



# Search for new antiviral compounds using fragment screening methodology

Zuzanna Kaczmarska



Aquesta tesi doctoral està subjecta a la llicència *Reconeixement 3.0. Espanya de Creative Commons.*

Esta tesis doctoral está sujeta a la licencia *Reconocimiento 3.0. España de Creative Commons.*

This doctoral thesis is licensed under the *Creative Commons Attribution 3.0. Spain License.*



UNIVERSITAT DE BARCELONA



Facultat de Farmàcia

Departament de Bioquímica i Biologia Molecular

# Search for new antiviral compounds using fragment screening methodology

Zuzanna Kaczmarska



Barcelona, 2014





UNIVERSITAT DE BARCELONA



Thesis submitted by Zuzanna Kaczmarska, enrolled in the Biotechnology program at the University of Barcelona, for the degree of Doctor of Philosophy.

This work was carried out in the Structural Biology of Protein & Nucleic Acid Complexes and Molecular Machines Group at the Institute for Research in Biomedicine (IRB-Barcelona) and Molecular Biology Institute of Barcelona (IBMB-CSIC), under the supervision of Prof. Miquel Coll Capella.

**Zuzanna Kaczmarska**  
IRB Barcelona  
IBMB-CSIC

**Miquel Coll Capella**  
IRB Barcelona  
IBMB-CSIC

Barcelona, 2014



## Preface

*Picornaviridae* are among the most diverse and oldest known viral families that include many important pathogens of humans and animals. They are small, icosahedral (+)ssRNA viruses, causing a variety of diseases, such as encephalitis, and poliomyelitis. Vaccines are available for poliovirus, hepatitis A virus and foot and mouth disease virus, but no effective prophylaxis is implemented for other picornaviruses. Thus far, anti-viral research has focused on the capsid, whereas inhibitors targeting non-structural proteins (i.e. proteases, helicases, polymerases) have remained largely unaddressed.

The project was focused on structural and biochemical characterization of the enterovirus-B93 (EV-B93) 3C protease alone and in complex with several covalent inhibitors. The second objective was to identify the first non-covalent potent inhibitors of the EV-B93 3C protease and their further biochemical, antiviral, and structural evaluation.

## List of abbreviations and symbols

#	number
°C	degree Celsius
ΔG	Gibbs energy
[v/v]	volume per volume
[w/v]	weight per volume
1,8-ANS	1-anilinonaphthalene- 8-sulfonic acid
2,6-TNS	2-p-toluidinyl-6-naphthalene sulfonate
3D	three dimensions
3CL protease	3C-like protease
5'UTR	5' untranslated region
Å	Angstrom
ADMET	absorption, distribution, metabolism, secretion, toxicity
Ac	acetyl
AiV	aichi virus
APS	ammonium persulfate
ATP	adenosine-5'-triphosphate
C-terminal	carboxy-terminal
Cbz	carboxybenzyl
CDK2	cyclin-dependent kinase 2
clogP	calculated logP
CoV	coronavirus
CPE	cytopathic effect
CV	column volume, coxsackievirus
Da	Dalton
DMAB	dimethylaminoborane
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
DTT	dithiothreitol
EC50	half maximal effective concentration
EDC	1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide
ERAV	equine rhinitis A virus
ERBV	equine rhinitis B virus
EV	enterovirus
FABP4	fatty acid binding protein 4
FBDD	fragment-based drug design
FDA	Food and Drug Administration
FMDV	Foot-and-mouth disease
g	gram
GAL4BD	GAL4 binding domain
h	hour

HAV	hepatitis A virus
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His-tag	poly(6) histidine-tag
HPLC	high performance/pressure liquid chromatography
HRV	human rhinovirus
HSQC	heteronuclear single quantum correlation
HTS	high-throughput screening
Hz	hertz
IC50	half maximal inhibitory concentration
IMAC	immobilized metal ion affinity chromatography
IRES	internal ribosomal entry
k	kilo ( $10^3$ )
Kd	dissociation constant
Ki	inhibitory constant
$k_{off}$	dissociation rate constant
l	liter
LB	Luria-Bertani
LE	ligand efficiency
$\mu$	micro ( $10^{-6}$ )
M	molar, marker
m	mili ( $10^{-3}$ )
m/z	mass-to-charge ratio
MALDI-TOF	matrix-assisted laser desorption/ionization time-of-flight
MD	molecular dynamics
min	minute
MPD	2-methyl-2,4-pentanediol
MS	mass spectrometry
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
MW	molecular weight
MWCO	molecular weight cut-off
n	nano ( $10^{-9}$ )
N-terminal	amino-terminal
Ni-NTA	nickel-nitrilotriacetic acid
NHS	N-hydroxysuccinimide
NMR	nuclear magnetic resonance
NROT	number of rotational bonds
P	partition coefficient between octanol and water
PAGE	polyacrylamide gel
PDB	protein data bank
PCR	polymerase chain reaction
PEG	polyethylene glycol
ppm	parts per million
PSA	polar surface area
PSV	porcine sapelovirus



PV	poliovirus
RA	relative activity
RLU	relative light unit
RMSD	root mean square deviation
RNA	ribonucleic acid
RPC	reverse phase chromatography
rpm	revolutions per minute
RU	response unit
RV	rhinovirus
s	second
S/N	signal-to-noise ratio
SAR	structure-activity relationship
SARS	severe acute respiratory syndrome
SDS	sodium dodecyl sulfate
SPR	surface plasmon resonance
(+)ssRNA	positive single-stranded RNA
STD	saturation transfer difference
TEMED	N,N,N,N-tetramethylethylenediamine
TFA	trifluoroacetic acid
TINS	target immobilized NMR screening
T <sub>m</sub>	melting temperature
Tris	Tris(hydroxymethyl)aminomethane
TSA	thermal shift assay
V	Volt
VP16AD	VP16 activation domain

## Amino acids abbreviations

1-letter code	3-letter code	amino acid
A	Ala	alanine
R	Arg	arginine
N	Asn	asparagine
D	Asp	aspartic acid
C	Cys	cysteine
E	Glu	glutamic acid
Q	Gln	glutamine
G	Gly	glycine
H	His	histidine
I	Ile	isoleucine
L	Leu	leucine
K	Lys	lysine
M	Met	methionine
F	Phe	phenylalanine
P	Pro	proline
S	Ser	serine
T	Thr	threonine
W	Try	tryptophan
Y	Tyr	tyrosine
V	Val	valine

## Table of contents

Preface .....	5
List of abbreviations and symbols.....	6
Amino acids abbreviations .....	9
Table of contents .....	10
1. Introduction .....	15
1.1 Drug discovery.....	17
1.1.1 Fragment-based drug discovery vs. high-throughput screening: a matter of chemical space.....	17
1.1.2 What is a fragment?.....	19
1.1.3 The drug-like molecule.....	19
1.1.4 Hit selection and lead optimization .....	20
1.1.5 Detection of fragment binding .....	21
1.1.5.1 Nuclear Magnetic Resonance.....	22
1.1.5.2 Surface plasmon resonance.....	22
1.1.5.3 X-ray crystallography .....	23
1.1.5.4 Fluorescence melting assay.....	23
1.1.5.5 Other methods.....	24
1.1.6 Elaboration of fragment hits .....	24
1.2 Enteroviruses as a part of <i>Picornaviridae</i> family .....	26
1.2.1 Viral proteins.....	27
1.2.2 3C protease .....	29
1.2.3 Drugs targeting the 3C protease .....	30
1.2.3.1 Peptidic inhibitors .....	31
1.2.3.1.1 Peptide aldehydes and ketones .....	31
1.2.3.1.2 Michael-acceptor-based inhibitors.....	32
1.2.3.1.3 Other peptidic inhibitors.....	35
1.2.3.2 Non-peptidic inhibitors .....	35
1.2.3.3 Non-covalent inhibitors.....	39

2. Objectives .....	41
3. Materials and methods.....	45
3.1 Materials .....	47
3.1.1 Expression vector.....	47
3.1.2 Plasmid purification .....	47
3.1.3 Growth media .....	47
3.1.4 Bacterial cloning and expression strains .....	48
3.1.5 Transformation of bacterial strains .....	48
3.1.6 Protein electrophoresis.....	49
3.1.7 Chromatography columns .....	49
3.1.8 Protein concentration .....	50
3.1.9 Mass spectrometry.....	50
3.1.10 Protein crystallization .....	50
3.1.11 X-ray data collection, processing and structure determination .....	52
3.1.12 Saturation transfer difference NMR screening.....	52
3.1.13 Thermal shift assay .....	52
3.1.14 Surface plasmon resonance experiment.....	53
3.1.15 Cell-based proteolytic assay and antiviral assay .....	53
3.1.16 Analogs selection, docking and classical molecular dynamics simulations.....	53
3.2 Methods.....	54
3.2.1 Construct design .....	54
3.2.2 Plasmid transformation .....	54
3.2.3 Protein expression.....	54
3.2.4 Protein purification .....	55
3.2.5 Reductive methylation.....	55
3.2.6 <i>In vitro</i> proteolytic activity assay .....	56
3.2.8 Crystallization .....	56
3.2.8.1 Protein sample preparation .....	56
3.2.8.2 Protein-covalent compound complex preparation .....	57
3.2.8.3 Protein-non-covalent compound complex preparation.....	57

3.2.8.4 Crystallization screening and optimization.....	57
3.2.8.5 Soaking experiment .....	59
3.2.8.6 Mounting and freezing crystals .....	59
3.2.9 Saturation transfer difference NMR.....	59
3.2.10 Thermal shift assay .....	60
3.2.11 Surface plasmon resonance .....	60
4. Results and discussion.....	61
4.1 Structural and functional studies of EV-B93 3C protease.....	63
4.1.1 Expression and purification of native EV-B93 3C protease.....	63
4.1.2 Biochemical characterization of native EV-B93 3C protease and its C147A mutant .....	64
4.1.3 Crystallization screening and crystal optimization of native EV-B93 3C protease ..	66
4.1.4 Structure determination of EV-B93 3C protease and data analysis .....	68
4.2 Rupintrivir and compound 1 as covalent inhibitors of EV-B93 3C protease .....	72
4.2.1 Biochemical characterization of rupintrivir and compound 1 inhibitory activity....	72
4.2.2 Crystallization screening and crystal optimization of EV-B93 3C protease in a complex with rupintrivir and compound 1 .....	74
4.2.3 Structure determination of EV-B93 3C protease in complex with rupintrivir and compound 1 and data analysis .....	75
4.3 Small covalent inhibitors of EV-B93 3C protease as a starting point for development of potent drugs.....	81
4.3.1 Biochemical characterization of NZN, NZO and DB5_60 inhibitory activity.....	81
4.3.2 Crystallization screening and crystal optimization of EV-B93 3C protease in complex with NZO, NZN and DB5_60.....	85
4.3.3 Structure determination of EV B93-3C protease in a complex with NZN.....	86
4.4 Discovery of non-covalent inhibitors of EV-B93 3C protease using fragment-screening methodology .....	92
4.4.1 Library design.....	93
4.4.2. STD NMR-based fragment screening.....	93
4.4.3 Biophysical validation of the selected hits – TSA .....	96
4.4.4 Biophysical validation of the selected hits – SPR.....	98

4.4.5 <i>In vitro</i> proteolytic assay .....	102
4.4.6 Elaboration of compound 241 .....	103
4.4.7 Chemical optimization of the best compound A4.....	116
4.4.8 Co-crystallization/soaking trials .....	121
4.4.9 Homologous 3C proteases as targets for the best compounds .....	123
4.4.10 Cell-based assays with the best compounds.....	133
4.4.10.1 Cell-based proteolytic assay with the best compounds .....	133
4.4.10.2 Cytopathic effect (CPE)-based antiviral assay .....	137
4.4.11 Bioinformatics simulations of A4_D binding mode.....	140
5. Conclusions .....	147
6. Bibliography .....	151



# 1. Introduction





## 1.1 Drug discovery

Drug discovery is the fascinating process through which potential new medicines are found. In the past, active ingredients were identified from traditional remedies or by unanticipated discovery. Later intact cells or whole organisms were treated by series of synthetic compounds, natural products or extracts in order to select substances with desirable therapeutic properties. The sequencing of the human genome enabled rapid cloning, expression and purification of a large number of proteins. Since then, high-throughput screening of large quantities of molecules against isolated biological targets has become a common practice. The selected hits from the screens are subsequently tested in cellular and animal models for efficacy. Most frequently, optimization of the most promising compounds is needed to improve their affinity, selectivity (reduction of possible side effects), potency, metabolic stability (increase of half-life), and oral bioavailability. Once a compound fulfills the requirements established for a drug candidate, it will undergo the process of drug development prior to clinical trials [1].

### 1.1.1 Fragment-based drug discovery vs. high-throughput screening: a matter of chemical space

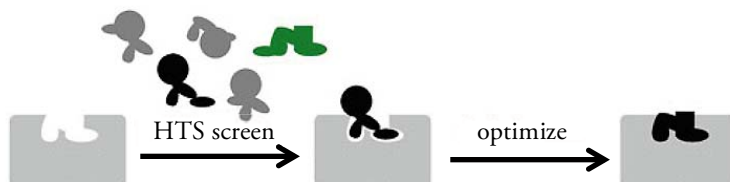
Drug discovery methods are divided into several techniques described below (Figure 1.1). Fragment-based drug discovery (FBDD) emerged as a novel strategy for hit identification in the past decade. The power of the FBDD-concept was first described by Fesik, Hajduk and colleagues in the late 1990s [2], [3]. Since then, the field has vastly expanded, and the fragment-based approach has started playing an important role in the discovery of drugs, many of which have now entered the clinic. Nowadays, this method is widely applied by scientists working in both academia and the industry [4].

The FBDD approach is recognized as a highly efficient tool for drug discovery, and offers an alternative to more traditional methods such as high-throughput screening (HTS). Despite the fact that HTS has taken hold in most pharmaceutical companies, and allows significantly a larger number of compounds to be screened against a target of interest, thus far the industry has not succeeded in generating the desired level of productivity. Undoubtedly, HTS has produced high quality lead molecules for many drug discovery programs, however, the hit rates are generally low, particularly when screening against challenging targets [5], [6]. In the case of fragment screening, the fragments exhibit lower affinities than HTS hits, but simultaneously show higher ligand efficiencies [7], thus proving that with these small molecules a larger proportion of the fragment is truly involved in the binding. In other words, smaller, less

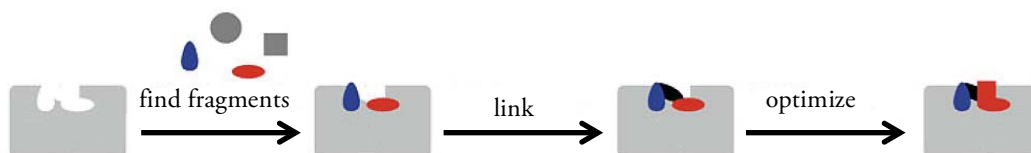
complex molecules have fewer possibilities to bind to the target, hence this leads to a more effective probing of the available chemical space.

Additionally, the optimization of lead compounds in HTS can be problematic owing to the often large and relatively lipophilic nature of the screening hits [8]. In contrast, the identification of a starting compound is guaranteed in the application of fragment screening that identifies smaller soluble ligands even in the most difficult places, such as protein-protein interaction regions. According to a theoretical model proposed by the Novartis Group, the observed hit rates in low molecular weight screenings were 10- to 1,000-fold higher than in conventional HTS assays [9]. Despite the fact that the size of a fragment library is typically smaller (500-10,000 compounds) than an HTS library (up to  $10^6$  molecules), it is easier to obtain a more structurally diverse collection of low molecular weight compounds that can cover the chemical ligand space more effectively. Therefore, the hit identification for most targets is often faster and allows an effective subsequent hit-to-lead elaboration (Figure 1.1).

#### Traditional HTS



#### Fragment-based screening, linking fragments



#### Fragment-based screening, growing fragments

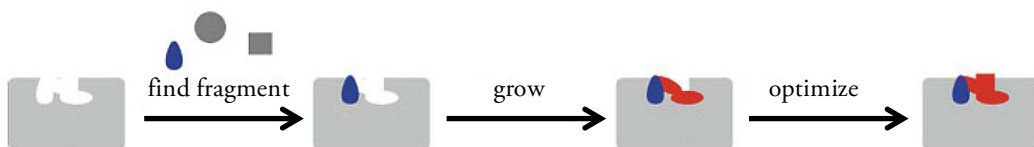


Figure 1.1 Comparison of high-throughput screening (HTS, *top*) with fragment-based screening (linking and growing strategy, *middle* and *bottom*). Adapted from [10].

### 1.1.2 What is a fragment?

Fragments used in FBDD are defined as low molecular weight ( $MW < 300$  Da), moderately lipophilic (calculated  $\log P \leq 3$ , where  $P$  is the partition coefficient between octanol and water), highly soluble organic molecules, which typically bind to their target with low affinity (in the  $\mu M$  to  $mM$  range). Also, the number of hydrogen-bond donors and acceptors should be  $\leq 3$ . Some data suggested the amount of rotational bonds (NROT)  $\leq 3$  and polar surface area (PSA) should not exceed  $60 \text{ \AA}^2$ . These general principles, known as “the rule of three” [11], could be helpful in designing fragment libraries for an efficient discovery of hits. Subsequently, they can be further optimized into drug-like molecules with good physicochemical properties compliant with Lipinski’s “rule of five” [12].

