

20. Stedman, C. Sofosbuvir, a NS5B polymerase inhibitor in the treatment of hepatitis C: a review of its clinical potential. *Ther Adv Gastroenter* **7,** 131–140 (2014).

21. Link, J. O. *et al.* Discovery of ledipasvir (GS-5885): A potent, once-daily oral NS5A inhibitor for the treatment of hepatitis C virus infection. *J Med Chem* **57,** 2033–2046 (2014).

22. Langenbacher-Niessing, D. Additional benefits for 90 % of all genotype-1-patients with hepatitis C: IQWiG recognizes positive "dramatic effects" for patients during therapy with Harvoni®. *Viszeralmedizin* **31,** 144 (2015).

23. Hayden, F. G. *et al.* Use of the selective oral neuraminidase inhibitor oseltamivir to prevent influenza. *New Engl J Med* **341,** 1336–1343 (1999).

24. Paeshuyse, J., Dallmeier, K. & Neyts, J. Ribavirin for the treatment of chronic hepatitis C virus infection: A review of the proposed mechanisms of action. *Curr Opin Virol* **1,** 590–598 (2011).

25. Van der Linden, L., Wolthers, K. C. & van Kuppeveld, F. J. Replication and inhibitors of enteroviruses and parechoviruses. *Viruses* **7,** 4529–4562 (2015).

26. Agol, V. I. & Gmyl, A. P. Viral security proteins: counteracting host defences. *Nat Rev Microbiol* **8,** 867–878 (2010).

27. De Crom, S. C., Rossen, J. W., van Furth, A. M. & Obihara, C. C. Enterovirus and parechovirus infection in children: a brief overview. *Eur J Pediatr* **175,** 1023–1029 (2016).

28. Mirand, A. *et al.* Outbreak of hand, foot and mouth disease/herpangina associated with coxsackievirus A6 and A10 infections in 2010, France: a large citywide, prospective observational study. *Clin Microbiol Infect* **18,** e110–e118 (2012).

29. Wong, A. H., Lau, C. S., Cheng, P. K., Ng, A. Y. & Lim, W. W. Coxsackievirus B3-associated aseptic meningitis: an emerging infection in Hong Kong. *J Med Virol* **83,** 483–989 (2011).

30. Nakamura, H. *et al.* Repetitive coxsackievirus infection induces cardiac dilatation in post-myocarditic mice. *Jpn Circ J* **63,** 794–802 (1999).

31. Badorff, C. & Knowlton, K. U. Dystrophin disruption in enterovirus-induced myocarditis and dilated cardiomyopathy: from bench to bedside. *Med Microbiol Immunol* **193,** 121–126 (2004).

32. Bowles, N. E., Richardson, P. J., Olsen, E. G. & Archard, L. C. Detection of coxsackie-B-virus-specific RNA sequences in myocardial biopsy samples from patients with myocarditis and dilated cardiomyopathy. *Lancet* **1,** 1120–1123 (1986).

33. Bergamin, C. S. & Dib, S. A. Enterovirus and type 1 diabetes: What is the matter? *World J Diabetes* **6,** 828–839 (2015).

34. Richer, M. J. & Horwitz, M. S. Coxsackievirus infection as an environmental factor in the etiology of type 1 diabetes. *Autoimmun Rev* **8,** 611–615 (2009).

35. Gkrania-Klotsas, E. *et al.* The association between prior infection with five serotypes of Coxsackievirus B and incident type 2 diabetes mellitus in the EPIC-Norfolk study. *Diabetologia* **55,** 967–970 (2012).

36. Mantri, S. & Shah, B. B. Enterovirus causes rapidly progressive dementia in a 28-year-old immunosuppressed woman. *J Neurovirol* **22,** 538–540 (2016).

37. Lau, S. K. P. *et al.* Enterovirus D68 infections associated with severe respiratory illness in elderly patients and emergence of a novel clade in Hong Kong. *Scientific Reports* **6** (2016).

38. Tokarz, R. *et al.* Worldwide emergence of multiple clades of enterovirus 68. *J Gen Virol* **93,** 1952–1958 (2012).

39. Leveque, N. & Semler, B. L. A 21st century perspective of poliovirus replication. *Plos Pathog* **11** (2015).

40. Garon, J. R., Cochi, S. L. & Orenstein, W. A. The challenge of global poliomyelitis eradication. *Infect Dis Clin N Am* **29,** 651 (2015).

41. Jacobs, S. E., Lamson, D. M., St George, K. & Walsh, T. J. Human Rhinoviruses. *Clin Microbiol Rev* **26,** 135–162 (2013).

42. Global Burden of Disease Study, C. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **386,** 743–800 (2015).

43. Benkovic, S. J. & Hammes-Schiffer, S. A perspective on enzyme catalysis. *Science* **301,** 1196–1202 (2003).

44. Imming, P., Sinning, C. & Meyer, A. Opinion - Drugs, their targets and the nature and number of drug targets. *Nat Rev Drug Discov* **5,** 821–834 (2006).

45. Zimmermann, H., Hewlett, G. & Rubsamen-Waigmann, H. Other inhibitors of viral enzymes and functions. *Handb Exp Pharmacol,* 155–176 (2009).

46. Dixon, H. B. F. & Cornishbowden, A. (I.-I. J. C. o. B. N. Revision of enzyme nomenclature - Listing enzymes. *Eur J Biochem* **133,** 479–479 (1983).

47. Fischer, E. *Untersuchungen über Kohlenhydrate und Fermente (1884 - 1908)* 912 (Springer, 1909).

48. Koshland, D. E. Application of a theory of enzyme specificity to protein synthesis. *Proc Natl Acad Sci USA* **44,** 98–104 (1958).

49. Lienhard, G. E. Enzymatic catalysis and transition-state theory. *Science* **180,** 149–154 (1973).

50. Copeland, R. A. *Evaluation of enzyme inhibitors in drug discovery. A guide for medicinal chemists and pharmacologists* 1–265 (Wiley, 2005).

51. Turk, B., Turk, D. & Salvesen, G. S. Regulating cysteine protease activity: Essential role of protease inhibitors as guardians and regulators. *Curr Pharm Design* **8,** 1623–1637 (2002).

52. Overall, C. M. & Dean, R. A. Degradomics: Systems biology of the protease web. Pleiotropic roles of MMPs in cancer. *Cancer Metast Rev* **25,** 69–75 (2006).