### 1.1.3 The drug-like molecule

Lipinski suggested the following guiding criteria for orally available small-molecule drugs:  $MW < 500$  Da, calculated  $\log P < 5$ , number of hydrogen bonds  $\leq 5$ , number of donors and acceptors  $\leq 10$ . This concept of drug-likeness is not always obeyed by natural products or natural-product like molecules (e.g. cyclosporine). Nevertheless, they may be easily absorbed after oral administration. The Lipinski’s rules were enhanced by Clark and Picket who discovered that the polar surface area (PSA) is another important property for a compound to exhibit high oral bioavailability [13]. Consequently, more polar, less lipophilic compounds with a PSA of  $110\text{--}140 \text{ \AA}^2$  can be absorbed more efficiently and display low toxicity [14]. Another observation was described by Veber and colleagues who noticed that the maximum number of rotatable bonds for optimal oral bioavailability is seven [15]. A potential drug candidate should not only fulfill the Lipinsky’s “rule of five”, but also possess appropriate ADMET properties (absorption, distribution, metabolism, secretion and toxicity). These parameters are exceptionally difficult to optimize in the case of large (ca. 500 Da) and often lipophilic compounds selected by HTS approach. About 50% of drugs in general fail in clinical trials due to inadequate ADMET attributes [16], [17], [18], [19], [20].

### 1.1.4 Hit selection and lead optimization

The concept of ligand efficiency (LE) introduced by Hopkins and colleagues [7] can be used for quality evaluation of the preliminary screening hits and further assessment of lead optimization. The ligand efficiency index (LE) is defined as:

$$LE = \frac{\Delta G_{binding}}{N_{non-hydrogen\ atoms}}$$

where  $\Delta G_{binding}$  is the free binding energy (Gibbs energy) of the ligand for the protein of interest while  $N_{non-hydrogen\ atoms}$  is the number of heavy (non-hydrogen) atoms in the ligand. Free binding energy is defined by:

$$\Delta G_{binding} = -RT \ln K_d$$

where  $R$  is the gas constant,  $T$ - the absolute temperature and  $K_d$  the dissociation constant. It is possible to describe LE more practically using parameters that can be measured directly:

$$LE = \frac{pK_i/pK_d/pIC_{50}}{N_{non-hydrogen\ atoms}}$$

The parameters:  $pK_i$ ,  $pK_d$ ,  $pIC_{50}$  as well as  $\Delta G_{binding}$  can define potency equally well. LE is a useful method to measure the impact of addition of more atoms to a molecule on its potency. Molecules that reach a desired potency but possess fewer heavy atoms are defined as more efficient. In successful optimization, potency increases linearly with molecular weight. For this reason, it is crucial to select the initial hit with the most optimal efficiency for further lead elaboration.

Figure 1.2 shows the general relationship of the potencies/efficacies and the molecular weight of selected HTS hits and fragments [21]. Fragment hits are usually in the range between 120-250 Da and of low potency (mM to 30  $\mu$ M). The potency of selected fragment needs to be increased in order to acquire a useful lead, nearly always resulting in a higher molecular weight. Hits coming from the HTS approach have a wider range of molecular weight (250-600 Da), and tend to be of higher potency (high nM range to low  $\mu$ M). These molecules have to be optimized in order to obtain a good quality lead of the same or higher potency and a reduced molecular weight. Years of experience show this is a challenging goal to achieve.

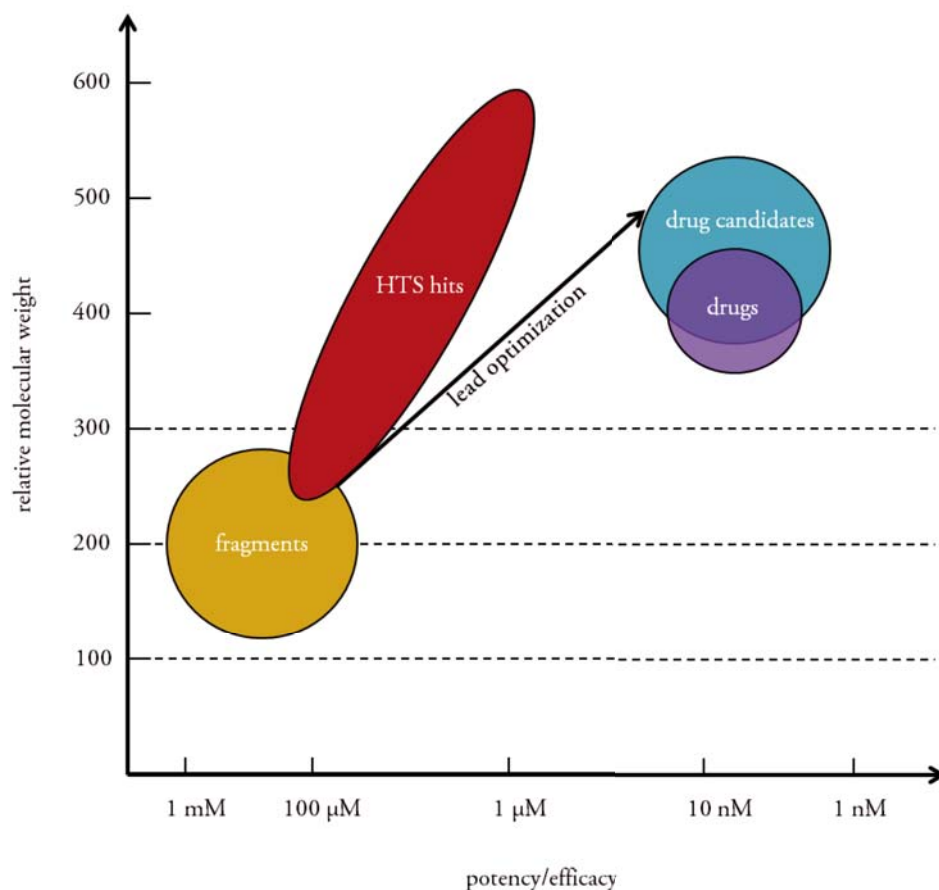


Figure 1.2 General relationship of potencies/efficacies and molecular weight of selected HTS hits and fragments as starting points for lead identification and drug discovery.

### 1.1.5 Detection of fragment binding

One of the major challenges of fragment-based drug discovery is an effective and reliable detection of low-affinity binders. Since fragments possess millimolar or high micromolar affinity to the target of interest, they need to be screened at a higher concentration using a very sensitive method. Typically, most biochemical and binding assays fail at this stage; however they are the simplest and cheapest approaches available. Performing a biochemical assay requires the use of highly concentrated compounds, typically 250-1,000 mM, and may cause various problems including target denaturation, toxicity to cells, false positives (due to compound aggregation) and false negatives (due to lack of solubility). For all these reasons, a variety of alternative biophysical methods have been developed to improve the selection of fragment hits.

### 1.1.5.1 Nuclear Magnetic Resonance

The first published description of Fragment-Based Lead Discovery (FBLD) was the Structure-Activity Relationship (SAR) by Nuclear Magnetic Resonance (NMR) method by the Abbott Group. They detected binding of ligands to specific regions by the perturbation of HSQC (Heteronuclear Single Quantum Correlation) spectrum of an isotopically-labeled protein [2, 22]. This technique relies on the changes in chemical shifts of specific protein residues in the absence and presence of small compounds. However, it requires large quantities of the protein (exceeding 200 mg), and is limited to relatively small proteins (less than 30-40 kDa) [2].

Another NMR-based method for fragment screening is the so-called ligand-detected approach, in which changes in the NMR properties of the fragments are monitored, rather than properties of the protein itself. There are several techniques in use [23]. The most common is saturation transfer difference (STD) [24], which is based on the differences in relaxation between small compounds and the target protein. This method uses smaller amounts of protein, require no isotope-labeling, and has no upper size limit for the protein target. Moreover, STD NMR is an excellent approach for determining the binding epitope of a ligand-the piece of information vital for directed drug development.

### 1.1.5.2 Surface plasmon resonance

Surface plasmon resonance (SPR) is another ligand-screening technique that has gained importance in recent years. This label-free study of interactions between molecules has been known for over 20 years [25]. Typically, a protein is immobilized on a chip, and varying concentrations of ligands are allowed to flow over it. Molecules that bind to the protein cause changes in the reflective properties that depend on the ratio of the ligand mass to the protein mass. In some cases, association and dissociation rates can be directly determined, however, in the case of fragments these are usually too rapid to be measured. Although SPR has been available for some time, the recent improvement in sensitivity and increased experience in robust attachment of a target to the chip surface have led to an extended use of this technique. It is important to note that although SPR experiments do not require extensive training, it is very easy to set up the method incorrectly or to be misled by the artifacts. When done properly, SPR can be a very useful tool, not only for obtaining dissociation constants, but also for stoichiometry estimation [26-28].

### 1.1.5.3 X-ray crystallography

Crystallography and protein-detected NMR are unique methods that provide detailed information on how (including low-affinity) fragments bind to protein targets. X-ray crystallography has the advantage over NMR of yielding high-resolution structural data, and it can be employed for large proteins. Nowadays, most companies use crystallography as a primary screening technique in FBDD campaigns. However, it may also be applied for characterization or confirmation of the binding mode of previously selected hits, thus providing valuable insight into SAR of the fragments and potential areas for a future development [10, 29-32]. It is worth remembering X-ray crystallography has also numerous limitations. Generation of a suitable protein in high yield requires an effective expression and purification of the target protein. This process is time-consuming and labor-demanding. Reproducible production of well-diffracting crystals and its optimization is another important step. There are two options to obtain co-crystal structures with compounds of interest: soaking the fragment hits into already-existing crystals and co-crystallization. The collected data result in a 3D model that may be useful although also misleading. In the case of low resolution structures, it is possible to misassign the position and conformation of a ligand. Frequently, the potential hits are not noticed (false negatives) due to occlusion of the binding sites by crystal contacts, or due to ligand binding requiring a conformational change of the protein that is not possible within the crystalline state. Finally, a crystal structure provides no or limited information on the binding affinity of the ligand, and requires the use of additional methods to confirm its presumed functional activity.

### 1.1.5.4 Fluorescence melting assay

The thermal shift (Thermofluor) method is another biophysical technique widely used as a primary screening. It is based on monitoring the changes in temperature at which the protein of interest unfolds in the presence of a putative ligand. This is possible when Sypro Orange, a dye that preferentially binds to unfolded hydrophobic parts of a denatured protein increasing its fluorescence, is added to the protein-ligand mixture. Through measuring the fluorescence emission over a range of temperatures, it is possible to determine the melting temperature ( $T_m$ ) of a protein. Ligands usually stabilize their hosts against thermal denaturation, so a molecule that increases a protein's melting temperature is considered a hit [33]. In spite of being fast and inexpensive, the thermal shift assay (TSA) is believed to be highly unreliable, and not all proteins are suited to this approach. Low-affinity molecules may not lead to



a strong stabilization of a protein and hence to a non-detectable shift in protein melting temperature, that results in a high false-negative hit rate [34].

#### 1.1.5.5 Other methods

Despite the unquestionable success of the aforementioned methods, continued efforts are being made to develop improved techniques for detection and characterization of fragment binding. The MS-based analysis deserves special attention, where low affinity ligands can be identified by measuring the abundance of the “free” ligand in the absence and presence of the target or by analyzing the non-covalent protein-fragment complex in gas phase. Another particularly interesting ligand-detecting NMR-based method is the TINS (Target Immobilized NMR Screening) approach when a target is immobilized onto a resin [35]. The advantage of this method is that a low amount of protein is required, and it may be applied for membrane proteins, which are generally challenging in NMR. Calorimetric methods [36] may gain importance as a primary screening in fragment-based approach due to their higher sensitivity and throughput. The recent advances in calorimetry allows a direct measurement of enthalpy and entropy contributions to the binding energy, and may become an important factor in the selection of fragments to progress [37].

#### 1.1.6 Elaboration of fragment hits

The confirmed fragment hits can be readily optimized into potent leads by iterative process of rational design, synthesis and/or exploration of commercially available compounds that pick up additional target-ligand interactions. This can result in an improved affinity while still maintaining the drug-like properties. The process is guided by empirical information based on 3D X-ray co-structures and binding affinity measurements. The development of fragment hits can follow many paths. The most important ones are the “Fragment Growth” and “Fragment Linking” strategies [29]. In “Fragment Growth”, high resolution crystallographic or NMR data is needed for the identification of close contacts between the hit and the protein of interest. A comparison of the binding modes and affinities is then used to infer the requirements for association and suggest novel fragments which may satisfy these. In parallel, the possibility of adding functional groups to the fragment is also examined. These modifications are directed towards increasing the binding affinity but at a certain point other considerations are taken into account such as selectivity, stability and physical properties, which are important for effective drug design. Fragment-growth strategies have proven particularly efficient when

targeting inner active site pockets, such as ATP-binding site of kinases, and allow a careful assessment of the effect of each new functional group or atom, thus maintaining optimal ligand efficiency. An example of hit improvement using this strategy is compound AT7519 developed by Astex [38]. It is a potent cyclin-dependent kinase 2 (CDK2) inhibitor that was successful in phase I clinical trials for patients with solid tumors, and has since then entered phase II studies. The compound elaboration started from a weak binder 1H-indazole with  $IC_{50} = 185 \mu\text{M}$  and  $LE = 0.57$ , and resulted in the much more potent compound AT7519 with  $IC_{50} = 0.03 \mu\text{M}$  and similar ligand efficiency  $LE = 0.40$  (Figure 1.3).

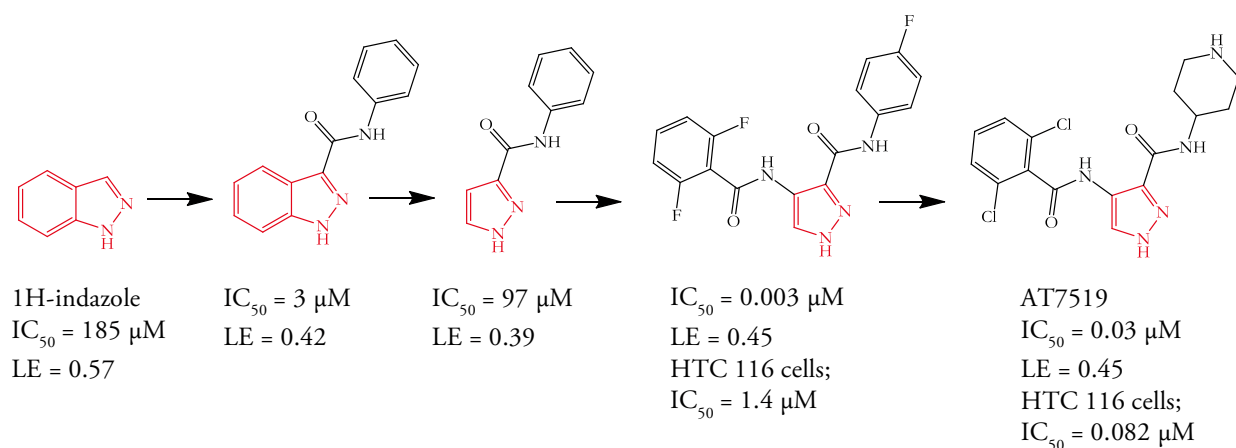


Figure 1.3 Evolution of weak fragment hit to potent drug for CDK2 as an example of “Fragment-Growing” strategy. Adapted from [10].

In the alternative “Fragment-Linking” strategy, additional screening is required in order to find a second fragment that can bind simultaneously with one or several fragments identified previously during the primary screening. In the following step, two concurrently binding molecules are linked in several different ways, using flexible linkers of variable length (Figure 1.4). In binding sites composed of multiple sub-pockets, “Fragment-Linking” strategy can lead to a rapid increase in ligand affinity and eventually to a potent therapeutic agent. Proteases are ideal candidates for the “Fragment-Linking” approach. They cleave peptides or proteins, thus their active sites are extended with each amino acid binding in its own sub-site (S). An elegant example of this strategy was published by Astex [39]. A crystallography-based fragment screening technique was used to develop the inhibitor for the protease thrombin. Two weak inhibitors were identified as binding in adjacent regions, one in P1 pocket and the other in the P2-P4 region. High-affinity inhibitor resulted in the linkage of these two molecules, and crystallographic studies showed the final compound has the same binding mode as the separate fragments.

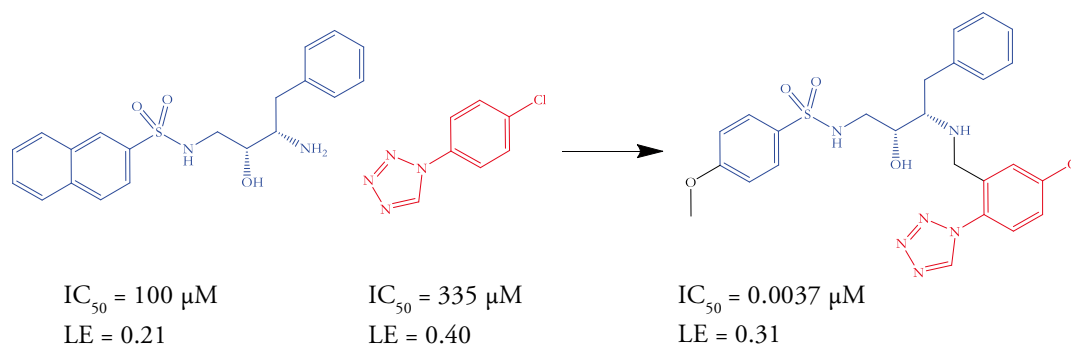


Figure 1.4 Development of the inhibitor for thrombin protease as an example of “Fragment-Linking” strategy. Adapted from [40].

Both the “Fragment-Growth” and “Fragment-Linking” strategies are valuable. “Fragment-Growth” gives the possibility to modify the molecule in different positions and have greater control over the ligand size and binding efficiencies. As mentioned before, this strategy is generally preferred for compact active sites, which are capable of highly efficient protein-ligand interactions. In contrast, the “Fragment-Linking” strategy is more suited to less ligand-efficient and more solvent-exposed sites, which contain several sub-pockets. Although the linking strategy appears more elegant, the limited scope of available linkers may hinder the binding the original fragments. In comparison, the fragment-growing strategy provides more freedom for introducing chemical modifications and further optimization.