53. Mohamed, M. M. & Sloane, B. F. Cysteine cathepsins: multifunctional enzymes in cancer. *Nature Reviews Cancer* **6,** 764–775 (2006).

54. Puente, X. S., Sanchez, L. M., Overall, C. M. & Lopez-Otin, C. Human and mouse proteases: a comparative genomic approach. *Nat Rev Genet* **4,** 544–558 (2003).

55. Deu, E., Verdoes, M. & Bogyo, M. New approaches for dissecting protease functions to improve probe development and drug discovery. *Nat Struct Mol Biol* **19,** 9–16 (2012).

56. Turk, B. Targeting proteases: successes, failures and future prospects. *Nat Rev Drug Discov* **5,** 785–799 (2006).

57. Schechter, I. & Berger, A. On active site of proteases . 3. Mapping active site of papain - Specific peptide inhibitors of papain. *Biochem Bioph Res Co* **32,** 898 (1968).

58. Berger, A. & Schechte.I. Mapping active site of papain with aid of peptide substrates and inhibitors. *Philos T Roy Soc B* **257,** 249 (1970).

59. Tyndall, J. D. A., Nall, T. & Fairlie, D. P. Proteases universally recognize beta strands in their active sites. *Chem Rev* **105,** 973–999 (2005).

60. Rawlings, N. D. & Salvesen, G. *Handbook of proteolytic enzymes* (Academic Press, 2013).

61. James, M. N. G. An X-ray crystallographic approach to enzyme structure and function. *Can J Biochem Cell B* **58,** 251–271 (1980).

62. Kraut, J. Serine proteases - Structure and mechanism of catalysis. *Annu Rev Biochem* **46,** 331–358 (1977).

63. Hedstrom, L. Serine protease mechanism and specificity. *Chem Rev* **102,** 4501–4523 (2002).

64. Polgar, L. Mercaptide-imidazolium ion-pair: The reactive nucleophile in papain catalysis. *FEBS Lett* **47,** 15–18 (1974).

65. Craik, C. S., Roczniak, S., Largman, C. & Rutter, W. J. The catalytic role of the active-site aspartic-acid in serine proteases. *Science* **237,** 909–913 (1987).

66. Barrett, A. J. & Rawlings, N. D. Evolutionary lines of cysteine peptidases. *Biol Chem* **382,** 727–733 (2001).

67. Mosimann, S. C., Cherney, M. M., Sia, S., Plotch, S. & James, M. N. G. Refined X-ray crystallographic structure of the poliovirus 3C gene product. *J Mol Biol* **273,** 1032–1047 (1997).

68. Cordingley, M. G., Callahan, P. L., Sardana, V. V., Garsky, V. M. & Colonno, R. J. Substrate requirements of human rhinovirus 3c protease for peptide cleavage in vitro. *J Biol Chem* **265,** 9062–9065 (1990).

69. Sun, D., Chen, S., Cheng, A. C. & Wang, M. S. Roles of the picornaviral 3C proteinase in the viral life cycle and host cells. *Viruses-Basel* **8** (2016).

70. Chau, D. H. *et al.* Coxsackievirus B3 proteases 2A and 3C induce apoptotic cell death through mitochondrial injury and cleavage of eIF4GI but not DAP5/p97/NAT1. *Apoptosis* **12,** 513–524 (2007).

71. Mukherjee, A. *et al.* The coxsackievirus B 3C protease cleaves MAVS and TRIF to attenuate host type I interferon and apoptotic signaling. *PLoS Pathog* **7** (2011).

72. Qu, L. *et al.* Disruption of TLR3 signaling Due to Cleavage of TRIF by the hepatitis A virus protease-polymerase processing intermediate, 3CD. *Plos Pathog* **7** (2011).

73. Gamarnik, A. V. & Andino, R. Interactions of viral protein 3CD and poly(rC) binding protein with the 5'-untranslated region of the poliovirus genome. *J Virol* **74,** 2219–2226 (2000).

74. Nayak, A. *et al.* Role of RNA structure and RNA binding activity of foot-and-mouth disease virus 3C protein in VPg uridylylation and virus replication. *J Virol* **80,** 9865–9875 (2006).

75. Ohlenschlager, O. *et al.* The structure of the stemloop D subdomain of coxsackievirus B3 cloverleaf RNA and its interaction with the proteinase 3C. *Structure* **12,** 237–248 (2004).

76. Claridge, J. K. *et al.* A picornaviral loop-to-loop replication complex. *J Struct Biol* **166,** 251–262 (2009).

77. Shih, S. R. *et al.* Mutations at KFRDI and VGK domains of enterovirus 71 3C protease affect its RNA binding and proteolytic activities. *J Biomed Sci* **11,** 239–248 (2004).

78. Wang, Q. M. Protease inhibitors as potential antiviral agents for the treatment of picornaviral infections. *Prog Drug Res* **Spec No,** 229–253 (2001).

79. Patick, A. K. *et al.* In vitro antiviral activity of AG7088, a potent inhibitor of human rhinovirus 3C protease. *Antimicrob Agents Ch* **43,** 2444–2450 (1999).

80. Webber, S. E. *et al.* Tripeptide aldehyde inhibitors of human rhinovirus 3C protease: Design, synthesis, biological evaluation, and cocrystal structure solution of P-1 glutamine isosteric replacements. *J Med Chem* **41,** 2786–2805 (1998).

81. Dragovich, P. S. *et al.* Structure-based design, synthesis, and biological evaluation of irreversible human rhinovirus 3C protease inhibitors. 1. Michael acceptor structure-activity studies. *J Med Chem* **41,** 2806–2818 (1998).

82. Dragovich, P. S. *et al.* Structure-based design, synthesis, and biological evaluation of irreversible human rhinovirus 3C protease inhibitors. 2. Peptide structure-activity studies. *J Med Chem* **41,** 2819–2834 (1998).

83. Dragovich, P. S. *et al.* Structure-based design, synthesis, and biological evaluation of irreversible human rhinovirus 3C protease inhibitors. 3. Structure - Activity studies of ketomethylene-containing peptidomimetics. *J Med Chem* **42,** 1203–1212 (1999).

84. Dragovich, P. S. *et al.* Structure-based design, synthesis, and biological evaluation of irreversible human rhinovirus 3C protease inhibitors. 4. Incorporation of P-1 lactam moieties as L-glutamine replacements. *J Med Chem* **42,** 1213–1224 (1999).

85. Dragovich, P. S. *et al.* Structure-based design, synthesis, and biological evaluation of irreversible human rhinovirus 3C protease inhibitors. 6. Structure-activity studies of orally bioavailable, 2-pyridone-containing peptidomimetics. *J Med Chem* **45,** 1607–1623 (2002).

86. Dragovich, P. S. *et al.* Structure-based design, synthesis, and biological evaluation of irreversible human rhinovirus 3C protease inhibitors. Part 7: Structure-activity studies of bicyclic 2-pyridone-containing peptidomimetics. *Bioorg Med Chem Lett* **12,** 733–738 (2002).

87. Hayden, F. G. *et al.* Phase 2, randomized, double-blind, placebo-controlled studies of ruprintrivir nasal spray 2-percent suspension for prevention and treatment of experimentally induced rhinovirus colds in healthy volunteers. *Antimicrob Agents Chemother* **47,** 3907–3916 (2003).

88. Patick, A. K. *et al.* In vitro antiviral activity and single-dose pharmacokinetics in humans of a novel, orally bioavailable inhibitor of human rhinovirus 3C protease. *Antimicrob Agents Ch* **49,** 2267–2275 (2005).

89. Dragovich, P. S. *et al.* Structure-based design, synthesis, and biological evaluation of irreversible human rhinovirus 3C protease inhibitors. 8. Pharmacological optimization of orally bioavailable 2-pyridone containing peptidomimetics. *J Med Chem* **46,** 4572–4585 (2003).

90. Zhang, X. N. *et al.* Rupintrivir is a promising candidate for treating severe cases of enterovirus-71 infection: Evaluation of antiviral efficacy in a murine infection model. *Antivir Res* **97,** 264–269 (2013).

91. Lim, B. K. *et al.* Soluble coxsackievirus B3 3C protease inhibitor prevents cardiomyopathy in an experimental chronic myocarditis murine model. *Virus Res* **199,** 1–8 (2015).

92. Zhang, X. N. *et al.* Rupintrivir is a promising candidate for treating severe cases of Enterovirus-71 infection. *World J Gastroentero* **16,** 201–209 (2010).

93. Smee, D. F., Evans, W. J., Nicolaou, K. C., Tarbet, E. B. & Day, C. W. Susceptibilities of enterovirus D68, enterovirus 71, and rhinovirus 87 strains to various antiviral compounds. *Antiviral Res* **131,** 61–65 (2016).

94. Kim, B. K. *et al.* Development of anti-coxsackievirus agents targeting 3C protease. *Bioorg Med Chem Lett* **22,** 6952–6956 (2012).

95. Zeng, D. B. *et al.* Synthesis and structure-activity relationship of alpha-keto amides as enterovirus 71 3C protease inhibitors. *Bioorg Med Chem Lett* **26,** 1762–1766 (2016).

96. Kawatkar, S. P. *et al.* Design and structure-activity relationships of novel inhibitors of human rhinovirus 3C protease. *Bioorg Med Chem Lett* **26,** 3248–3252 (2016).

97. Tan, J. Z. *et al.* 3C Protease of enterovirus 68: Structure-based design of Michael acceptor inhibitors and their broad-spectrum antiviral effects against picornaviruses. *J Virol* **87,** 4339–4351 (2013).

98. Reich, S. H. *et al.* Substituted benzamide inhibitors of human rhinovirus 3C protease: Structure-based design, synthesis, and biological evaluation. *J Med Chem* **43,** 1670–1683 (2000).

99. Webber, S. E. *et al.* Design, synthesis, and evaluation of nonpeptidic inhibitors of human rhinovirus 3C protease. *J Med Chem* **39,** 5072–5082 (1996).

100. Johnson, T. O. *et al.* Structure-based design of a parallel synthetic array directed toward the discovery of irreversible inhibitors of human rhinovirus 3C protease. *J Med Chem* **45,** 2016–2023 (2002).

101. Ma, G. H. *et al.* Identification and biochemical characterization of DC07090 as a novel potent small molecule inhibitor against human enterovirus 71 3C protease by structure-based virtual screening. *Eur J Med Chem* **124,** 981–991 (2016).

102. Kim, B. K. *et al.* Benserazide, the first allosteric inhibitor of coxsackievirus B3 3C protease. *Febs Lett* **589,** 1795–1801 (2015).

103. Baxter, A. *et al.* Non-covalent inhibitors of rhinovirus 3C protease. *Bioorg Med Chem Lett* **21,** 777–780 (2011).

104. Kuo, C. J. *et al.* Individual and common inhibitors of coronavirus and picornavirus main proteases. *Febs Lett* **583,** 549–555 (2009).

105. Ramajayam, R., Tan, K. P., Liu, H. G. & Liang, P. H. Synthesis and evaluation of pyrazolone compounds as SARS-coronavirus 3C-like protease inhibitors. *Bioorgan Med Chem* **18,** 7849–7854 (2010).

106. Segel, I. H. Enzyme-Kinetics. *Bioscience* **26,** 425–426 (1976).

107. Lineweaver, H. & Burk, D. The determination of enzyme dissociation constants. *J Am Chem Soc* **3,** 658–666 (1934).

108. Dixon, M. Graphical determination of equilibrium constants. *Biochem J* **94,** 760 (1965).

109. Cornish-Bowden, A. A simple graphical method for determining the inhibition constants of mixed, uncompetitive and non-competitive inhibitors. *Biochem J* **137,** 143–4 (1974).

110. Cheng, Y. & Prusoff, W. H. Relationship between inhibition constant (K1) and concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic-reaction. *Biochem Pharmacol* **22,** 3099–3108 (1973).

111. Roth, G. J., Stanford, N. & Majerus, P. W. Acetylation of prostaglandin synthase by aspirin. *P Natl Acad Sci USA* **72,** 3073–3076 (1975).

112. Bryan, J. Carbapenems have stood the test of time so far but resistance is emerging. *Pharm J* **292** (2014).

113. Mullin, R. Before the storm. *Chem Rev* **89,** 12 (2011).

114. Nakayama, S. *et al.* A zone classification system for risk assessment of idiosyncratic drug toxicity using daily dose and covalent binding. *Drug Metab Dispos* **37,** 1970–1977 (2009).