## 1.2 Enteroviruses as a part of *Picornaviridae* family

Enteroviruses (EVs) are part of the *Picornaviridae* family, one of the oldest and most diversified families of viral pathogens of vertebrates, including humans [41]. They are small, non-enveloped, positive-sense, single-stranded RNA viruses with an icosahedral capsid. The *Enterovirus* genus is divided into 12 species where 7 of them contain human-infecting viruses: EV-A, EV-B, EV-C, and EV-D, including three *Rhinovirus* species: RV-A, RV-B, and RV-C. The echovirus and coxsackievirus B (CV-B) are classified within the EV-B species and polioviruses (PVs) within EV-C species. EVs that infect primates, cattle and swine have also been described and may lead to zoonotic spread and the emergence of new human pathogens. The *Enterovirus* genus is remarkably heterogenic, with more than 250 different serotypes. Despite the fact that enteroviruses and rhinoviruses share numerous genetic features, rhinoviruses are restricted to the respiratory tract, whereas the vast majority of enteroviruses

cause gastrointestinal infections, and can spread to other organs such as the heart and the central nervous system. Infections caused by EVs in humans may range from asymptomatic to more severe illnesses manifested as encephalitis, aseptic meningitis, myocarditis, gastroenteritis, paralysis, poliomyelitis, with high mortality rates in infected newborn infants [42]. There are over five billion cases of EV infections annually worldwide and several mass outbreaks in Australia, United Kingdom and China have been reported [43], [44], [45]. For instance, EV-A71 was responsible for serious complications such as encephalitis, myocarditis and deaths during the epidemics of hand, foot, and mouth disease in Asia in 1997, 1998, 2000, and 2008 [46]. The disease manifestation of acute flaccid paralysis is associated with the nonpoliovirus EVs within the EV-B species, including the newly discovered viruses like EV-B93 [47], which is the subject of this thesis.

### 1.2.1 Viral proteins

The genome of EVs is a positive-sense, single-stranded RNA and varies between 7.4 and 7.5 kb with a single, large open reading frame that encodes a long polyprotein of approximately 2,200 amino acids (~250 kDa) [48], [49]. The organization of the polyprotein follows a common pattern that is shared among all *Picornavirus*. It contains the structural proteins of the capsid, followed by the non-structural proteins that are involved in replication. Co- and posttranslational cleavage of the polyprotein occurs resulting in three precursor molecules, P1, P2 and P3. These are further processed releasing mature viral proteins: the structural proteins VP4 to VP1 from P1 and the non-structural proteins associated with replication, 2A to 2C and 3A to 3D from P2 and P3, respectively [50] (Figure 1.5).

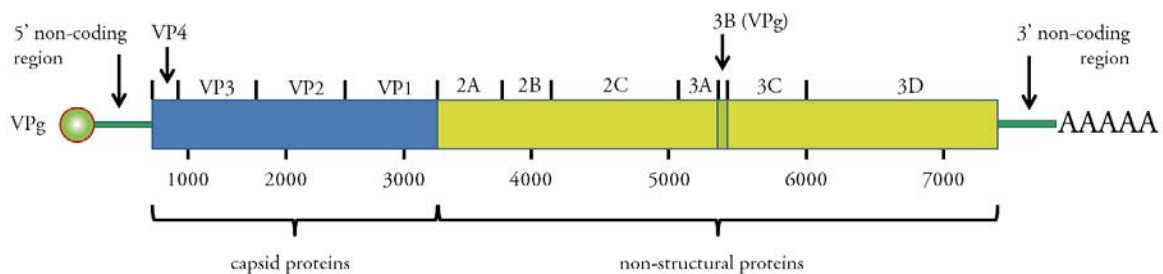


Figure 1.5 The organization of the enteroviral polyprotein.

The cleavages are mediated by the viral proteases 2A and 3C/3CD protease. The 2A protease cleaves the bond between the P1 and P2 segments of the polyprotein, whereas the 3C protease and its precursor 3CD, are responsible for generating the majority of precursors and mature proteins [41]. The 5' end of the genome is covalently linked to the small viral protein VPg [51], [52] that contains a long 5' untranslated region (5'UTR), a 5' terminal domain involved in replication and an internal ribosomal entry site (IRES) necessary for cap-independent translation. The typical cycle of picornavirus infection involves attachment to the specific receptor on the cell membrane, internalization, uncoating and releasing of RNA genome in the cytoplasm where it is translated by host machinery. The resulting polyprotein is then processed, releasing structural and non-structural proteins. The genome is replicated in membrane vesicles by the RNA-dependent RNA polymerase 3D, creating a negative strand intermediate template, which is then used for positive strand replication through the same mechanism. The positive RNA genome is packed into the assembled capsid to form provirion and released upon maturation by cell lysis (Figure 1.6).

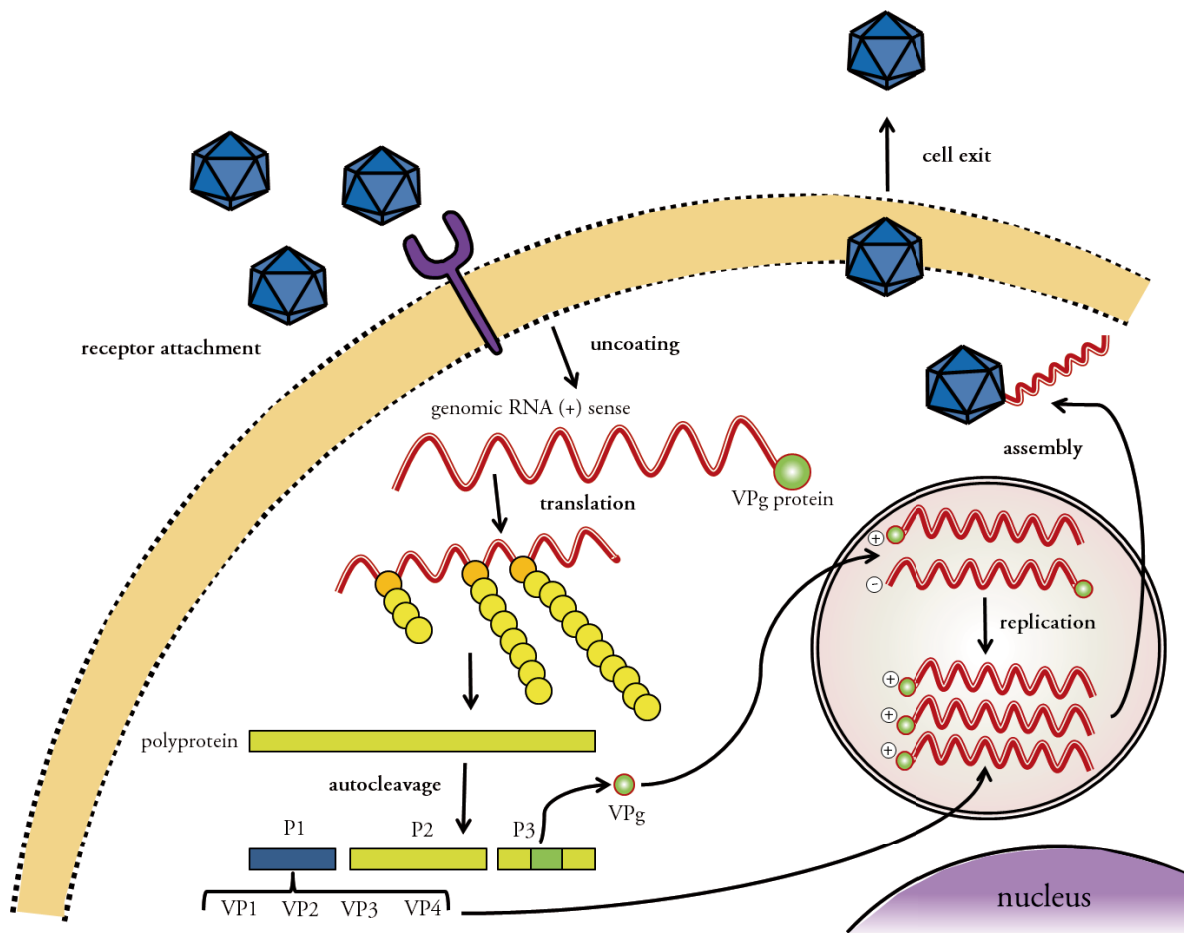


Figure 1.6 The enteroviral infection cycle. Adapted from [53].

### 1.2.2 3C protease

A typical enteroviral 3C protease is a small monomeric protein with molecular weight of around 20 kDa. It is considered to be the main protease due to its key role in polyprotein processing. In addition, the 3C protease cleaves various host proteins to alter essential cellular functions and facilitate virus replication [54-59]. Apart from its proteolytic activity, the 3C protease possesses RNA and DNA binding activities [60-64], which confirm its multitasking activity. Structural studies of several enteroviral (EV-A71, HRV-A2, HRV-A14, PV-1, CV-B3) 3C proteases [64-67] have revealed an overall fold that resembles an architecture of chymotrypsin-like serine proteases. The proteolytic active site is located in a shallow groove between two topologically equivalent six-stranded  $\beta$ -barrels. It contains Cys-147, His-40, and Glu-71 catalytic triad that is linked together by a hydrogen bond network. The His-40 residue with help of Glu-71 and Arg-39 prepares the Cys-147 side chain for nucleophilic attack on the substrate's carbonyl carbon by abstracting a proton. The role of Glu-71 is to provide electrostatic stabilization of the resulting positive charge on His-40, while Arg-39 neutralizes the charging of Glu-71 side chain. Substrate hydrolysis occurs through a covalent tetrahedral intermediate formed by the nucleophile from the active site (sulphur from Cys-147 side chain) and the carbonyl carbon of the scissile bond. The resulting oxyanion is stabilized by the hydrogen-bonding interactions with backbone NH groups of Cys-147, Gln-146, and Gly-145 amino acids, which are collectively called the oxyanion hole (Figure 1.7) [65, 67]. The 3C protease is able to process a limited number of cleavage sites. Most cleavages occur preferentially between the Q↓G pair at the P1 ↓ P1' position in peptides that consist of minimum 6 amino acids, and have at least 4 amino acids upstream from the scissile bond [68].

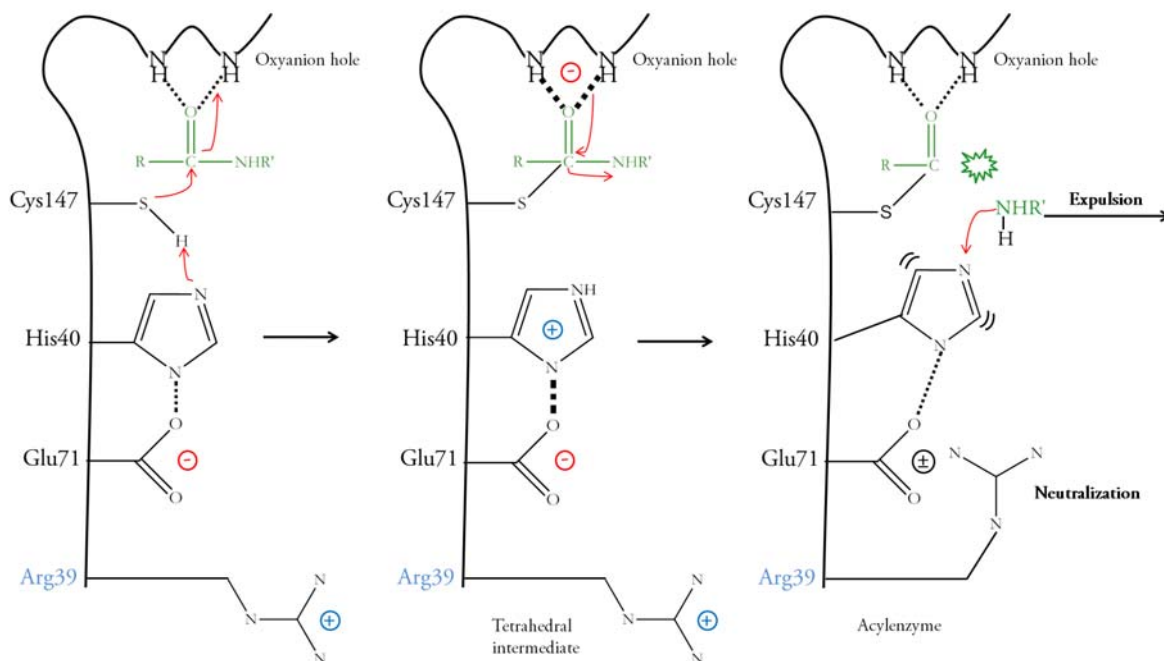


Figure 1.7 The mechanism of proteolysis of the EV93 3C protease. The catalytic triad (Cys-147, His-40 and Glu-71) is shown in black. The additional amino acid (Arg-39) helping during the proteolysis event is shown in blue. The substrate is shown in green. The electron relays are indicated by red arrows. Thick dashed lines indicate strong interactions, while thin dashed lines indicate weak interactions. Adapted from [65].

### 1.2.3 Drugs targeting the 3C protease

Enteroviral 3C proteases are highly conserved [69], and play a critical role in virus replication. They represent a unique class of enzymes with narrow substrate specificity, which makes them a promising target for antiviral therapy. Inhibitors of this enzyme are expected to possess a broad spectrum activity as well as high selectivity since no cellular homologue is known [70]. Inhibitors of proteases can be divided into two structurally distinct groups: peptidic inhibitors that mimic the substrate, and non-peptidic inhibitors. There are two known mechanisms of protease inhibition: reversible tight binding reactions and irreversible trapping reactions [71]. During the first mechanism, the reversible inhibitor binds to the active site of the enzyme, and competes with the natural peptide substrate. This inhibitor can dissociate from the enzyme in an unchanged form or after modification by the protease. Binding via the trapping mechanism leads to an irreversible protein-inhibitor reaction, which blocks the active site of the enzyme in a way that neither the inhibitor nor the protease can participate in further reactions.

### 1.2.3.1 Peptidic inhibitors

The first class of inhibitors was created based on substrate specificity. Modifications of chemical structures have been done with the aid of 3D structures of 3C proteases alone or in complex with potential inhibitors in order to get molecules with higher inhibitory and antiviral activity (i.e. SAR studies).

#### 1.2.3.1.1 Peptide aldehydes and ketones

Peptide aldehydes were one of the earliest and fairly potent inhibitors of cysteine and serine proteases. They bind as reversible adducts, in which the carbonyl carbon of the aldehyde forms a covalent bond with nucleophilic cysteine or serine resulting in stable tetrahedral species. Several peptidic aldehydes targeting enteroviral 3C proteases have been reported [72-76]. They were designed based on the knowledge that this enzyme recognizes and cuts P1-Q-P1'-G sequence with at least 4 amino acids upstream from the cleavage site, and possesses small non-polar amino acid in the position P4 (Figure 1.8). It was essential to replace the amide group in the side chain of the glutamine in the P1 position in order to prevent from intramolecular cyclization with reactive terminal aldehyde group, which results in the formation of hemiaminal [74] (compound 1a, 1b, 2 and 3, Figure 1.8). Following a similar course of thought, Dragovich and colleagues synthesized ketone-containing peptidic inhibitors [77], which are also not able to form hemiaminals (compound 4, Figure 1.8). Both peptide aldehydes and ketones exhibit low micromolar antiviral activity against several rhinovirus serotypes, low cytotoxicity and a reasonable therapeutic index.



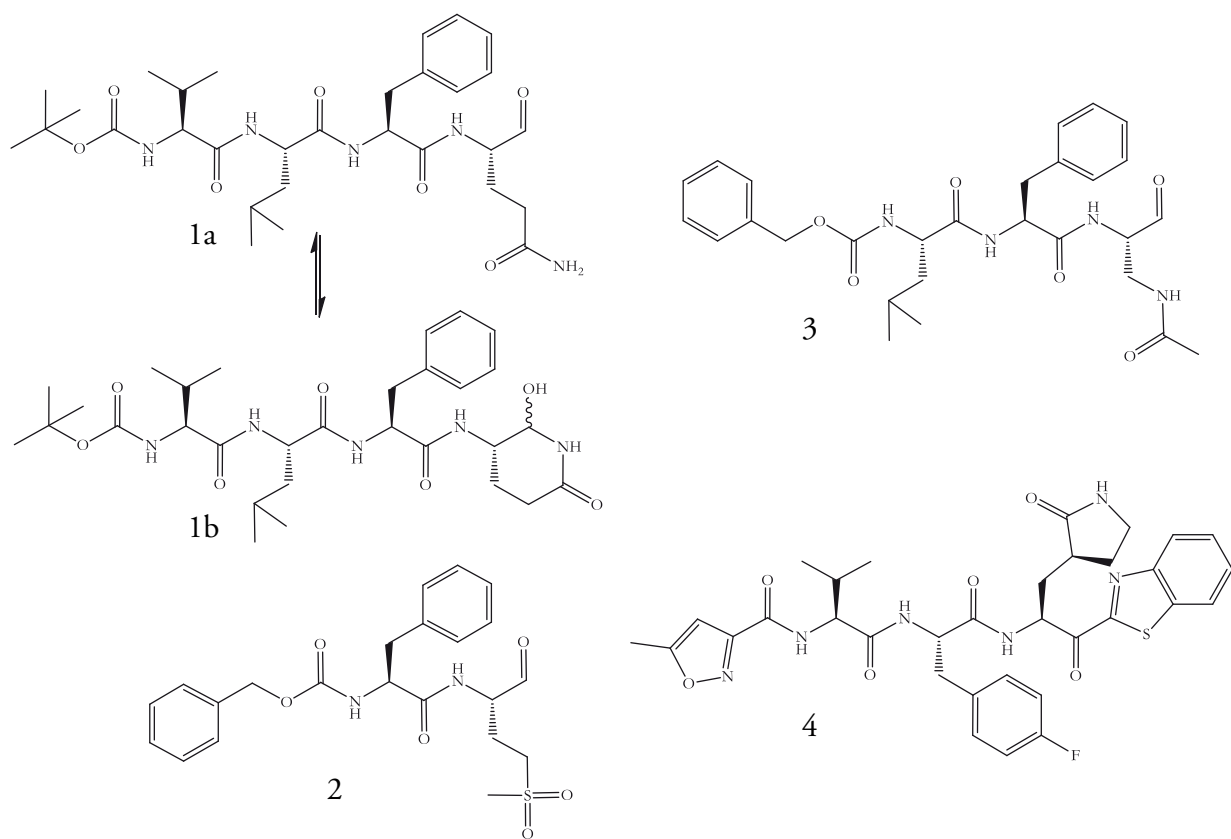


Figure 1.8 Peptide aldehydes and ketones as examples of reversible inhibitors of enteroviral 3C proteases.