115. Singh, J., Petter, R. C., Baillie, T. A. & Whitty, A. The resurgence of covalent drugs. *Nat Rev Drug Discov* **10,** 307–317 (2011).

116. Bauer, R. A. Covalent inhibitors in drug discovery: From accidental discoveries to avoided liabilities and designed therapies. *Drug Discov Today* **20,** 1061–1073 (2015).

117. Copeland, R. A. The dynamics of drug-target interactions: Drug-target residence time and its impact on efficacy and safety. *Expert Opin Drug Discov* **5,** 305–310 (2010).

118. Lammert, C. *et al.* Relationship between daily dose of oral medications and idiosyncratic drug-induced liver injury: Search for signals. *Hepatology* **47,** 2003–2009 (2008).

119. Smith, A. J., Zhang, X., Leach, A. G. & Houk, K. N. Beyond picomolar affinities: Quantitative aspects of noncovalent and covalent binding of drugs to proteins. *J Med Chem* **52,** 225–233 (2009).

120. Johnson, D. S., Weerapana, E. & Cravatt, B. F. Strategies for discovering and derisking covalent, irreversible enzyme inhibitors. *Future Med Chem* **2,** 949–964 (2010).

121. Lim, S. M. *et al.* Therapeutic targeting of oncogenic K-Ras by a covalent catalytic site inhibitor. *Angew Chem Int Edit* **53,** 199–204 (2014).

122. Miller, R. M., Paavilainen, V. O., Krishnan, S., Serafimova, I. M. & Taunton, J. Electrophilic fragment-based design of reversible covalent kinase inhibitors. *J Am Chem Soc* **135,** 5298–5301 (2013).

123. Mah, R., Thomas, J. R. & Shafer, C. M. Drug discovery considerations in the development of covalent inhibitors. *Bioorg Med Chem Lett* **24,** 33–39 (2014).

124. Potashman, M. H. & Duggan, M. E. Covalent modifiers: an orthogonal approach to drug design. *J Med Chem* **52,** 1231–46 (2009).

125. Krantz, A. A classification of enzyme-inhibitors. *Bioorg Med Chem Lett* **2,** 1327–1334 (1992).

126. Powers, J. C., Asgian, J. L., Ekici, O. D. & James, K. E. Irreversible inhibitors of serine, cysteine, and threonine proteases. *Chem Rev* **102,** 4639–4750 (2002).

127. Shannon, D. A. & Weerapana, E. Covalent protein modification: The current landscape of residue-specific electrophiles. *Curr Opin Chem Biol* **24,** 18–26 (2015).

128. Smoum, R., Rubinstein, A., Dembitsky, V. M. & Srebnik, M. Boron containing compounds as protease inhibitors. *Chem Rev* **112,** 4156–4220 (2012).

129. Otto, H. H. & Schirmeister, T. Cysteine proteases and their inhibitors. *Chem Rev* **97,** 133–172 (1997).

130. Erlanson, D. A., Fesik, S. W., Hubbard, R. E., Jahnke, W. & Jhoti, H. Twenty years on: The impact of fragments on drug discovery. *Nat Rev Drug Discov* (2016).

131. Bollag, G. *et al.* Vemurafenib: The first drug approved for BRAF-mutant cancer. *Nat Rev Drug Discov* **11,** 873–886 (2012).

132. Mullard, A. Pioneering apoptosis-targeted cancer drug poised for FDA approval. *Nat Rev Drug Discov* **15,** 147–149 (2016).

133. Jencks, W. P. On the attribution and additivity of binding-energies. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **78,** 4046–4050 (1981).

134. English, A. C., Groom, C. R. & Hubbard, R. E. Experimental and computational mapping of the binding surface of a crystalline protein. *Protein Engineering* **14,** 47–59 (2001).

135. Goodford, P. J. A computational-procedure for determining energetically favorable binding-sites on biologically important macromolecules. *J Med Chem* **28,** 849–857 (1985).

136. Shuker, S. B., Hajduk, P. J., Meadows, R. P. & Fesik, S. W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **274,** 1531–1534 (1996).

137. Hann, M. M., Leach, A. R. & Harper, G. Molecular complexity and its impact on the probability of finding leads for drug discovery. *J Chem Inf Comp Sci* **41,** 856–864 (2001).

138. Kuntz, I. D., Chen, K., Sharp, K. A. & Kollman, P. A. The maximal affinity of ligands. *P Natl Acad Sci USA* **96,** 9997–10002 (1999).

139. Ruddigkeit, L., van Deursen, R., Blum, L. C. & Reymond, J. L. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *J Chem Inf Model* **52,** 2864–2875 (2012).

140. Erlanson, D. A., McDowell, R. S. & O'Brien, T. Fragment-based drug discovery. *J Med Chem* **47,** 3463–3482 (2004).

141. Hopkins, A. L., Keseru, G. M., Leeson, P. D., Rees, D. C. & Reynolds, C. H. The role of ligand efficiency metrics in drug discovery. *Nat Rev Drug Discov* **13,** 105–121 (2014).

142. Hajduk, P. J. Fragment-based drug design: How big is too big? *J Med Chem* **49,** 6972–6976 (2006).

143. Keseru, G. M. *et al.* Design principles for fragment libraries: Maximizing the value of learnings from pharma fragment-based drug discovery (FBDD) programs for use in academia. *J Med Chem* (2016).

144. Jhoti, H., Williams, G., Rees, D. C. & Murray, C. W. The 'rule of three' for fragment-based drug discovery: Where are we now? *Nat Rev Drug Discov* (2013).

145. Davis, B. J. & Erlanson, D. A. Learning from our mistakes: the 'unknown knowns' in fragment screening. *Bioorg Med Chem Lett* **23,** 2844–2852 (2013).

146. Scott, D. E., Coyne, A. G., Hudson, S. A. & Abell, C. Fragment-based approaches in drug discovery and chemical biology. *Biochemistry* **51,** 4990–5003 (2012).

147. Barelier, S. *et al.* Increasing chemical space coverage by combining empirical and computational fragment screens. *Acs Chem Biol* **9,** 1528–1535 (2014).

148. Verdonk, M. L. *et al.* Virtual screening using protein-ligand docking: Avoiding artificial enrichment. *J Chem Inf Comp Sci* **44,** 793–806 (2004).

149. Marcou, G. & Rognan, D. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J Chem Inf Model* **47,** 195–207 (2007).

150. Da, C. & Kireev, D. Structural protein-ligand interaction fingerprints (SPLIF) for structure-based virtual screening: Method and benchmark study. *J Chem Inf Model* **54,** 2555–2561 (2014).

151. Wolber, G., Seidel, T., Gökhan, I. & Bendix, F. Strategies for 3D pharmacophore-based virtual screening. *Drug Discov Today Technol* **7,** e221–e223 (2010).

152. Schiebel, J. *et al.* One question, multiple answers: Biochemical and biophysical screening methods retrieve deviating fragment hit lists. *ChemMedChem* **10,** 1511–1521 (2015).

153. Teplitsky, E. *et al.* High throughput screening using acoustic droplet ejection to combine protein crystals and chemical libraries on crystallization plates at high density. *J Struct Biol* **191,** 49–58 (2015).

154. Joseph-McCarthy, D., Campbell, A. J., Kern, G. & Moustakas, D. Fragment-based lead discovery and design. *J Chem Inf Model* **54,** 693–704 (2014).

155. Ferenczy, G. G. & Keseru, G. M. How are fragments optimized? A retrospective analysis of 145 fragment optimizations. *J Med Chem* **56,** 2478–2486 (2013).

156. Chung, S., Parker, J. B., Bianchet, M., Amzel, L. M. & Stivers, J. T. Impact of linker strain and flexibility in the design of a fragment-based inhibitor. *Nat Chem Biol* **5,** 407–413 (2009).

157. Ward, R. A. *et al.* Design and synthesis of novel lactate dehydrogenase A inhibitors by fragment-based lead generation. *J Med Chem* **55,** 3285–3306 (2012).

158. Woolford, A. J. *et al.* Exploitation of a novel binding pocket in human lipoprotein-associated phospholipase A2 (Lp-PLA2) discovered through X-ray fragment screening. *J Med Chem* **59,** 5356–5367 (2016).

159. Yokokawa, F. *et al.* Discovery of potent non-nucleoside inhibitors of dengue viral RNA-dependent RNA polymerase from a fragment hit using structure-based Drug Design. *J Med Chem* **59,** 3935–3952 (2016).

160. Coyne, A. G., Scott, D. E. & Abell, C. Drugging challenging targets using fragment-based approaches. *Curr Opin Chem Biol* **14,** 299–307 (2010).

161. Wermuth, G., Ganellin, C. R., Lindberg, P. & Mitscher, L. A. Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). *Pure Appl Chem* **70,** 1129–1143 (1998).

162. Kirchmair, J. *et al.* Fast and efficient in silico 3D screening: Toward maximum computational efficiency of pharmacophore-based and shape-based approaches. *J Chem Inf Model* **47,** 2182–2196 (2007).

163. Kurogi, Y. & Guner, O. F. Pharmacophore modeling and three-dimensional database searching for drug design using catalyst. *Curr Med Chem* **8,** 1035–1055 (2001).

164. Wolber, G. & Langer, T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J Chem Inf Model* **45,** 160–169 (2005).

165. Labute, P., Williams, C., Feher, M., Sourial, E. & Schmidt, J. M. Flexible alignment of small molecules. *J Med Chem* **44,** 1483–1490 (2001).

166. Dixon, S. L., Smondyrev, A. M. & Rao, S. N. PHASE: A novel approach to pharmacophore modeling and 3D database searching. *Chem Biol Drug Des* **67,** 370–372 (2006).

167. Spitzer, G. M. *et al.* One concept, three implementations of 3D pharmacophore-based virtual screening: Distinct coverage of chemical search space. *J Chem Inf Model* **50,** 1241–1247 (2010).

168. Wolber, G., Dornhofer, A. A. & Langer, T. Efficient overlay of small organic molecules using 3D pharmacophores. *J Comput Aid Mol Des* **20,** 773–788 (2006).

169. Schulz, R. *Design of novel covalent coxsackievirus B3 3C protease inhibitors.* Master thesis (Martin-Luther-Universitaet Halle-Wittenberg, Halle, Germany, 2015).

170. Mysinger, M. M., Carchia, M., Irwin, J. J. & Shoichet, B. K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J Med Chem* **55,** 6582–6594 (2012).

171. Nicholls, A. What do we know and when do we know it? *J Comput Aid Mol Des* **22,** 239–255 (2008).

172. Triballeau, N., Acher, F., Brabet, I., Pin, J. P. & Bertrand, H. O. Virtual screening workflow development guided by the "receiver operating characteristic" curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *J Med Chem* **48,** 2534–2547 (2005).

173. Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* **30,** 2785–2791 (2009).

174. Meier, R., Pippel, M., Brandt, F., Sippl, W. & Baldauf, C. ParaDockS: a framework for molecular docking with population-based metaheuristics. *J Chem Inf Model* **50,** 879–889 (2010).

175. Ewing, T. J., Makino, S., Skillman, A. G. & Kuntz, I. D. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des* **15,** 411–428 (2001).

176. Kramer, B., Rarey, M. & Lengauer, T. Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. *Proteins* **37,** 228–241 (1999).

177. Friesner, R. A. *et al.* Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* **47,** 1739–1749 (2004).

178. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. & Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins* **52,** 609–623 (2003).

179. Kitchen, D. B., Decornez, H., Furr, J. R. & Bajorath, J. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nat Rev Drug Discov* **3,** 935–949 (2004).

180. Huang, S. Y., Grinter, S. Z. & Zou, X. Scoring functions and their evaluation methods for protein-ligand docking: Recent advances and future directions. *Phys Chem Chem Phys* **12,** 12899–12908 (2010).

181. Warren, G. L. *et al.* A critical assessment of docking programs and scoring functions. *J Med Chem* **49,** 5912–5931 (2006).

182. Zhou, Z. Y., Felts, A. K., Friesner, R. A. & Levy, R. M. Comparative performance of several flexible docking programs and scoring functions: Enrichment studies for a diverse set of pharmaceutically relevant targets. *J Chem Inf Model* **47,** 1599–1608 (2007).