### 1.2.3.1.2 Michael-acceptor-based inhibitors

More potent inhibitors against 3C proteases have been developed by introducing an electron-withdrawing group (Michael acceptor) that replaced the scissile bond between P1-Q-P1'-G amino acids in the peptide-like molecules. The Michael acceptor moiety is able to bind irreversibly to the cysteine in the active site, forming a stable covalent complex. Hanzlik and co-workers [78, 79] were first to report this kind of chemical group in peptidic inhibitors for cysteine proteases and also later for 3C proteases.

Additionally, a series of tripeptidic inhibitors with different Michael acceptor moieties were reported by scientist from Pfizer (formerly Agouron). A *trans*- $\alpha,\beta$ -unsaturated ethyl ester was

incorporated into Cbz protected tripeptide giving a non-toxic, potent irreversible inhibitor 5 (Figure 1.9) with moderate antiviral activity against three HRV serotypes in cell culture [80].

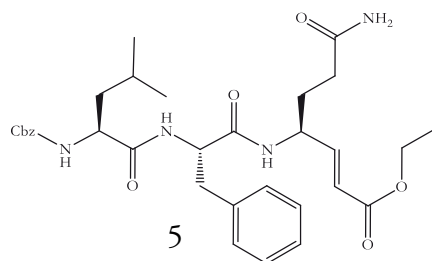


Figure 1.9 Cbz-protected tripeptide bearing  $\alpha,\beta$ -unsaturated ethyl ester moiety as an example of irreversible inhibitor of HRV 3C protease.

Various modifications of this molecule were studied in order to enhance its antiviral activity [80]. The results can be summarized in a few points:

- Related  $\alpha,\beta$ -unsaturated carboxylic acids and amides showed highly reduced inhibitory activity towards 3C protease, poor or no antiviral activity and increased toxicity.
- $\alpha,\beta$ -unsaturated esters substituted in position  $\alpha$  or *cis*- $\alpha,\beta$ -unsaturated esters displayed lower inhibitory effect when compared to the initial compound 5.
- Aliphatic and aryl  $\alpha,\beta$ -unsaturated ketones appeared to have an increased anti-3C activity but possessed reduced antiviral properties and were toxic to cells.
- Vinyl sulfones, nitriles, phosphonates, oximes, and vinyl heterocycles exhibited low levels of 3C protease inhibition and were weak antiviral agents.
- Michael acceptors possessing lactam, acyl oxazolidinone, and acyl urea functionalities demonstrated good inhibitory activity although they were inactivated by the exposure to non-enzymatic thiols.

Trans- $\alpha,\beta$ -unsaturated esters emerged as the most promising inhibitors against 3C proteases. Due to their significant inhibitory activity, high antiviral activity, stability in the presence of non-enzymatic thiols, low cellular toxicity, and ease of synthesis, they were chosen for optimization. As a result, Pfizer (formerly Agouron) managed to obtain an excellent irreversible inhibitor of 3C protease rupintrivir (AG7088) with a strong antiviral capacity against broad spectrum of picornaviruses and low toxicity [66, 77, 81-84].

### 1.2.3.1.2.1 Rupintrivir

Rupintrivir (AG7088), a peptidomimetic analog, contains an  $\alpha,\beta$ -unsaturated ester as a Michael acceptor moiety at P1' position, a five-member lactam ring mimicking Gln at the P1 position, a fluoro-phenylalanine and a methylene group instead of the backbone amide in the P2 position and an 5-methyl-3-isoxazole group at the P4 position (compound 6, Figure 1.10). Rupintrivir as a highly specific molecule against picornaviral 3C proteases demonstrated no activity towards cellular serine and cysteine proteases. It was shown that AG7088 is potent antiviral compound *in vitro* against all tested HRV and EV clinical isolates and it was selected for clinical trials [69, 84, 85]. Due to poor oral bioavailability, rupintrivir was administered intranasally [86], and appeared to be well tolerated, reduce the severity of illness and viral load in HRV challenge trials [87, 88]. However, in naturally infected patients, rupintrivir was not able to considerably affect the virus load or reduce the severity of disease, thus was halted from further clinical studies [89].

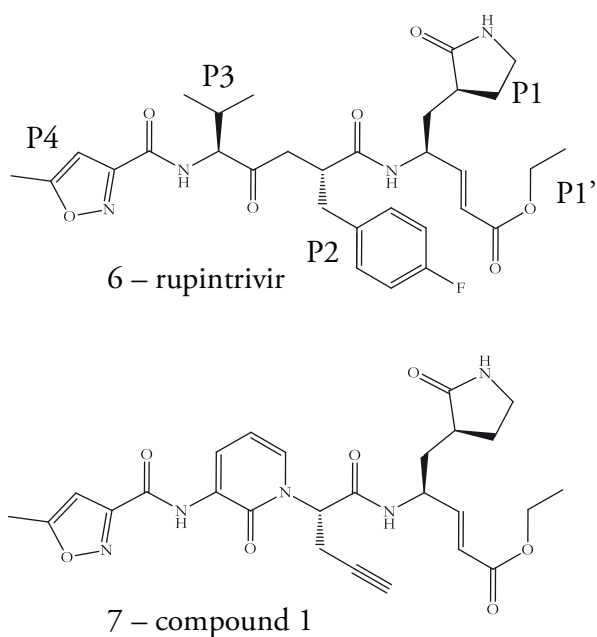


Figure 1.10 Rupintrivir and compound 1 – irreversible, highly-specific inhibitors against enteroviral 3C protease.

### 1.2.3.1.2.2 Compound 1

Since rupintrivir did not show the desirable therapeutic effect, several modifications were introduced to improve its efficacy. Compound 1 (AG7404), the orally bioavailable analogue of rupintrivir with similar mechanism of action was synthesized [89, 90] (compound 7, Figure 1.10). Compound 1 also displayed broad spectrum anti-HRV activity, and was safe and well-tolerated during phase I clinical trials. However, its subsequent clinical development of AG7404 was also discontinued [89].

### 1.2.3.1.3 Other peptidic inhibitors

Despite the fact that the most promising drug candidates: rupintrivir and compound 1 failed in clinical development,  $\alpha,\beta$ -unsaturated esters or terminal aldehydes still play a major role in enteroviral drug discovery. Many peptide-like molecules with Michael acceptor moiety were described in recent years as promising and potent drugs, including anti-coxsackievirus B3 [91-93], anti-EV-A71 [94, 95], anti-EV-D68 and anti-PV [95] 3C protease inhibitors. Furthermore, there are many other peptide-based inhibitors designed for combating enteroviral 3C proteases such as azapeptides [96, 97], diazomethyl ketones [98], S-nitrosothiols [99], tripeptidil  $\alpha$ -ketoamides [100].

### 1.2.3.2 Non-peptidic inhibitors

In order to develop orally available 3C protease inhibitors with more favorable pharmacokinetic properties, a series of reversible non-peptidic compounds was explored.  $\alpha,\beta$ -unsaturated ketobenzamides and mono-, di- and tri-substituted benzamides appeared to possess good inhibitory property against HRV 3C protease [101, 102] (compound 8 and 9, Figure 1.11).

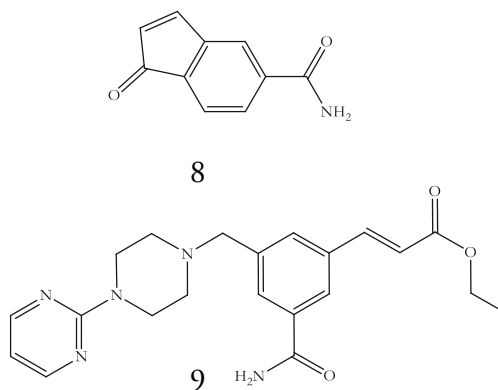


Figure 1.11  $\alpha,\beta$ -Unsaturated ketobenzamide 8 and 5-substituted benzamide 9.

A combination of structure-based design, molecular modelling, and structure-activity relationship analysis led to the discovery of 2,3-dioxindole-based reversible HRV inhibitors (isatins). Despite their high affinity for the protease pocket, this class of compounds were lacking antiviral activity and were very toxic, most probably due to the high electrophilic reactivity [103] (compound 10, Figure 1.12).

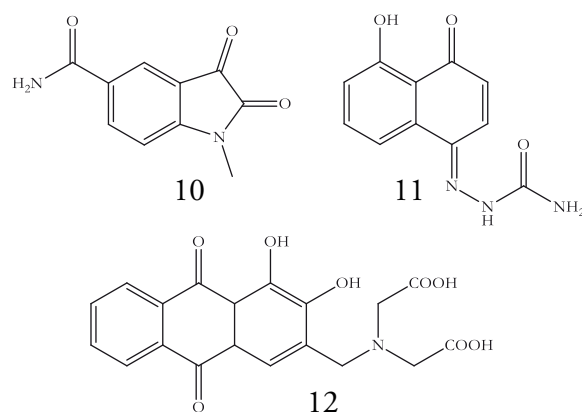


Figure 1.12 Isatin 10, benzoquinone 11 and anthraquinone 12.

Compounds with quinone moiety and quinone analogs were described as useful inhibitors of cysteine proteases including caspases and 3C proteases [Patent WO-02076939]. The compounds 11 and 12 (Figure 1.12) tested against several HRV showed moderate *in vitro*

activity. The mechanism of action is based on the assumption that quinone-like molecules act as Michael acceptors and are prone to attack by cysteine residue.

A blind screening revealed that homophthalimides are promising anti-HRV compounds with certain antiviral activity [104, 105]. MS analysis suggested that representative compound of this class forms a tight binding complex with 3C protein (compound 13, Figure 1.13).

The ability of heteroaromatic esters to inhibit HRV 3C proteases was recently described [106]. The best molecule 14, active also against SARS 3CL protease and HAV 3C protease, was modified resulting in similarly potent inhibitor with 4-quinolone moiety (compound 15, Figure 1.13).

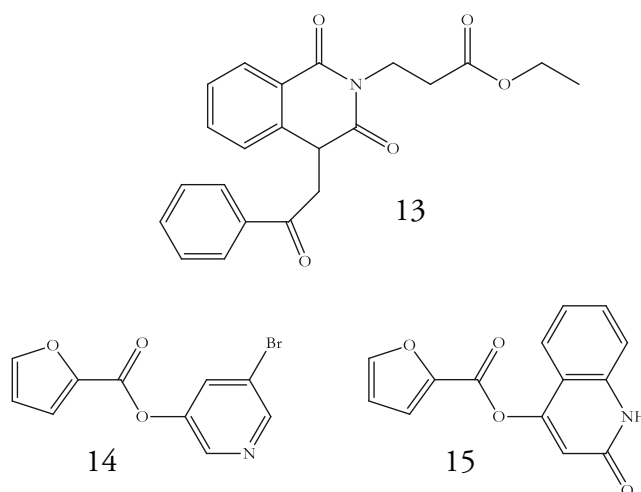


Figure 1.13 Homophthalimide 13 and heteroaromatic esters 14 and 15.

High-throughput screening of a small molecule library led to the discovery of several novel pyrazole-based inhibitors active towards SARS-CoV 3CL protease. One of them (compound 16, Figure 1.14) equally inhibited both the 3C protease and 3CL protease from PV and CoV, respectively. Further development of this compound gave a molecule 17 inhibiting various 3C(L) proteases with low  $\mu\text{M}$   $\text{IC}_{50}$  values, including SARS-CoV 3CL protease, CoV-229E 3CL protease, CV-B3 3C protease, EV-A71 3C protease, and HRV-A14 3C protease [107].

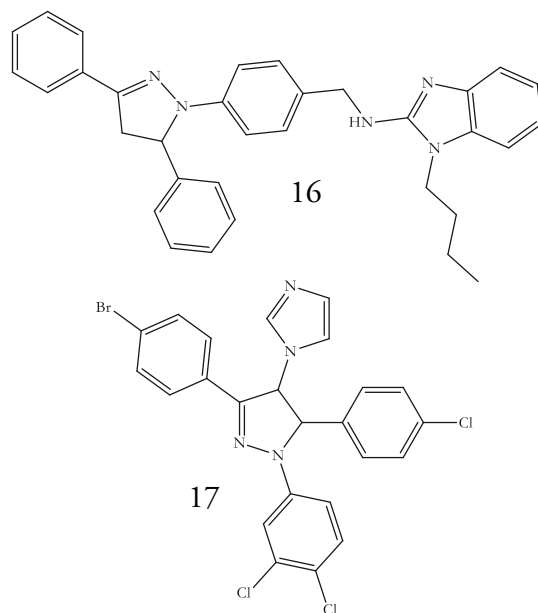


Figure 1.14 Pyrazole-based compounds.

Another series of molecules pseudoxazolones appeared to be potent time-dependent inhibitors of HAV and HRV 3C protease at micromolar levels (compound 18, Figure 1.15) [108, 109]. However, their reactivity towards thiols (e.g. DTT) does not allow them to be a drug candidate.

Since previously mentioned non-peptidic HRV 3C protease inhibitors, like isatines and homoftalimides, have suffered from cellular toxicity and average antiviral activity, a new class of low-molecular Michael acceptor-containing, irreversible HRV inhibitors was developed [110]. The selection of the core molecule was done based on the previously described potent peptidomimetic inhibitors such as rupintrivir. A small molecule with  $\alpha,\beta$ -unsaturated ethyl ester and lactam moiety was used for high-throughput synthesis of large number of potential active compounds. The best inhibitors (molecule 19, Figure 1.15) were highly active against HRV14, possessed good antiviral properties and were non-toxic to cells.

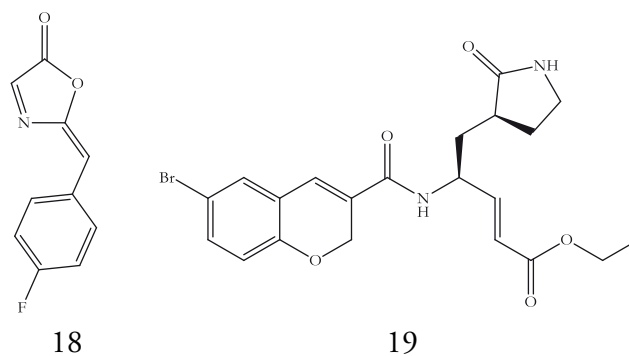


Figure 1.15 Pseudoxazolone 18 and Michael acceptor-containing small molecule 19.

It is worth mentioning microbial fermentation extracts as a source of antiviral inhibitors. Anti-HRV molecules include naphthoquinonelactol [111], a quinine like citrinin [112], radicinin [112], and triterpene sulfates [113].

### 1.2.3.3 Non-covalent inhibitors

All previously mentioned inhibitors possess electrophilic functionality that forms reversible or irreversible covalent interaction with the reactive cysteine (Cys-147) in the active site. Despite its high potency and selectivity, there is an increased risk of additional reactions with host proteins due to the electrophilic nature of the compounds. There is only one class of inhibitors mentioned in the literature that were proven by X-ray crystallography and NMR not to bind covalently to the enteroviral 3C protease [114]. 2-Phenylquinolone (compound 20, Figure 1.16) was selected through fragment-based screening, and appeared to be an effective non-covalent binder of HRV 3C protease with  $IC_{50} = 0.2 \mu\text{M}$ . A couple of analogs were studied in order to improve their inhibitory activity (compound 21 with  $IC_{50} = 10 \text{ nM}$ ) and solubility (compound 22 with  $IC_{50}$  comparable to the compound 20). The results showed that compound 18 binds to the P1 pocket in the proteolytic active site of the 3C protease.



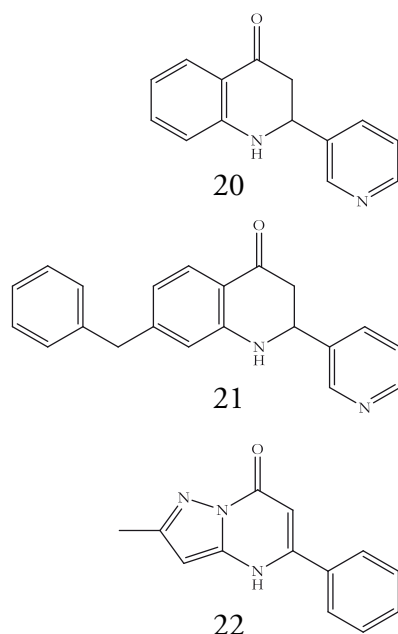


Figure 1.16 Non-covalent inhibitors of HRV 3C protease.

Researchers have to face several challenges to develop antiviral agents to treat infections caused by enteroviruses. Firstly, there are more than 250 different serotypes within the *Enterovirus* genus, which render vaccine development impractical. Secondly, to make a drug effective, good oral bioavailability and tissue distribution are essential to reach sufficient drug quantity at the infected site. Furthermore, the treatment for acute viral infections must be administered at a proper time for an optimal effect. Since most of the symptoms occur within 3 days of illness, the drug must be capable of reducing the severity of symptoms within the first 24 hours. Finally, the drug must have an excellent safety profile to ensure an appropriate risk/benefit ratio. To date, numerous anti-enteroviral agents have been identified to target the 3C protease, which is highly conserved among different serotypes. Unfortunately, none of the described in the literature compounds fulfilled all aforementioned requirements to be considered a safe and effective drug. Thus, there are no EV-specific antivirals on the market, which have been approved by the FDA. That is why it is essential to continue the development of drugs for enteroviral infections.

## 2. Objectives



The objectives of this thesis were:

1. Structural and functional characterization of EV-B93 3C protease.

The study involved: protein expression and purification, activity evaluation in the proteolytic activity assay, protein crystallization, determination of the crystal structure, and functional interpretation of the results.

2. Structural and biochemical characterization of potent irreversible inhibitors rupintrivir and compound 1 in complex with EV-B93 3C protease.

The study involved: protein-inhibitor complex preparation, inhibitory activity evaluation in the proteolytic activity assay, complex crystallization, determination of the complex crystal structure, and interpretation of the results.

3. Structural and biochemical characterization of small irreversible inhibitors NZO, NZN and DB5\_60 in complex with EV-B93 3C protease.

The study involved: protein-inhibitor complex preparation, inhibitory activity evaluation in the proteolytic activity assay, complex crystallization, determination of the complex crystal structure, and interpretation of the results.

4. Discovery of new non-covalent inhibitors against EV-B93 3C protease and its further biochemical, antiviral, and structural characterization.

The study involved: STD NMR based fragment screening, hit validation using Thermofluor and SPR, hit optimization and its further activity evaluation in the proteolytic activity assay and antiviral assay, complex preparation and crystallization, determination of the complex crystal structure, bioinformatics simulations of binding mode, interpretation of the results.



# 3. Materials and methods



### 3.1 Materials

This section lists the materials, instruments and common techniques used throughout the study. All reagents for media and buffer preparation were purchased from Merck, Fluka, Sigma, Panreac, Invitrogen, Roche, BioRad and Fermentas.

#### 3.1.1 Expression vector

The protein was expressed using the below characterized vector (Table 3.1).

**Table 3.1 Expression vector used in the study.**

<b>vector type</b>	pDEST, Gateway T7 Vectors
<b>cloning method</b>	Gateway®
<b>promoter</b>	T7
<b>protein tag</b>	HisTag (6x)
<b>antibiotic resistance</b>	ampicillin

#### 3.1.2 Plasmid purification

Plasmid purification was performed with the Qiagen Miniprep kit (Qiagen) on a table-top centrifuge following the manufacturer's guidelines.

#### 3.1.3 Growth media

- Luria-Bertani (LB) medium: 1 % [w/v] tryptone, 0.5 % [w/v] yeast extract, 1 % NaCl
- Autoinduction medium [115]

stock solutions:

1 M MgSO<sub>4</sub>

1000x Trace Metals (1L): 8 ml 5 M NCl, 5 g FeCl<sub>2</sub>·4H<sub>2</sub>O, 184 mg CaCl<sub>2</sub>·H<sub>2</sub>O, 64 mg H<sub>3</sub>BO<sub>3</sub>, 18 mg CoCl<sub>2</sub>·6H<sub>2</sub>O, 4 mg CuCl<sub>2</sub>·2H<sub>2</sub>O, 340 mg ZnCl<sub>2</sub>, 605 mg Na<sub>2</sub>MoO<sub>4</sub>·2H<sub>2</sub>O, 40 mg MnCl<sub>2</sub>·4H<sub>2</sub>O



50x 5052 (200 ml): 25 % [v/v] glycerol, 2.5 % [w/v] glucose, 10 % [w/v] lactose

50x M (250 ml): 1.25 M Na<sub>2</sub>HPO<sub>4</sub>·2H<sub>2</sub>O, 1.25 M KH<sub>2</sub>PO<sub>4</sub>, 2.5 M NH<sub>4</sub>Cl, 0.25 M Na<sub>2</sub>SO<sub>4</sub>

400 ml of autoinducing media contains: 4 g tryptone, 2 g yeast extract, 380 ml MiliQ water, 800 µl 1 M MgSO<sub>4</sub>, 800 µl 1000x Trace Metals, 8 ml 50x 5052, 8 ml 50x M, and was supplemented with 400 µl of each antibiotic and 4 ml pre-culture.

- LB agar plates: 1 % [w/v] tryptone, 0.5 % [w/v] yeast extract, 1 % NaCl, 1.5 % [w/v] agar
- Antibiotics: ampicillin (stock solution: 100 mg/ml in MiliQ water, working concentration: 100 µg/ml); chloramphenicol (stock solution: 35 mg/ml in 100 % [v/v] ethanol, working concentration: 35 µg/ml)

### 3.1.4 Bacterial cloning and expression strains

Bacterial cloning and expression strains used in the study are listed in Table 3.2.

**Table 3.2 Cell strains used in the study.**

cell strain	genotype	remarks
DH5α	F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK- mK <sub>+</sub> ), λ-	Cloning strain. Novagen
Rosetta (DE3)	F- ompT hsdSB(RB- mB-) gal dcm λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5]) pLysSRARE (CamR)	Expression strain with DE3, and pLysSRARE. pLysSRARE contains tRNA genes argU, argW, ileX, glyT, leuW, proL, metT, thrT, tyrU, and thru. Novagen

### 3.1.5 Transformation of bacterial strains

Transformations were performed with competent *Escherichia coli* cells obtained through the calcium chloride-method [116]. Competence was induced by heat-shock treatment at 42 °C for 45 s.

### 3.1.6 Protein electrophoresis

Protein electrophoresis (SDS-PAGE) was carried out using BioRad electrophoresis tanks and power sources applying currents that ranged between 150 V-250 V. Obtained stained gels were subsequently scanned. Reagents for gel preparation, running and staining are detailed below:

- 15 % SDS PAGE-gel: separative gel (15 % [w/v] acrylamide, 0.4 % [w/v] bis-acrylamide, 0.1 % [w/v] SDS in 0.38 M Tris pH 8.8), concentrative gel (6 % [w/v] acrylamide, 0.16 % [w/v] bis-acrylamide, 0.1 % [w/v] SDS in 0.13 M Tris pH 6.8. Polymerization occurred in the presence of 0.06 % [v/v] cross-linking reagent N,N,N',N'-tetramethylethylenediamine (TEMED) and 0.05 % [w/v] initiator ammonium persulfate (APS).
- running buffer (0.025 M Tris, 0.2 M Glycine, 0.05 % [w/v] SDS)
- loading buffer (55 mM Tris pH 6.8, 1.6 % [w/v] SDS, 8.3 % [v/v] glycerol, 0.002 % [w/v] Bromophenol Blue, 5 % [v/v]  $\beta$ -mercaptoethanol)
- molecular weight markers: SeeBlue (Invitrogen)
- Coomassie staining solution (0.5 % [w/v] Coomassie Blue G-250 (Sigma), 50 % [v/v] methanol, 10 % [v/v] acetic acid)
- Coomassie destain solution (10 % [v/v] acetic acid, 10 % [v/v] isopropanol)

### 3.1.7 Chromatography columns

Protein purification and peptide analysis columns were purchased from GE Healthcare and were used on ÄKTA Purifier systems (GE Helathcare). Protein purification and peptide analysis protocols were carried out using the following materials:

- HisTrap HP column (5 ml)
- Superdex 75 column (10/300)
- Superdex 75 column (16/60)
- SOURCE™ 5RPC ST 4.6/150
- SOURCE™ 15RPC ST 4.6/100

### 3.1.8 Protein concentration

Protein concentration was carried out with Vivaspin (GE Helathcare) and Amicon (Merck Millipore) devices of different sizes (0.5 ml, 2 ml, 4 ml, 6 ml, 20 ml) with different cut-off values (3000, 10000 MWCO).

Protein concentration was measured on a Nanodrop 2000 device (Thermoscientific).

Protein purity was assessed by analyzing the SDS-PAGE gels, gel filtration chromatograms and MS spectra.

### 3.1.9 Mass spectrometry

MALDI-TOF Mass Spectrometry (MS) analysis was performed in the Proteomics Platform of Barcelona Science Park.

### 3.1.10 Protein crystallization

Reagents and crystallization materials were purchased from Hampton Research and Molecular Dimensions. Crystallization experiments were performed at the Crystallography Platform of Barcelona Science Park.

The crystals screens used in this study are listed in Table 3.3.

**Table 3.3 Crystal screens used in the study.**

<b>code</b>	<b>screen</b>	<b># of conditions</b>	<b>reference</b>
<b>PAC1</b>	Crystal Screen I	48	Hampton Research
	Crystal Screen II	48	
<b>PAC2</b>	Wizard I	48	Emerald Bio
	Wizard II	48	
<b>PAC3</b>	Index	96	Hampton Research
<b>PAC4</b>	Salt RX	96	Hampton Research
<b>PAC15</b>	JBS1	24	Jena Bioscience
	JBS2	24	
	JBS3	24	
	JBS4	24	
<b>PAC16</b>	JBS5	24	Jena Bioscience
	JBS6	24	
	JBS7	24	
	JBS8	24	
<b>PAC17</b>	JBS9	24	Jena Bioscience
	JBS10	24	Molecular Dimensions
	Clear Strategy I – 8.5	24	
	Clear Strategy II – 8.5	24	
<b>PAC21</b>	PACT premier HT-96	96	Molecular Dimensions
<b>PAC22</b>	Pi-PEG Screen	96	Jena Bioscience
<b>PAC23</b>	Pi – minimal screen	96	Jena Bioscience
	Midas	96	Molecular Dimensions
	Morpheus	96	Molecular Dimensions

Crystallization experiments were performed using the sitting or hanging drop vapor diffusion method.

Initial screens were performed in 96-well sitting drop MRC plates. Crystallization screenings were prepared with Tecan robots (Tecan). Crystallization drops were handled by Cartesian (Cartesian Technologies) and Crystal Phoenix robots (ArtRobbins), both providing microscale liquid handling for high-throughput crystallization screening. The best conditions were optimized on 24-well Linbro plates (Hampton Research). Particular conditions and settings are indicated for every experiment in following sections.

### 3.1.11 X-ray data collection, processing and structure determination

X-ray data collection was performed in the European Synchrotron Radiation Facility (ESRF, Grenoble, France) at the beamlines ID23-2, ID14-4, ID29 and the ALBA synchrotron (Cerdanyola de Vallès, Spain) at the beamline XALOC.

Diffraction data were indexed and integrated using XDS [117], and scaled and merged with SCALA [118]. Structure elucidation and refinement were carried out using the CCP4 suite of programs [119]. The initial phases were obtained by molecular replacement using PHASER [120] with previously-generated EV-B93 3C structure as a model (PDB code: 3Q3X). REFMAC [121] was employed for structure refinement and Coot [122] for interpretation of electron density and model building. Geometry restraint information for covalent ligands was generated with SKETCHER within CCP4.

### 3.1.12 Saturation transfer difference NMR screening

STD-NMR screening was performed at the PCB NMR facility within collaboration with the group of Ernest Giralt (IRB Barcelona, Spain). STD-NMR spectra were acquired on a Varian Inova 500 MHz spectrometer with 5 mm PFG Penta Probe equipped with automatic sample changer (S/N for 1 H was 815:1 (0.1% ethylbenzene in CDCl<sub>3</sub>)) at 25 °C. The data were analyzed with MestreNova software.

All common materials and instruments needed for the NMR experiment were provided by the laboratory of Ernest Giralt. Specific deuterated chemicals were purchased from cortecnet, Deutero GmbH, euriso-top.

Fragment library consisted of 527 compounds mixed in a cocktails of 5 was designed and prepared by Michael Goldflam in the laboratory of Ernest Giralt (IRB Barcelona, Spain).

### 3.1.13 Thermal shift assay

Thermal shift assay (TSA) measurements were carried out in an iCycler iQ Real Time PCR Detection system (BioRad). Data evaluation and melting point determination were performed using BioRad Optical System Software.

### 3.1.14 Surface plasmon resonance experiment

Surface plasmon resonance (SPR) analysis was performed at the Scientific and Technological Centers, University of Barcelona, Spain). SPR experiments were carried out on a Biacore T100 instrument (GE Healthcare) with the use of Series S Sensor Chip CM5 (GE Healthcare). The data were analyzed with BiaEvaluation 1.1 software (GE Healthcare) and Scrubber 2.0 software (BioLogic).

All common materials needed for the SPR experiment were provided by the Cytometry Unit. Amine Coupling Kit and His Capture Kit were purchased from GE Healthcare.

### 3.1.15 Cell-based proteolytic assay and antiviral assay

Cell-based proteolytic assay [123] and antiviral assay were performed by Rachel Ulferts (prof. Frank van Kuppeveld's laboratory, Utrecht University, The Netherlands).

### 3.1.16 Analogs selection, docking and classical molecular dynamics simulations

Bioinformatics calculations including analogs selection, docking and classical molecular dynamics (MD) simulations were performed by Daniel Álvarez (Xavier Barril's laboratory, University of Barcelona, Spain). Analogs selection was performed through the sub-structural similarity search in the in-house database or ZINC database using SMILE/SMART definitions and BABEL software. Docking was performed using rDOCK software and classical MD simulations were prepared using pyMDMix (<http://mdmix.sourceforge.net/>) and run with AMBER software.

## 3.2 Methods

This section lists the materials and methods selected and designed for the purpose of this work and used throughout the study.

### 3.2.1 Construct design

The construct of EV-B93 3C protease was designed and prepared by Bruno Coutard (Architecture et Fonction des Macromolécules Biologiques, Marseille, France). The 3C protease gene was cloned into a pDEST14 expression vector with an N-terminal MK and C-terminal His<sub>6</sub> tag.

The C147A mutant of EV-B93 3C protease was previously obtained via site-directed mutagenesis. The protein was expressed and purified by Lionel Costenaro (IBMB-CSIC, Barcelona).

### 3.2.2 Plasmid transformation

1 µl of the plasmid DNA containing the sequence for the EV-B93 3C protease was mixed with 50 µl of competent cells. The mixture was incubated 30 min. on ice and then briefly heated to 42 °C during 45 s. The mixture was placed back on ice for 2 min and 500 µl of LB medium was added. The cells were incubated in 37 °C for 1 h and plated out for marker selection.

### 3.2.3 Protein expression

EV-B93 3C protein expression was performed in Rosetta *E. coli* cells. Two flasks filled with 150 ml LB medium each were used for pre-cultures preparation. To each flask one colony of freshly transformed bacteria and two antibiotics ampicillin and chloramphenicol were added. The flasks were incubated overnight at 37 °C at 220 rpm agitation speed. The cultures were grown in 2 l flasks containing 750 ml of autoinduction media supplemented with two antibiotics ampicillin and chloramphenicol. The cells were initially incubated at 37 °C for 3 h with 220 rpm agitation speed and then the temperature was lowered to 16 °C. The induction time was 24 h. Cells were harvested at 7000 rpm, and cellular pellet was frozen in -20 °C.

### 3.2.4 Protein purification

#### Ni-NTA affinity purification and gel filtration

Cell pellet coming from the large scale cultures was resuspended in the lysis buffer (50 mM Tris pH 8.0, 200 mM NaCl, 30 mM imidazole, 2 mM  $\beta$ -mercaptoethanol) supplemented with Complete Protease Inhibitor Cocktail Tablets (Roche) (1 tablet/50 ml) and 0.75 mg/ml DNase I. The cells were lysed using cell disruptor (PECF Constant Systems Ltd.) and centrifuged at 13000 rpm or 24000 rpm. The supernatant was filtered and loaded on the pre-equilibrated HisTrap HP column. Column was washed with 15 CV of binding buffer (50 mM Tris pH 8.0, 200 mM NaCl, 30 mM imidazole, 2 mM  $\beta$ -mercaptoethanol) and eluted with elution buffer (50 mM Tris pH 8.0, 200 mM NaCl, 500 mM imidazole, 2 mM  $\beta$ -mercaptoethanol) in a 20 CV linear gradient from 30 to 500 mM imidazole. Eluted fractions from the affinity column were concentrated and loaded on the Superdex 75 (16/60) column. Protein was eluted with 1.3 CV gel filtration buffer (10 mM Tris pH 8.0, 200 mM NaCl, 2 mM DTT). All above steps were performed at 4 °C. Collected fractions were flash frozen in liquid nitrogen and stored in -20 °C.

The C147A mutant of EV-B93 3C protease was previously produced, purified, flash-frozen in liquid nitrogen and stored in -20 °C by Lionel Costenaro (IBMB-CSIC, Barcelona).

### 3.2.5 Reductive methylation

1250 ml of 4.4 mg/ml EV-B93 3C protease solution was used for buffer exchange (50 mM Hepes pH 7.4, 10 mM NaCl) using PD-10 Desalting Column (GE Helathcare). Resulting solution was concentrated to volume of 1.3 ml and transferred to a 2 ml tube with a small stirring bar. 20  $\mu$ l of freshly prepared 1M DMAB (Fluka) was added followed by 40  $\mu$ l 1M methanol-free formaldehyde (Thermo Scientific). Mixture was incubated for 2h with gentle stirring. 20  $\mu$ l of 1M DMAB and 40  $\mu$ l of 1M formaldehyde were added, and the mixture was left for additional 2h. The final aliquot of 10  $\mu$ l 1M DMAB was added and the reaction mixture was left overnight. Finally, 80  $\mu$ l of 1M glycine was added and stirred for 1 h. The protein solution was repurified on Superdex 75 10/300 column using 1.3 CV of the elution buffer (10 mM Tris pH 8.0, 200 mM NaCl, 8 mM DTT). All steps were performed at 4 °C.



### 3.2.6 *In vitro* proteolytic activity assay

100  $\mu$ l of reaction mixture containing 1.9  $\mu$ M 3C protease, 25  $\mu$ M substrate peptide, 5 % DMSO, buffer (30 mM Tris pH 8.0 and 50 mM NaCl or 50 mM HEPES pH 7.4 and 50 mM NaCl) and (optionally) different concentrations of inhibitor (ranging from 0 to 350  $\mu$ M) was incubated in 37 °C during 3 h or 20 h. The reaction was quenched by addition of 0.5 % (final concentration) trifluoroacetic acid (TFA). Samples were analyzed by reverse-phase high-performance liquid chromatography (HPLC) on SOURCE™ 5RPC ST 4.6/150 or SOURCE™ 15RPC ST 4.6/100 column using: 2 % to 90 % linear gradient of acetonitrile in 0.1 % TFA

The 3C protease activity in the absence of inhibitor was defined as 100 %.