183. Bostrom, J., Berggren, K., Elebring, T., Greasley, P. J. & Wilstermann, M. Scaffold hopping, synthesis and structure-activity relationships of 5,6-diarylpyrazine-2-amide derivatives: A novel series of CB1 receptor antagonists. *Bioorgan Med Chem* **15,** 4077–4084 (2007).

184. Rush, T. S., Grant, J. A., Mosyak, L. & Nicholls, A. A shape-based 3D scaffold hopping method and its application to a bacterial protein-protein interaction. *J Med Chem* **48,** 1489–1495 (2005).

185. Singh, J. *et al.* Successful shape-based virtual screening: The discovery of a potent inhibitor of the type I TGF beta receptor kinase (T beta RI). *Bioorg Med Chem Lett* **13,** 4355–4359 (2003).

186. Grant, J. A., Gallardo, M. A. & Pickup, B. T. A fast method of molecular shape comparison: A simple application of a Gaussian description of molecular shape. *J Comput Chem* **17,** 1653–1666 (1996).

187. Mortier, J. *et al.* The impact of molecular dynamics on drug design: applications for the characterization of ligand-macromolecule complexes. *Drug Discov Today* **20,** 686–702 (2015).

188. Adcock, S. A. & McCammon, J. A. Molecular dynamics: Survey of methods for simulating the activity of proteins. *Chem Rev* **106,** 1589–1615 (2006).

189. Mackerell, A. D. Empirical force fields for biological macromolecules: Overview and issues. *J Comput Chem* **25,** 1584–1604 (2004).

190. Cornell, W. D. *et al.* A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* **118,** 2309–2309 (1996).

191. MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* **102,** 3586–3616 (1998).

192. Oostenbrink, C., Villa, A., Mark, A. E. & Van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J Comput Chem* **25,** 1656–1676 (2004).

193. Jorgensen, W. L. & Tiradorives, J. The OPLS potential functions for proteins - Energy minimizations for crystals of cyclic-peptides and crambin. *J Am Chem Soc* **110,** 1657–1666 (1988).

194. Dalby, A. *et al.* Description of several chemical-structure file formats used by computer-programs developed at molecular design limited. *J Chem Inf Comp Sci* **32,** 244–255 (1992).

195. Weininger, D. Smiles, a chemical language and information-system .1. Introduction to methodology and encoding rules. *J Chem Inf Comp Sci* **28,** 31–36 (1988).

196. James, C. A. & Weininger, D. *Daylight Theory Manual* 2006.

197. Strelow, J. *et al.* in *Assay Guidance Manual* (eds Sittampalam, G. S. *et al.*) (Bethesda, 2004).

198. Brooks, H. B. *et al.* in *Assay Guidance Manual* (eds Sittampalam, G. S. *et al.*) (Bethesda, 2004).

199. Zhang, G. in *Assay Guidance Manual* (eds Sittampalam, G. S. *et al.*) (Bethesda, 2004).

200. Zimmerman, M., Ashe, B., Yurewicz, E. C. & Patel, G. Sensitive assays for trypsin, elastase, and chymotrypsin using new fluorogenic substrates. *Anal Biochem* **78,** 47–51 (1977).

201. Forster, T. *Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann Phys-Berlin* **2,** 55–75 (1948).

202. Milman, B. L. General principles of identification by mass spectrometry. *Trac-Trend Anal Chem* **69,** 24–33 (2015).

203. Dunn, G. H. Electron-impact ionization. *Ieee T Nucl Sci* **23,** 929–933 (1976).

204. Ho, C. S. *et al.* Electrospray ionisation mass spectrometry: principles and clinical applications. *Clin Biochem Rev* **24,** 3–12 (2003).

205. Karas, M. *et al.* Principles and applications of matrix-assisted UV laser desorption ionization mass-spectrometry. *Anal Chim Acta* **241,** 175–185 (1990).

206. Cottrell, J. S. & Greathead, R. J. Extending the mass range of a sector mass-spectrometer. *Mass Spectrom Rev* **5,** 215–247 (1986).

207. Guilhaus, M. Principles and instrumentation in time-of-flight mass spectrometry - Physical and instrumental concepts. *J Mass Spectrom* **30,** 1519–1532 (1995).

208. Chernushevich, I. V., Loboda, A. V. & Thomson, B. A. An introduction to quadrupole-time-of-flight mass spectrometry. *J Mass Spectrom* **36,** 849–865 (2001).

209. Schwartz, J. C., Senko, M. W. & Syka, J. E. P. A two-dimensional quadrupole ion trap mass spectrometer. *J Am Soc Mass Spectr* **13,** 659–669 (2002).

210. Backus, K. M. *et al.* Proteome-wide covalent ligand discovery in native biological systems. *Nature* **534,** 570–574 (2016).

211. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28,** 235–242 (2000).

212. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7** (2011).

213. Lu, G. W. *et al.* Enterovirus 71 and coxsackievirus A16 3C proteases: Binding to rupintrivir and their substrates and anti-hand, foot, and mouth disease virus drug design. *J Virol* **85,** 10319–10331 (2011).

214. Vanhaverbeke, Y., Maquestiau, A. & Vandeneynde, J. J. Synthesis of 1H-pyrazolo[3,4-b]pyridines and pyrazolo[1,5-a]pyrimidines. *J Heterocyclic Chem* **16,** 773–777 (1979).

215. Lisurek, M. *et al.* Design of chemical libraries with potentially bioactive molecules applying a maximum common substructure concept. *Mol Divers* **14,** 401–408 (2010).

216. Schmidt, M. F., El-Dahshan, A., Keller, S. & Rademann, J. Selective identification of cooperatively binding fragments in a high-throughput ligation assay enables development of a picomolar caspase-3 inhibitor. *Angew Chem Int Edit* **48,** 6346–6349 (2009).

217. Riano, V. M. & Rademann, J. Formylglycine-peptides for the identification of novel pTyr mimetics through a fragment based dynamic ligation screening (DLS) approach. *J Pept Sci* **18** (2012).

218. Fernandez-Bachiller, M. I., Horatscheck, A., Lisurek, M. & Rademann, J. Alzheimer's disease: Identification and development of β-secretase (BACE-1) binding fragments and inhibitors by dynamic ligation screening (DLS). *ChemMedChem* **8,** 1041–1056 (2013).