The following peptides (GL Biochem) were dissolved in DMSO at a concentration of 5 mM and used for testing 3C proteases activity (Table 3.4).

**Table 3.4 Peptide sequences used for testing various 3C proteases.**

<b>protease</b>	<b>peptide sequence</b>
EV-B93 3C	Ac-RHSVGATLEALFQ↓GPPVYREIKIS-NH <sub>2</sub>
EV-B93 3C (C147A)	Ac-RHSVGATLEALFQ↓GPPVYREIKIS-NH <sub>2</sub>
ERBV-1 3C	Ac-AKDGGIFAQ↓SRDRHYLVGTV-NH <sub>2</sub>
HRV-A49 3C	Ac-RHSVGATLEALFQ↓GPPVYREIKIS-NH <sub>2</sub>
EV-D68 3C	Ac-RHSVGATLEALFQ↓GPPVYREIKIS-NH <sub>2</sub>
AiV 3C	Ac-RHSVGATLEALFQ↓GPPVYREIKIS-NH <sub>2</sub>
PSV 3C	Ac-RHSVGATLEALFQ↓GPPVYREIKIS-NH <sub>2</sub>

### 3.2.8 Crystallization

#### 3.2.8.1 Protein sample preparation

The protein solution coming from the purification step was centrifuged at 14500 rpm for 10 min.

### 3.2.8.2 Protein-covalent compound complex preparation

2.0-2.8 mg/ml protein solution containing a 2-fold molar excess (for rupintrivir and compound 1) or 12-fold molar excess (for NZN, NZO and DB5\_60) of compound were incubated overnight in 20 °C (rupintrivir, compound 1) or 30 °C (NZN, NZO and DB5\_60). The mixture was loaded on the Superdex 75 (10/300) column and protein was eluted using 1.3 CV of the buffer (10 mM Bicine pH 8.5, 300 mM NaCl, 1 mM DTT for rupintrivir/compound 1 and 10 mM Tris pH 8.0 100 mM NaCl for DB5\_60). The collected fractions were concentrated to 2.3-3.2 mg/ml.

Rupintrivir and compound 1 were provided by Pfizer, while NZO, NZN and DB5\_60 were synthesized by Daniel Becker (Jörg Rademann's laboratory, Freie Universität Berlin, Germany). All compounds were dissolved in DMSO at concentrations of 0.5 mM (rupintrivir), 50 mM (compound 1) or 100 mM (NZO, NZN and DB5\_60).

### 3.2.8.3 Protein-non-covalent compound complex preparation

1.5-5.0 mg/ml protein solution was incubated with 5, 10 or 20-fold molar excess of compound and (optionally) with equal amount of peptide substrate in room temperature for 0.5 h.

The compounds from the screening library were dissolved in a mixture of deuterated DMSO:H<sub>2</sub>O (9:1 ratio) at a concentration of 100 mM, while the analogs (purchased from Enamine, Ambinter, FCH or synthesized by Daniel Becker) were dissolved in DMSO at concentrations of 12.5 mM, 50 mM or 100 mM.

### 3.2.8.4 Crystallization screening and optimization

Crystallization trials were performed as indicated in Materials section 3.1.10 at 20 °C. Drops were prepared by mixing 150 µl of protein solution with 150 µl of reservoir solutions. Crystal optimization was performed manually with 1+1 µl sitting drops.

All the conditions and PAC screens, which were used in protein and protein-complex crystallization are listed below (Table 3.5)

**Table 3.5 Protein or protein/complex crystallization screening conditions.**

protein/complex	excess of compound	additive	concentration [mg/ml]	temperature [°C]	screen
EV-B93 3C	-	-	2.3	20	PAC1-4, PAC15-17, PAC21-23, Midas, Morpheus
EV-B93 3C (methylated)	-	-	4.7	20	PAC1-4
EV-B93 3C – rupintrivir	2x	-	2.5	20	PAC1-4
EV-B93 3C – compound 1	2x	-	3.2	20	PAC1-4
EV-B93 3C – NZN	12x	-	2.4	20	PAC1-4
EV-B93 3C – NZO	12x	-	2.3	20	PAC1-4
EV-B93 3C – DB5_60	12x	-	2.3	20	
EV-B93 3C – 164, 188, 290, 306, 484, 544	20x	-	5.0	20	PAC1-4, PAC15-17
EV-B93 3C –EN2, EN12	10x	-	1.5	20	PAC1-4
EV-B93 3C –EN2, A4	20x	-	2.9	20	PAC1-4
EV-B93 3C –A4	5x	-	2.9	20	PAC1-4
EV-B93 3C –A4	5x	peptide substrate (1x)	2.3	20	PAC1-4
EV-B93 3C –A4_D	5x, 20x	peptide substrate (1x)	2.3	20	PAC1-4

### 3.2.8.5 Soaking experiment

2  $\mu$ l drops with previously prepared crystals of native protein were soaked for 10 s, 10 min, 1 day or 3 days before freezing with non-covalent compound by addition of 0.5  $\mu$ l of its 50 mM or 100 mM DMSO solution.

### 3.2.8.6 Mounting and freezing crystals

Protein crystals were fished with nylon cryoloops, and immediately soaked in solutions containing reservoir solution supplemented with 25 % of glycerol, MPD or PEG400 and (optionally) with appropriate non-covalent compound at a final concentration of 25 mM or 50 mM. Cryoprotected crystals were flash frozen in liquid nitrogen.

### 3.2.9 Saturation transfer difference NMR

Buffer exchange for the EV-B93 3C protease was performed using PD-10 Desalting Column (GE Helathcare). The deuterated buffer for saturation transfer difference NMR (STD NMR) contained 30 mM Tris- $d_{11}$  and 50 mM NaCl in deuterated water (99.8 %) at pH 8.0.

Samples of 600  $\mu$ l were prepared in an 3 mm NMR tube and contain 10  $\mu$ M EV-B93 3C protease, 500  $\mu$ M of each compound, 10  $\mu$ M DSS, 30 mM Tris pH 8.0 and 50 mM NaCl with the total concentration of DMSO below 3 %. In the case the spectrum contained STD signals, the covalent inhibitor rupintrivir was added to the mixture reaching the final concentration of 20  $\mu$ M and additional STD-NMR experiment was performed.

The  $^1\text{H}$  NMR spectra were acquired with presat water suppression and 32 increments. For STD-NMR experiments a pulse sequence with water suppression by the Watergate technique [124] was selected and 1,024 increments were recorded. The length of the mixing time to remove residual protein signals was set to 10 ms. Saturation time was set to 2.5 ms and the relaxation delay to 2 s. On and off resonance frequency for EV-B93 3C protease were 0 and 35 ppm. All spectra were recorded using an 8000 Hz spectral width.

### 3.2.10 Thermal shift assay

The aliquots were prepared on ice in a 96-well PCR plates (Bio-Rad) sealed with optically clear PCR sealers (Bio-Rad) and contain 12.2  $\mu$ M EV-B93 3C, 30 mM Tris pH 8.0, 50 mM NaCl and 42x Sypro Orange protein gel stain (Invitrogen). Then compounds dissolved in DMSO were added at a final concentration of 1 mM. For the control sample the same volume of DMSO were added to reach final concentration of 5 %. Samples were heated from 20 °C to 95 °C at a rate of 0.5 °C/min, and the change in absorbance ( $\lambda_{\text{ex}} = 490 \text{ nm}$ ;  $\lambda_{\text{em}} = 575 \text{ nm}$ ) was monitored over time in an iCycler iQ Real Time PCR Detection System (Bio-Rad).

### 3.2.11 Surface plasmon resonance

Running buffer containing 10 mM Tris 8.0, 100 mM NaCl and 5 % of DMSO was used. The EV-B93 3C protease was amine coupled (EDC/NHS, Amine Coupling Kit) using a 10 mM NaAc pH 6.0 as a coupling agent. Immobilization level was ranging from 1,700 to 2,050 RU. For the protein captured via its His-tag, firstly the antibody was covalently immobilized on the sensor chip (EDC/NHS, His Capture Kit) using a 10 mM NaAc pH 4.5 as a coupling agent at 10,000 RU. Then the protein was captured with the immobilization level around 1,900 RU.

The stock solutions of used compounds were taken from the screening library. Concentration series (1.0, 0.5, 0.25, 0.125 and 0 mM) were prepared in the running buffer in a 96-well plate (Greiner Bio-One) and sealed (GE Healthcare). The flow rate was 30  $\mu$ l/min. Test substances were injected at 20 °C with a contact time of 60 s and dissociation time of 180 s. The new portion of protein was coupled every 5 measurements.

# 4. Results and discussion



## 4.1 Structural and functional studies of EV-B93 3C protease

Expression and purification of native EV-B93 3C protease and its C147A mutant were performed in order to characterize them biochemically and structurally. Preparation of C147A mutant was aimed at investigating the importance of the Cys-147 residue in the proteolytic activity of the enzyme. The Cys-147, a member of a putative catalytic triad, was replaced by Ala-147. Large amount of native EV-B93 3C protease was needed for subsequent fragment screening and testing of possible hits. The C147A mutant was previously expressed and purified by Lionel Costenaro (IBMB-CSIC).

### 4.1.1 Expression and purification of native EV-B93 3C protease

The protein was expressed in *E. coli* cells and then purified using Ni-NTA affinity column and gel filtration. IMAC (Immobilized Metal ion Affinity Chromatography) purification provided two elution peaks (Figure 4.1 A, left side). The first one was very small and corresponds to the fractions giving multiple bands in SDS-PAGE gel (Figure 4.1 A, right side). The second peak, much higher than the first one, corresponds to the fractions giving broad bands in SDS-PAGE gel (Figure 4.1 A, right side). The collected fractions from the second peak were concentrated and loaded on the gel filtration column. The protein appeared as a single peak (Figure 4.1 B, left side) which corresponds to pure, highly concentrated 3C protease (Figure 4.1 B, right side). The molecular mass of the protein eluted during both purification steps corresponds to the mass of monomeric EV-B93 3C protease of 21.3 kDa. The purification yield was ~ 25 mg per liter of cell culture. The presence of EV-B93 3C protease was confirmed by MS-MS fragmentation.



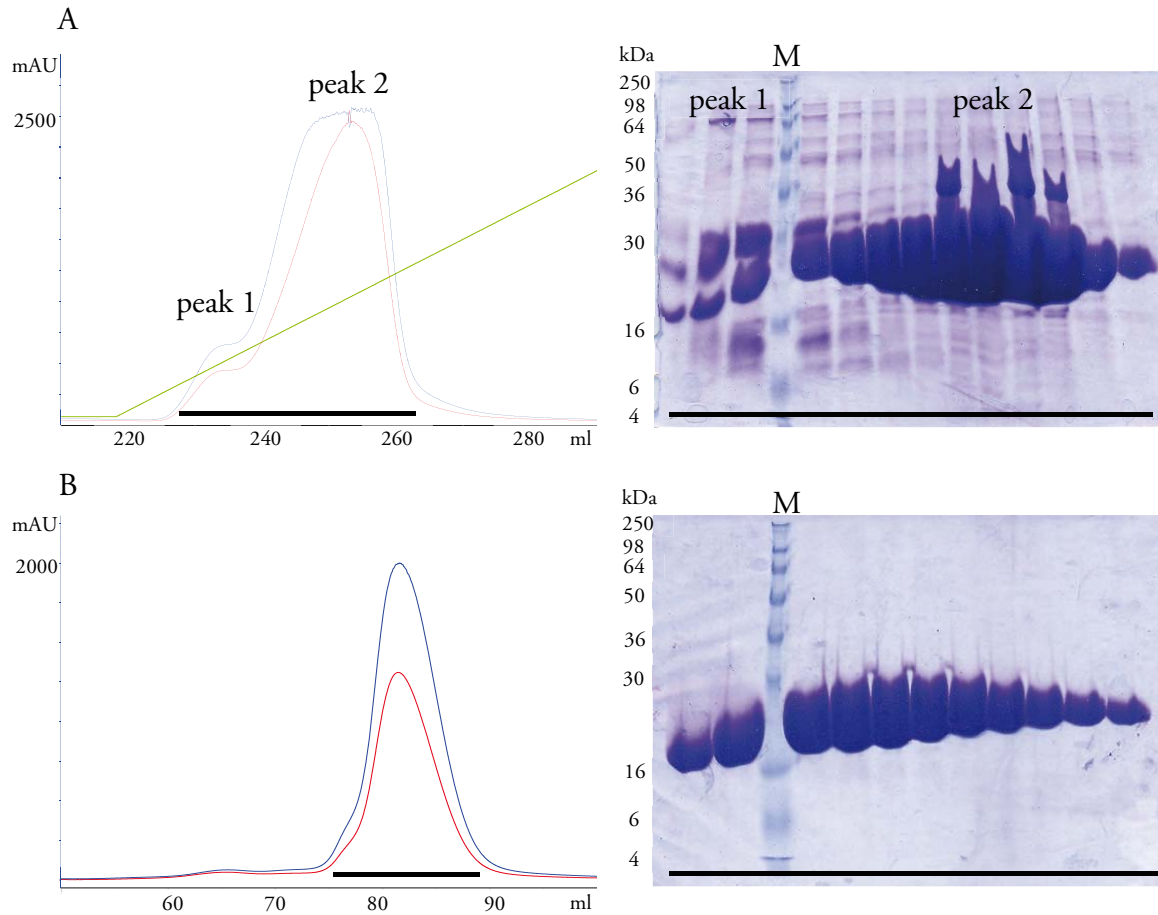


Figure 4.1 Purification of EV-B93 3C protease. (A) Chromatogram from IMAC purification and corresponding SDS-PAGE gel. (B) Chromatogram from gel filtration and corresponding SDS-PAGE gel. Green line corresponds to imidazole gradient, blue line to absorbance at 280 nm, red line to absorbance at 260 nm and black bar indicates fractions collected and loaded on the SDS-PAGE gel.

#### 4.1.2 Biochemical characterization of native EV-B93 3C protease and its C147A mutant

The proteolytic activity of EV-B93 3C protease and its C147A mutant was tested in the *in vitro* assay with the use of peptide substrate (Ac-RHSVGATLEALFQ↓GPPVYREIKIS-NH<sub>2</sub>) that mimics 2C↓3A cleavage site present in the enteroviral polyprotein (Figure 4.2).

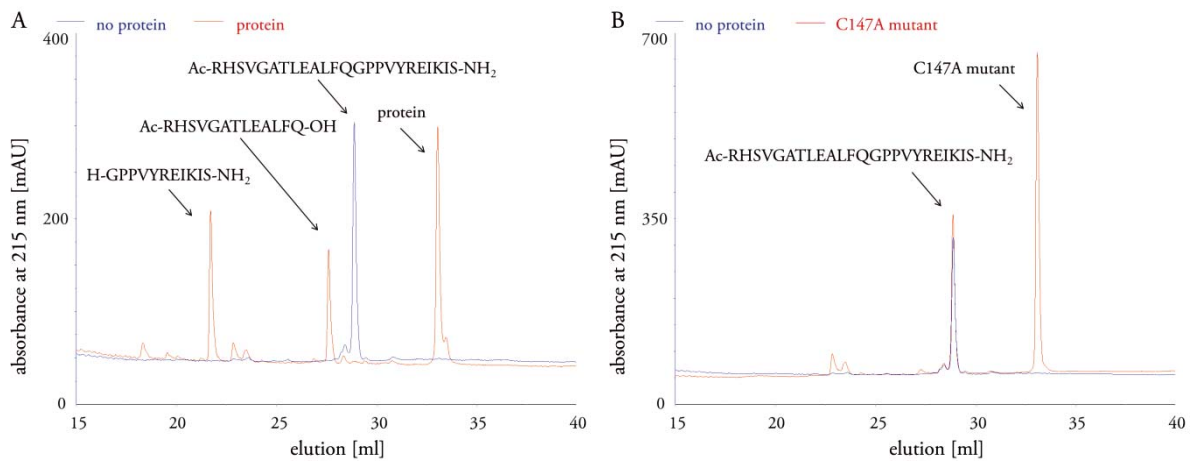


Figure 4.2 *In vitro* proteolytic activity assay of EV-B93 3C protease (A) and its C147A mutant (B). Reverse-phase chromatograms show in red the products of the digestion of the peptide (Ac-RHSVGATLEALFQ↓GPPVYREIKIS-NH<sub>2</sub>) by EV-B93 3C protease and its C147A mutant. Chromatograms of the peptide substrate alone are shown in blue.

The reverse phase HPLC analyses of the overnight incubations of the peptide with EV-B93 3C protease revealed the presence of expected cleavage products and the protein. The results were confirmed by MALDI-TOF MS (Matrix Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry) for the cleavage substrate and the products (Figure 4.3 B, C, D), and for the protein (Figure 4.3 A). The proteolytic activity assay performed with the C147A mutant of EV-B93 3C protease showed no effect on the substrate, since no cleavage products were observed; thereby confirming that presence of the Cys-147 is critical for the proteolytic activity of EV-B93 3C protease.