219. Burda, E. & Rademann, J. Catalytic activation of pre-substrates via dynamic fragment assembly on protein templates. *Nat Commun* **5** (2014).

220. Becker, D. *et al.* Irreversible inhibitors of the 3C protease of Coxsackie virus through templated assembly of protein-binding fragments. *Nat Commun* **7** (2016).

221. Vodolazhenko, M. A. *et al.* Application of stable fused dienolates for diversity oriented synthesis of 2,5-dioxo-5,6,7,8-tetrahydro-2H-chromene-3-carboxamides. *Rsc Advances* **2,** 1106–1111 (2012).

222. Choong, I. C. *et al.* Identification of potent and selective small-molecule inhibitors of caspase-3 through the use of extended tethering and structure-based drug design. *Journal of Medicinal Chemistry* **45,** 5005–5022 (2002).

223. Schneider, G. De novo design - hop(p)ing against hope. *Drug Discov Today Technol* **10,** 453–460 (2013).

224. Hartenfeller, M. *et al.* Probing the bioactivity-relevant chemical space of robust reactions and common molecular building blocks. *J Chem Inf Model* **52,** 1167–1178 (2012).

225. Hartenfeller, M. *et al.* A collection of robust organic synthesis reactions for in silico molecule design. *J Chem Inf Model* **51,** 3093–3098 (2011).

226. Roughley, S. D. & Jordan, A. M. The medicinal chemist's toolbox: An analysis of reactions used in the pursuit of drug candidates. *J Med Chem* **54,** 3451–3479 (2011).

227. Ouyang, X. *et al.* CovalentDock: Automated covalent docking with parameterized covalent linkage energy estimation and molecular geometry constraints. *J Comput Chem* **34,** 326–336 (2013).

228. Warshaviak, D. T., Golan, G., Borrelli, K. W., Zhu, K. & Kalid, O. Structure-based virtual screening approach for discovery of covalently bound ligands. *J Chem Inf Model* **54,** 1941–1950 (2014).

229. Willett, P., Barnard, J. M. & Downs, G. M. Chemical similarity searching. *J Chem Inf Comp Sci* **38,** 983–996 (1998).

230. Kozakov, D. *et al.* The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat Protoc* **10,** 733–755 (2015).

231. Amero, C. D., Arnold, J. J., Moustafa, I. M., Cameron, C. E. & Foster, M. P. Identification of the oriI-binding site of poliovirus 3C protein by nuclear magnetic resonance spectroscopy. *J Virol* **82,** 4363–4370 (2008).

232. Walker, P. K., Leong, L. E. C. & Porter, A. G. Sequence and structural determinants of the interaction between the 5'-noncoding region of picornavirus RNA and rhinovirus protease 3C. *J Biol Chem* **270,** 14510–14516 (1995).

233. Blair, W. S., Nguyen, J. H. C., Parsley, T. B. & Semler, B. L. Mutations in the poliovirus 3CD proteinase S1-specificity pocket affect substrate recognition and RNA binding. *Virology* **218,** 1–13 (1996).

234. Sydow, D. *Dynophores: Novel dynamic pharmacophores - Implementation of pharmacophore generation based on molecular dynamics trajectories and their graphical representation.* Master thesis (Humboldt-Universitaet zu Berlin, Berlin, Germany, 2015).

235. Ekonomiuk, D. & Caflisch, A. Activation of the West Nile virus NS3 protease: Molecular dynamics evidence for a conformational selection mechanism. *Protein Sci* **18,** 1003–1011 (2009).

236. Hajduk, P. J. & Greer, J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat Rev Drug Discov* **6,** 211–219 (2007).

237. Sadowski, J., Gasteiger, J. & Klebe, G. Comparison of automatic 3-dimensional model builders using 639 X-ray structures. *J Chem Inf Comp Sci* **34,** 1000–1008 (1994).

238. Hawkins, P. C. D., Skillman, A. G., Warren, G. L., Ellingson, B. A. & Stahl, M. T. Conformer generation with OMEGA: Algorithm and validation using high quality structures from The Protein Databank and Cambridge Structural Database. *J Chem Inf Model* **50,** 572–584 (2010).

239. Korb, O., Stutzle, T. & Exner, T. E. Empirical scoring functions for advanced protein-ligand docking with PLANTS. *J Chem Inf Model* **49,** 84–96 (2009).

240. Bowers, K. J. *et al. Scalable algorithms for molecular dynamics simulations on commodity clusters* in *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06)* (2006).

241. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J Mol Graph Model* **14,** 33–38 (1996).

242. *KNIME community contributions* <https://tech.knime.org/community> (2016).

243. *CDK* <https://sourceforge.net/projects/cdk/> (2016).

244. *Erl Wood Cheminformatics nodes for KNIME* <https://tech.knime.org/community/erlwood> (2016).

245. *The Indigo cheminformatics toolkit* <http://lifescience.opensource.epam.com/indigo/> (2016).

246. *RDKit* <http://www.rdkit.org/> (2016).

247. WDI. *Derwent World Drug Index* (2005).

List of Abbreviations

αK_i	Equilibrium dissociation constant of ESI
ΔG	Free binding energy
$3C^{pro}$	3C Protease
[E]	Enzyme concentration
[ES]	Concentration of enzyme substrate complex
[I]	Inhibitor concentration
[P]	Product concentration
[S]	Substrate concentration
A	Number of actives in a data set
AIDS	Acquired immune deficiency syndrome
BB	Building block
Cre	Cis-acting replication element
CV	Coxsackievirus
CVB3	Coxsackievirus B3
Da	Dalton
DCC	Dicyclocarbodiimde
DCE	Dichlorethane
DCM	Dichlormethane
DMAP	Dimethylaminopyridine
DMF	N,N-Dimethylformamide
DTT	Dithiothreitol
Dynophores	Dynamic 3D pharmacophores
E	Enzyme
E-I	Covalent enzyme inhibitor complex
EC	Enzyme Commission
EF	Enrichment factor
EI	Enzyme inhibitor complex
eIF4G	Eukaryotic translation initiation factor 4G
EP	Enzyme product complex
ES	Initial encounter complex
$ES^{\ddagger}$	Transition state
ESI	Electrospray ionization
ESI	Enzyme substrate inhibitor complex
EV	Enterovirus
Fit	Feature fit score