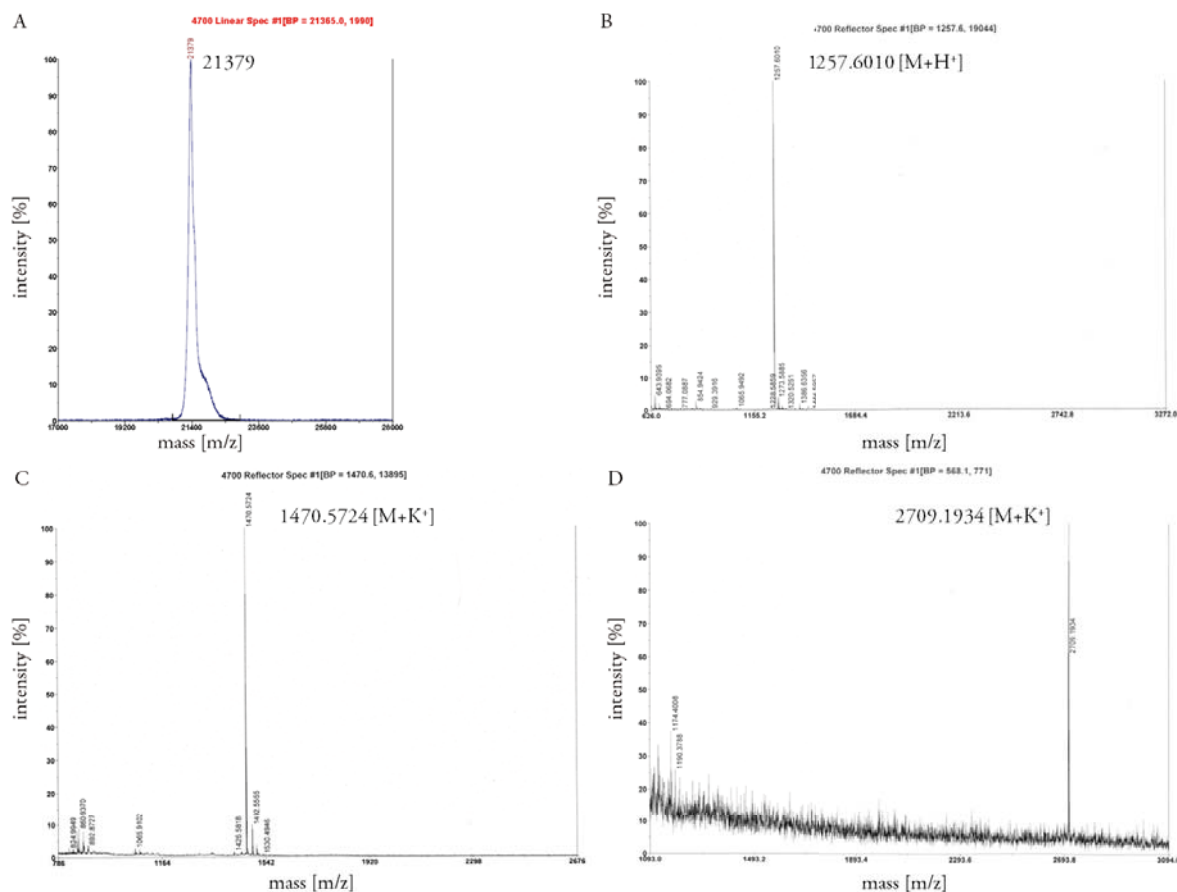
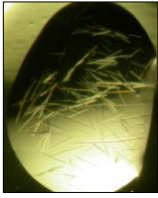


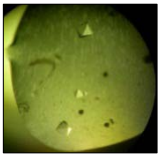


Figure 4.3 MALDI-TOF mass spectra of native EV-B93 3C protease (A), cleavage peptide products (B and C) and peptide substrate (D).

#### 4.1.3 Crystallization screening and crystal optimization of native EV-B93 3C protease

The active and native EV-B93 3C protease was subjected to crystallization screening which resulted in many conditions containing thin needle-like crystals, a couple of rod-like crystals and only one with octahedron-like crystals. The most promising conditions were chosen for further optimizations. The best crystal conditions and corresponding crystal data are summarized in Table 4.1.

Table 4.1 EV-B93 3C protease crystallization conditions and the corresponding crystal data.

PAC	initial condition	optimized condition	crystal shape	resolution [Å]	space group	cell dimension
PAC1 B6	20 % PEG8000 0.2 M Mg(Ac) <sub>2</sub> 0.1 M sodium cacodylate pH 6.5	18 % PEG8000 0.2 M Mg(Ac) <sub>2</sub> 0.1 M sodium cacodylate pH 6.0		1.90	P2 <sub>1</sub>	a=38.90 b=64.62 c=66.27 β=90.90
PAC22 A9	15 % PEG2000 0.1 Tris pH 8.0	12 % PEG2000 0.1 Tris pH 8.0		1.57	P2 <sub>1</sub>	a=38.95 b=56.94 c=84.62 β=94.99
PAC4 F10	1.5 M LiSO <sub>4</sub> 0.1 M NaAc pH 4.6	1.5 M LiSO <sub>4</sub> 0.1 NaAc pH 5.6		2.70	P6 <sub>5</sub>	a=135.9 b=135.9 c=51.57
PAC1 C8	2M (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	2M (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>		2.70	P6 <sub>5</sub>	a=123.27 b=123.27 c=102.63

The protein in the presence of PEGs was crystallizing in the monoclinic P2<sub>1</sub> space group and was able to give high-resolution diffraction data (below 2 Å). The crystals obtained in the acidic conditions and/or the presence of sulfates appeared to be of lower resolution (2.70 Å) and belong to the hexagonal P6<sub>5</sub> space group. The crystals grown in 18 % PEG8000, 0.2 M Mg(Ac)<sub>2</sub> and 0.1 M sodium cacodylate pH 6.0 were very thin and susceptible to radiation damage, which may have had an impact on the resolution (1.90 Å). However, the crystals obtained in 12 % PEG2000 and 0.1 Tris pH 8.0 were thicker and more resistant, which led to the improvement in the data resolution (1.57 Å).

#### 4.1.4 Structure determination of EV-B93 3C protease and data analysis

Many data sets of different EV-B93 3C protease crystals were collected and then solved with molecular replacement (see Materials and Methods section) using previously obtained structure of EV-B93 3C protease as a search model (PDB code: 3Q3X). This model structure has inactivating the enzyme oxidized cysteine residue in the active site, which was also observed in many data sets collected within this study. However, it was possible to obtain a few EV-B93 3C protease structures with no or partial oxidation. It is believed that the oxidation occurs during crystallization or X-ray data collection, since the purified protein is active *in vitro* (see 4.1.2 section). The selected structure of EV-B93 3C protease was refined at 1.6 Å to a final  $R_{\text{work}}$  of 16.2 % and  $R_{\text{free}}$  of 19.6 % (see Table 4.2 for data collection and refinement statistics). It was possible to model all residues except the His<sub>6</sub> tag and the last two and one C-terminal residues for chains A and B, respectively. The Ramachandran plot shows 96.06 % of the residues to be in the preferred regions, while 3.66 % and 0.28 % in the allowed and the outlier regions, respectively. The asymmetric unit contains two molecules that are related by a non-crystallographic 2-fold axis. The refined structure has 208 water molecules and 2 glycerol molecules.

**Table 4.2 Data collection and refinement statistics for EV-B93 3C protease native crystal.**

parameter	EV-B93 3C protease
resolution range [Å]	84.30-1.57 (1.61-1.57)
space group	P2 <sub>1</sub>
cell dimensions [Å, °]	a=38.95 b=56.94 c=84.62 β=94.99
no. of observed reflections	164878
no. of unique reflections	51255 (3742)
completeness [%]	99.31 (99.4)
mean multiplicity	3.2 (3.3)
R <sub>merge</sub>	0.074 (0.498)
<I>/σ<I>	10.06 (2.01)
R <sub>work</sub>	0.162
R <sub>free</sub>	0.196
no. of protein atoms (non-H)	3111
no. of water molecules	208
no. of hetero compounds	2 glycerol molecules
RMSD for bond length [Å]	0.022
RMSD for bond angles [°]	2.038
mean B value [Å <sup>2</sup> ]	15.391
Ramachandran analysis:	96.06/3.66/0.28
preferred/allowed/outliers [%]	

Values in parenthesis correspond to the outer resolution shell.

$R_{merge} = \frac{\sum_h \sum_i |I_{i,h} - \langle I_h \rangle|}{\sum_h \sum_i I_{i,h}}$ , where  $I_{i,h}$  is the  $i$ th-intensity measurement of reflection  $h$  and  $\langle I_h \rangle$  is the average intensity for multiple measurements.

$R_{work}$  and  $R_{free} = \frac{\sum_h ||F_o| - |F_c||}{\sum |F_o|}$ .  $R_{free}$  was calculated for 5% of the reflections not used for refinement.

EV-B93 3C protease folds into two antiparallel β-barrels that are orientated 90° apart, linked by a 20-amino-acid loop containing a short α-helix and flanked by other two α-helices at the N- and C-termini of 14 and 3 amino acids long, respectively. The two barrels are topologically equivalent and are formed by six antiparallel β strands (Figure 4.4).

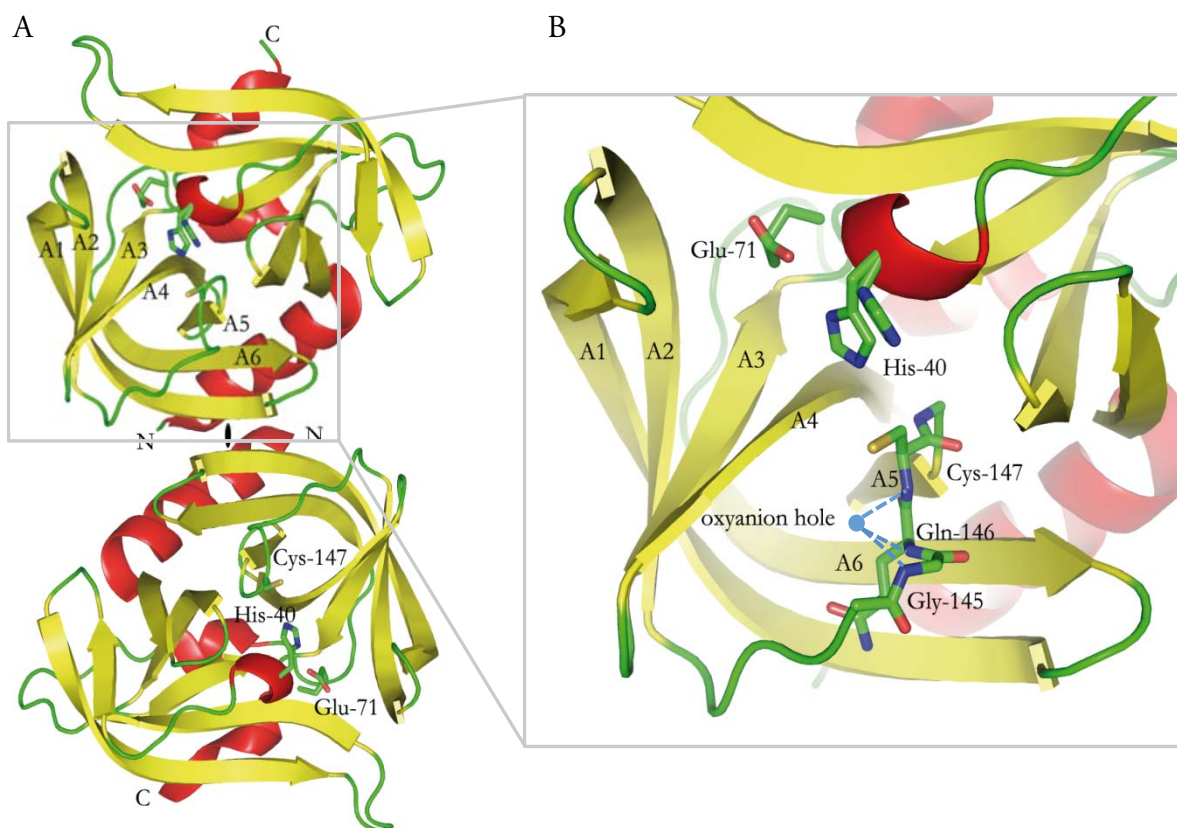


Figure 4.4 Crystal structure of EV-B93 3C protease. (A) Ribbon representation of the two molecules present in the asymmetric unit with the non-crystallographic 2-fold axis perpendicular to the plane. The protease folds into two antiparallel  $\beta$ -barrels, forming the chymotrypsin-like fold. The catalytic triad is highlighted as stick representations. (B) Active site of EV-B93 3C protease. Key residues are highlighted as stick representation. Oxyanion hole formed by the main-chain amides of Cys-147, Gln-146 and Glu-145 is indicated as a blue dot.

The structure confirms that EV-B93 3C protease adapts a chymotrypsin-like fold similarly to other picornavirus 3C proteases. The active site is placed in the shallow cleft between the two barrels and is made of a nucleophilic Cys-147 coming from the C-terminal barrel and the acid-base pair His-40 and Glu-71 located in the N-terminal barrel. There are two conformations of the His-40, one corresponds to the catalytically competent form where the sulfur atom from the Cys-147 is in plane with the imidazole ring from the His-40 and the carboxylate group of the Glu-71. In the second conformation the His-40 is rotated about  $120^\circ$  and exposed to the solvent. Substrate hydrolysis occurs through an attack of the active-site nucleophilic Cys-147 onto the carbonyl carbon of the scissile bond. The resulting covalent tetrahedral transition state is stabilized by the oxyanion hole formed by the amides of the main chain residues: Cys-147,

Gln-146 and Gly-145. In the structure the oxyanion hole is occupied by a water molecule that interacts with the amide of Cys-147 or Gln-146. EV-B93 3C protease like the other 3C proteases recognizes amino acid residues in the close proximity to the cleavage site, mostly at the P4, P3, P2, P1 and P1' positions that fit into the corresponding specific binding sub-sites (S4, S3, S2, S1 and S1'). The comparison of the HRV-A14 3C protease structure covalently bound to a peptide (Ac-LEALFQ-ethyl propionate) inhibitor containing the P6 to P1 sub-site residues [125], with EV-B93 3C protease allowed the proposal of the possible substrate binding pocket of EV-B93 3C protease. The active site is formed by the residues belonging to the  $\beta$  strands A1, A3, and A4 and the loop containing the amino acids that make the oxyanion hole.



## 4.2 Rupintrivir and compound 1 as covalent inhibitors of EV-B93 3C protease

### 4.2.1 Biochemical characterization of rupintrivir and compound 1 inhibitory activity

Rupintrivir and its orally bioavailable analogue compound 1 were developed as irreversible inhibitors of 3C protease from HRV and EV species [84, 89] (Figure 4.5).

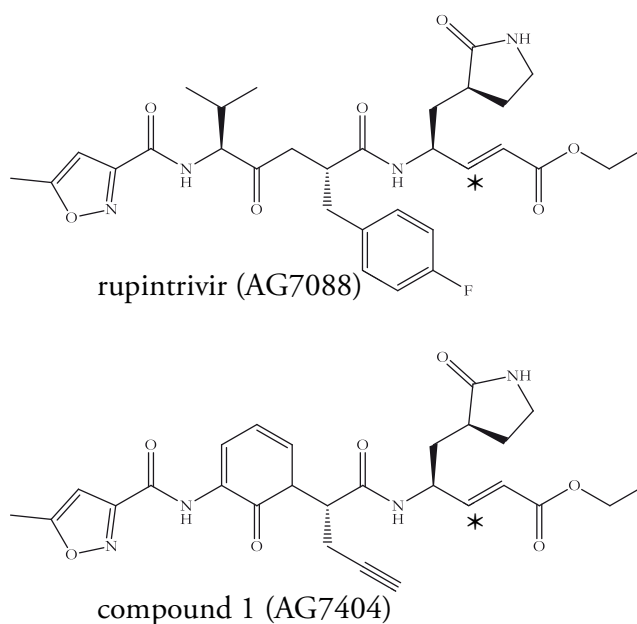


Figure 4.5 Chemical structure of rupintrivir (AG7088) and its orally bioavailable analogue compound 1 (AG7404). Asterisks indicate carbon atoms that make an irreversible covalent bond with the active-site cysteine residue of EV-B93 3C protease.

Since the proteolytic active site is highly conserved within *Enterovirus* species, both compounds should bind as well to EV-B93 3C protease. In order to verify it, the inhibitory effect of each antiviral agent was tested in the *in vitro* proteolytic activity assay (Figure 4.6).