FN	False negative
FP	False positive
GSH	Glutathione
h	Hill slope
HA	Number of heavy atoms
HAART	Highly active antiretroviral therapy
HCV	Hepatitis C virus
HIV	Human immunodeficiency virus
HRV	Human rhinovirus
HSV	Herpes simplex virus
HTS	High throughput screening
IC_{50}	Inhibitor concentration, which leads to a reduction of 50% in the enzyme reaction velocity
IRES	Internal ribosome entry site
ITC	Isothermal titration calorimetry
k_1	Rate constant for the formation of EI for covalent inhibitor
k_2	Rate constant for the dissociation of EI for covalent inhibitor
k_3	Rate constant for the formation of E-I
k_{cat}	Rate constant of the catalytic reaction
K_d	Equilibrium dissociation constant
k_{inact}	Maximum rate of covalent enzyme inhibition at infinite inhibitor concentration
k_{inact}/K_I	second-order rate constant of enzyme inactivation
K_I	Inhibitor concentration displaying half-maximum covalent enzyme inactivation
K_i	Equilibrium dissociation constant of EI
K_M	Michaelis constant
k_{obs}	Observed pseudo-first order rate constant
k_{off}	Dissociation rate constant
k_{on}	Association rate constant
K_s	Equilibrium dissociation constant of ES
LE	Ligand efficiency
m/z	Mass to charge ratio
MALDI	Matrix-assisted laser desorption/ionization
MD	Molecular dynamics
MD	Molecular dynamics
MOE	Molecular Operating Environment
MS	Mass spectrometry
N	Total number of molecules in a data set
N_{MFP}	Number of geometrically matched feature pairs

NMR	Nuclear magnetic resonance
NNRTI	Non-nucleoside reverse transcriptase inhibitors
NRTI	Nucleoside reverse transcriptase inhibitors
NS5B	Hepatitis C virus RNA-polymerase
NSAID	Nonsteroidal anti-inflammatory drugs
ORF	Open reading frame
P	Product
PAbP	Poly(A)-binding proteins
PAGE	Polyacrylamide gel electrophoresis
PDB	Protein Data Bank
PLIFS	Protein-ligand interaction fingerprints
PLP	Pyridoxal phosphate
PTTB	Phenyltrimethylammonium tribromide
PV	Poliovirus
R	Ideal gas constant
RdRp	RNA-dependent RNA polymerase
RMS	Root mean square
RMS_{FP}	Feature pair distances RMS
RMSD	Root mean square deviation
ROC	Receiver Operating Characteristics
ROCS	Rapid Overlay of Chemical Structures
ROCS	Rapid Overlay of Chemical Structures
S	Substrate
S_{FCR}	Feature count/RMS distance score,
S_{RMS}	Matched feature pair RMS distance score
SAR	Structure-activity relationship
SDS	Sodium dodecyl sulfate
Se	Sensitivity
SLD	Stem loop D
Sp	Specificity
SPR	Surface plasmon resonance
T	Absolute temperature
THF	Tetrahydrofuran
TN	True negative
TOF	Time of flight
TP	True positive
UTR	Untranslated regions
v	Velocity of an enzymatic reaction
v_0	Velocity of an uninhibited enzymatic reaction
v_i	Steady state velocity of an enzymatic reaction

V_{max}	Enzyme reaction velocity at an infinite substrate concentration
v_s	Velocity of an enzymatic reaction in equilibrium with a time-dependent inhibitor
VS	Virtual screening
VS	Virtual screening
Ya	Yield of actives

List of Figures

List of Tables

Publications

Peer-reviewed articles

1. Schulz, R. *et al.* Phenylthiomethyl ketone-based fragments show selective, irreversible inhibition of enteroviral 3C proteases. *Submitted*

2. Schaller, D. *et al.* Systematic Data Mining Reveals Synergistic H3R/MCHR1 Ligands. *ACS Med. Chem. Lett. Article ASAP* (2017).

3. Becker, D. *et al.* Irreversible inhibitors of the 3C protease of Coxsackie virus through templated assembly of protein-binding fragments. *Nat Commun* **7**, 12761 (2016).

Oral presentations

1. Schulz, R., Gottschalk, F., Rademann, R., Wolber, G. Thinking outside the box: Allosteric fragments targeting coxsackievirus 3C protease. 251st ACS National Meeting and Exposition, March 13-18, 2016, San Diego, USA.

2. Schulz, R., Becker, D., Suliman, A.A., Rademann, R., Wolber, G. Reaction-driven design of viral protease inhibitors from covalently binding fragments. 251st ACS National Meeting and Exposition, 2016, March 13-18, San Diego, USA.

3. Schulz, R., Becker, D., Rademann, R., Wolber, G. Fragment-based design of viral protease inhibitors by virtual screening and chemical space sampling. 11th German Conference on Chemoinformatics 2015, November 8-10, Fulda, Germany.

4. Schulz, R., Becker, D., Rademann, R., Wolber, G. Covalently binding fragments: Reaction-driven design of viral protease inhibitors. 8th Vienna Summer School on Drug-Design, September 20-25, 2015, Vienna, Austria.

5. Schulz, R., Becker, D., Rademann, R., Wolber, G. Reaction-driven fragment-based design of viral protease inhibitors. Annual EUROPIN meeting, September 26, 2014, Gdansk, Poland.

6. Schulz, R., Wolber, G. Design of viral protease inhibitors through in-situ fragment linking. Annual EUROPIN meeting, September 20, 2013, Vienna, Austria.

Poster presentations

1. Schulz, R., Gottschalk, F., Rademann, R., Wolber, G. Rationalization of allosteric 3C protease modulation: Fragment-based design and molecular dynamics simulations. Tag der Pharmazie, July 1, 2016, Berlin, Germany.

2. Schulz, R., Becker, D., Rademann, R., Wolber, G. Covalently binding fragments: Computer-aided design of viral protease inhibitors. 5th RSC-BMCS Fragment-based Drug Discovery Meeting, March 22-24, 2015, Cambridge, United Kingdom.

3. Schulz, R., Becker, D., Rademann, R., Wolber, G. Pharmacophore-based virtual screening for the identification of covalent coxsackievirus 3C protease inhibitors. 10th ICCS/GCC Conference, June 1-5, 2014, Noordwijkerhout, The Netherlands.