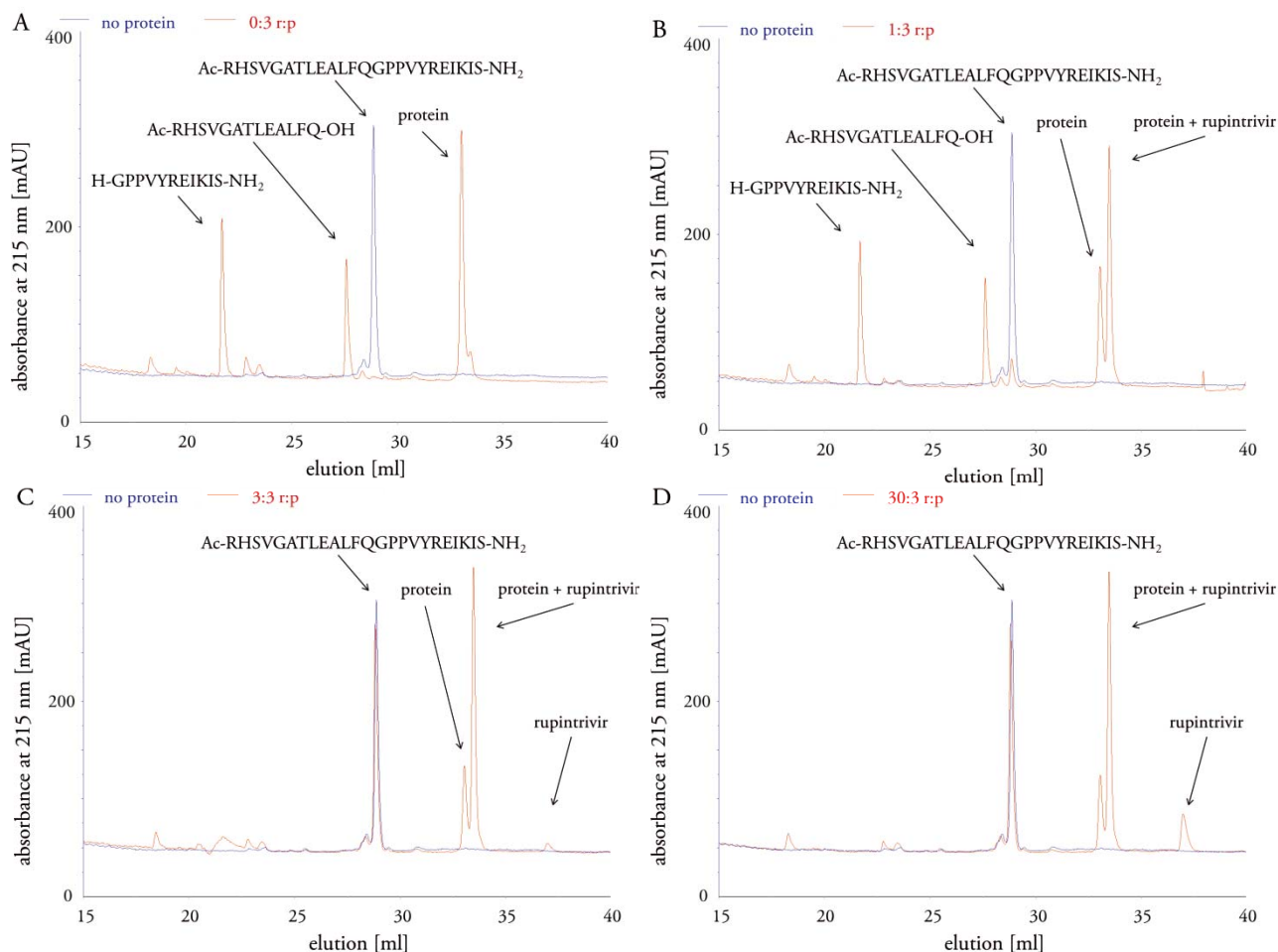


Figure 4.6 Inhibition of the *in vitro* proteolytic activity of EV-B93 3C protease by rupintrivir. Reverse-phase chromatograms show in red the products of the digestion of the peptide (Ac-RHSVGATLEALFQ↓GPPVYREIKIS-NH<sub>2</sub>) by EV-B93 3C protease (p) without rupintrivir (r) (A), with rupintrivir at 1:3 inhibitor-to-protein ratio (B) with equimolar ratio rupintrivir to protein (B) and with rupintrivir at 30:3 inhibitor-to-protein ratio (D). Chromatograms of the peptide substrate alone are shown in blue.

The native EV-B93 3C protease was incubated with three different inhibitor-to-protease molar ratios in the presence of the peptide substrate. The reverse-phase HPLC analyses showed that both rupintrivir and compound 1 efficiently inhibited EV-B93 3C protease *in vitro* (the results obtained for compound 1 are not shown, since they were very similar to those obtained for rupintrivir). No cleavage products were detected with the equimolar and 10-fold excess of the inhibitors over the enzyme (Table 4.3). These results are consistent with the prior statements regarding the potency and irreversible inhibition of these compounds towards 3C proteases

from HRVs and EVs [84, 89], and confirm their strong antiviral effect against broad spectrum of enteroviruses.

**Table 4.3 Cleavage efficiencies of the peptide substrate by EV-B93 3C protease.**

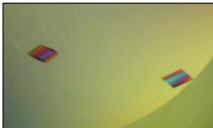
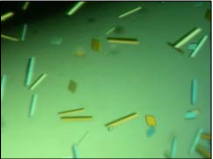
inhibitor	inhibitor-to-protein molar ratio	relative activity RA <sup>i</sup>
-	0:3	100
rupintrivir	1:3	100
	3:3	0
	30:3	0
compound 1	1:3	100
	3:3	0
	30:3	0

<sup>i</sup> relative activity is defined as the efficiency of a proteolytic reaction in the presence of inhibitor divided by the efficiency of the proteolytic reaction in the absence of inhibitor multiply by 100 %.

#### 4.2.2 Crystallization screening and crystal optimization of EV-B93 3C protease in a complex with rupintrivir and compound 1

In order to characterize the molecular interactions of rupintrivir and compound 1 with EV-B93 3C protease, the previously purified protein was incubated with both inhibitors. After the purification with the use of size exclusion chromatography, the protein solutions were used in the crystallization screenings. Many crystal hits for both complexes appeared. They were containing the thin needle-like crystals (for both compounds), the rhombohedron-like crystals (for rupintrivir) and the rod-like and rhombohedron-like crystals for the complex with compound 1. In the both cases the rhombohedron-like crystals were chosen for further optimizations. The best crystal conditions and corresponding crystal data are summarized in Table 4.4.

**Table 4.4 Crystallization conditions and corresponding crystal data for EV-B93 3C protease in complex with rupintrivir and compound 1.**

	PAC	initial condition	optimized condition	crystal shape	resolution [Å]	space group	cell dimension
rupintrivir	PAC2	20 % PEG8000	25 % PEG8000		1.50	P2 <sub>1</sub>	a=39.00
	E3	0.2 M MgCl <sub>2</sub> 0.1 M Tris pH 8.5	0.25 M MgCl <sub>2</sub> 0.1 M Tris pH 8.5				b=63.91
							c=66.36 β=90.43
compound	PAC4	2.4 M (NH <sub>4</sub> )HPO <sub>4</sub>	1.8 M (NH <sub>4</sub> )HPO <sub>4</sub>		1.32	P2 <sub>1</sub>	a=39.04
	E4	0.1 M Tris pH 8.5	0.1 M Tris pH 8.3				b=64.45 c=68.74 β=90.81

#### 4.2.3 Structure determination of EV-B93 3C protease in complex with rupintrivir and compound 1 and data analysis

The structure of both complexes: EV-B93 3C protease-rupintrivir and EV-B93 3C protease-compound 1 were solved with molecular replacement using previously obtained structure of EV-B93 3C protease as a search model (PDB code: 3Q3X). Both complexes of EV-B93 3C protease with rupintrivir and compound 1 were crystallizing in the monoclinic P2<sub>1</sub> space group giving the high-resolution diffraction data of 1.50 Å and 1.32 Å, respectively. The structures were refined to a final R<sub>work</sub> of 16.6 % and R<sub>free</sub> of 19.8 % for the complex with rupintrivir and to R<sub>work</sub> of 12.5 % and R<sub>free</sub> of 17.0 % for the complex with compound 1 (see Table 4.5 for refinement statistics). The final model for EV-B93 3C protease-rupintrivir and for EV-B93 3C protease-compound 1 encompasses all amino acids except the first N-terminal and the last three C-terminal (including His<sub>6</sub> tag) residues. One molecule of rupintrivir and one molecule of compound 1 per protein chain were unequivocally and precisely defined for the complexes of EV-B93 3C protease with rupintrivir and compound 1, respectively. In the both cases the asymmetric unit contains two protein molecules. The refined structure of EV-B93 3C protease-rupintrivir has 375 water molecules, 2 Mg<sup>2+</sup> ions and 1 Cl<sup>-</sup> ion, while the refined structure of EV-B93 3C protease-compound 1 has 543 water molecules, 8 ethylene glycol molecules, 3 HPO<sub>4</sub><sup>2-</sup> ions and 2 NH<sub>4</sub><sup>+</sup> ions.

**Table 4.5 Data collection and refinement statistics for EV-B93 3C protease-rupintrivir and EV-B93 3C protease-compound 1 complex crystals.**

parameter	EV-B93 3C protease-rupintrivir	EV-B93 3C protease-compound 1
resolution range [Å]	30.0-1.50 (1.55-1.50)	39.0-1.32 (1.39-1.32)
space group	P2 <sub>1</sub>	P2 <sub>1</sub>
cell dimensions [Å, °]	a=39.00 b=63.91 c=66.39 β=90.43	a=39.04 b=64.45 c=68.74 β=90.81
no. of observed reflections	194865	268892
no. of unique reflections	52513 (5196)	79978 (11666)
completeness [%]	100 (100)	100 (100)
mean multiplicity	3.7 (3.6)	3.4 (3.3)
R <sub>merge</sub>	0.063 (0.488)	0.085 (0.508)
<I>/σ<I>	17.7 (2.5)	5.2 (1.5)
R <sub>work</sub>	0.166	0.125
R <sub>free</sub>	0.198	0.170
no. of protein atoms (non-H)	3011	3322
no. of water molecules	375	543
no. of hetero compounds	2 rupintrivir molecules, 2 Mg <sup>2+</sup> ions, Cl <sup>-</sup> ion	2 compound 1 molecules, 8 ethylene glycol molecules, 3 HPO <sub>4</sub> <sup>2-</sup> ions, 2NH <sup>+</sup> ions
RMSD for bond length [Å]	0.008	0.012
RMSD for bond angles [°]	1.41	1.60
mean B value [Å <sup>2</sup> ]	12.21	9.29
mean B value for inhibitor molecules [Å <sup>2</sup> ]	16.8, 22.0	20.2, 19.7
Ramachandran analysis: preferred/allowed/outliers [%]	97.26/2.74/0.0	89.1/9.6/1.3
PDB accession code	3RUO	3Q3Y

Values in parenthesis correspond to the outer resolution shell.

$R_{merge} = \sum_h \sum_i |I_{i,h} - \langle I_h \rangle| / \sum_h \sum_i I_{i,h}$ , where  $I_{i,h}$  is the  $i$ th-intensity measurement of reflection  $h$  and  $\langle I_h \rangle$  is the average intensity for multiple measurements.

$R_{work}$  and  $R_{free} = \sum_h ||F_o| - |F_c|| / \sum |F_o|$ .  $R_{free}$  was calculated for 5% of the reflections not used for refinement.

The electron density allowed for the precise building of one molecule of rupintrivir and compound 1 per protein. In both cases, the electrophilic  $\beta$ -carbon (Figure 4.5, asterisk) is covalently bound to the Cys-147 active site after its Michael addition, forming a stable tetrahedral adduct and resulting in an irreversible inactivation of the protease. Rupintrivir binds to EV-B93 3C protease in a partially extended conformation with its peptidomimetic backbone making antiparallel  $\beta$ -sheet-type hydrogen bonds with the part of the solvent-exposed  $\beta$  strand, A4, of the protein. The P1' carbonyl oxygen of the ethyl ester is placed above the oxyanion hole formed by the amide groups of the Cys-147, Gln-146, Gly-145 but makes a hydrogen bond only with the latter (Figure 4.7). The P1 part of rupintrivir is deeply inserted between the  $\beta$ -strand A4 and the loop formed by the Thr-142, Arg-143 and Ala-144. The P1 lactam ring interacts with the Thr-142 and His-161, mimicking the recognition site for glutamine, which is highly conserved in the 3C cleavage sequences. The P2 1-fluorobenzen-4-yl ring forms a stacking interaction with the His-40 and Glu-71 and its fluorine atom interacts with the Thr-130, Arg-39 and Leu-127. The valine side chain from the P3 part is solvent exposed, thus does not interact with the protein. The P4 part of rupintrivir lies in a deep groove formed by the  $\beta$ -strands A4, A3 and A1 and interacts with the protein residues Leu-125, Gly-128, Asn-165, Glu-168 and Phe-170.

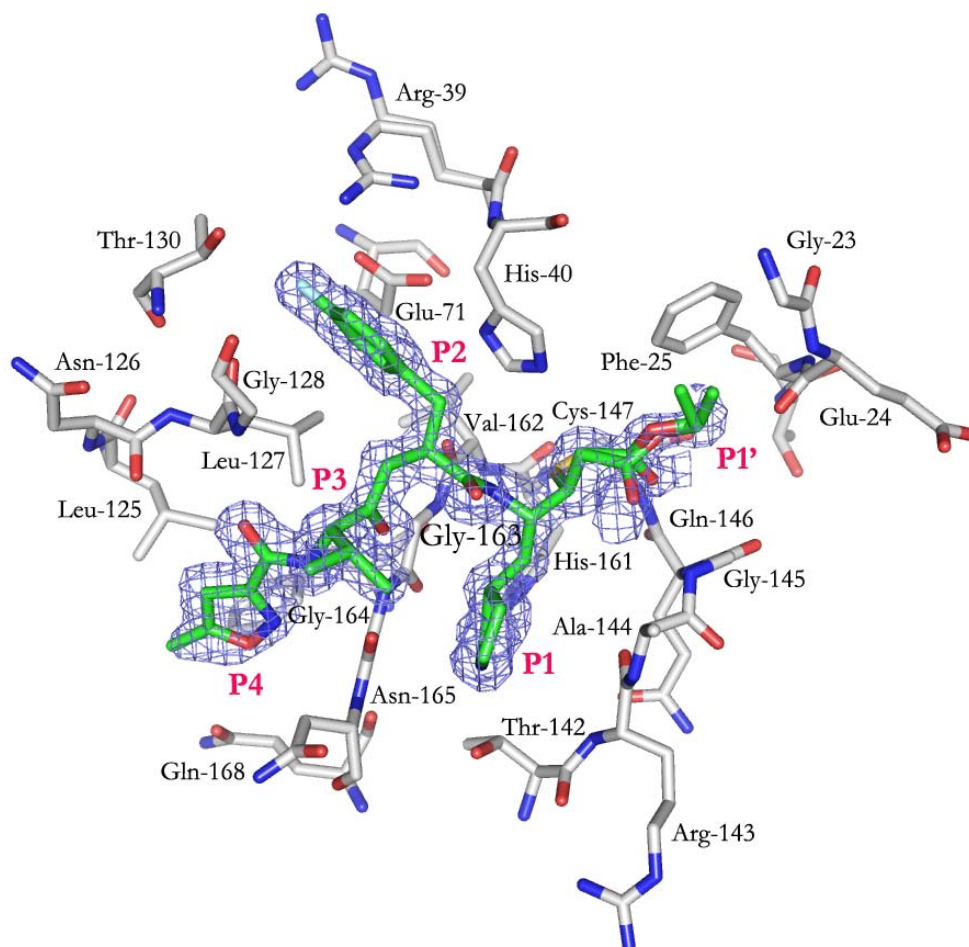


Figure 4.7 Refined structure of EV-B93 3C protease in complex with rupintrivir. The inhibitor is represented as sticks, with its  $2F_o-F_c$ -weighted electron density contoured at  $1.0 \sigma$  and represented as a blue mesh. EV-B93 3C protease residues interacting with rupintrivir are shown as sticks and are labeled.

The binding mode of compound 1 to EV-B93 3C protease is very similar to that of rupintrivir (Figure 4.8). The differences between the two antivirals lie in the least-conserved P2 and P3 positions. In the P2 position the 1-fluorobenzen-4-yl group is substituted by the 2-propynyl residue, which stacks between the His-40, Glu-71 and Leu-127. In the case of the P3 position, the valine residue is replaced by the ring, which makes two main-chain hydrogen bonds with the Gly-163 and Gly-164.

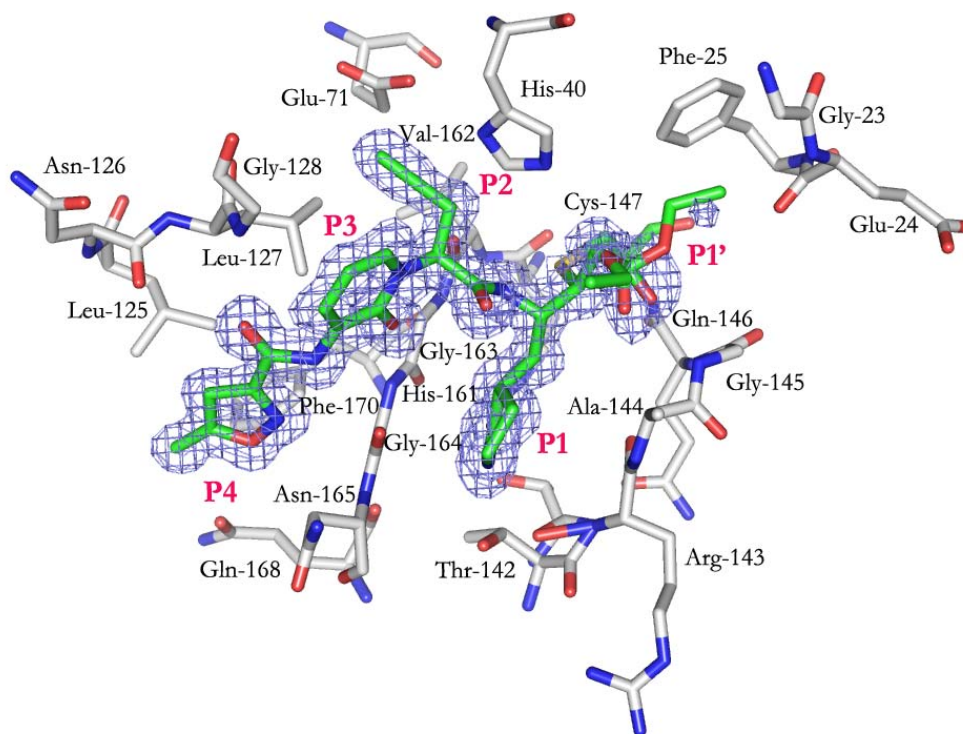


Figure 4.8 Refined structure of EV-B93 3C protease in complex with compound 1. The inhibitor is represented as stick, with its  $2F_o-F_c$ -weighted electron density contoured at  $1.0 \sigma$  and represented as blue mesh. EV-B93 3C protease residues interacting with compound 1 are shown as sticks and are labeled.

Rupintrivir generates slightly more interactions with the binding pocket of EV-B93 3C protease than compound 1, thus should exhibit higher potency. Indeed, rupintrivir appeared to inhibit EV-B93 replication in infected RD cells with a mean  $EC_{50}$  of 33 nM, while the  $EC_{50}$  for compound 1 was 93 nM [126].



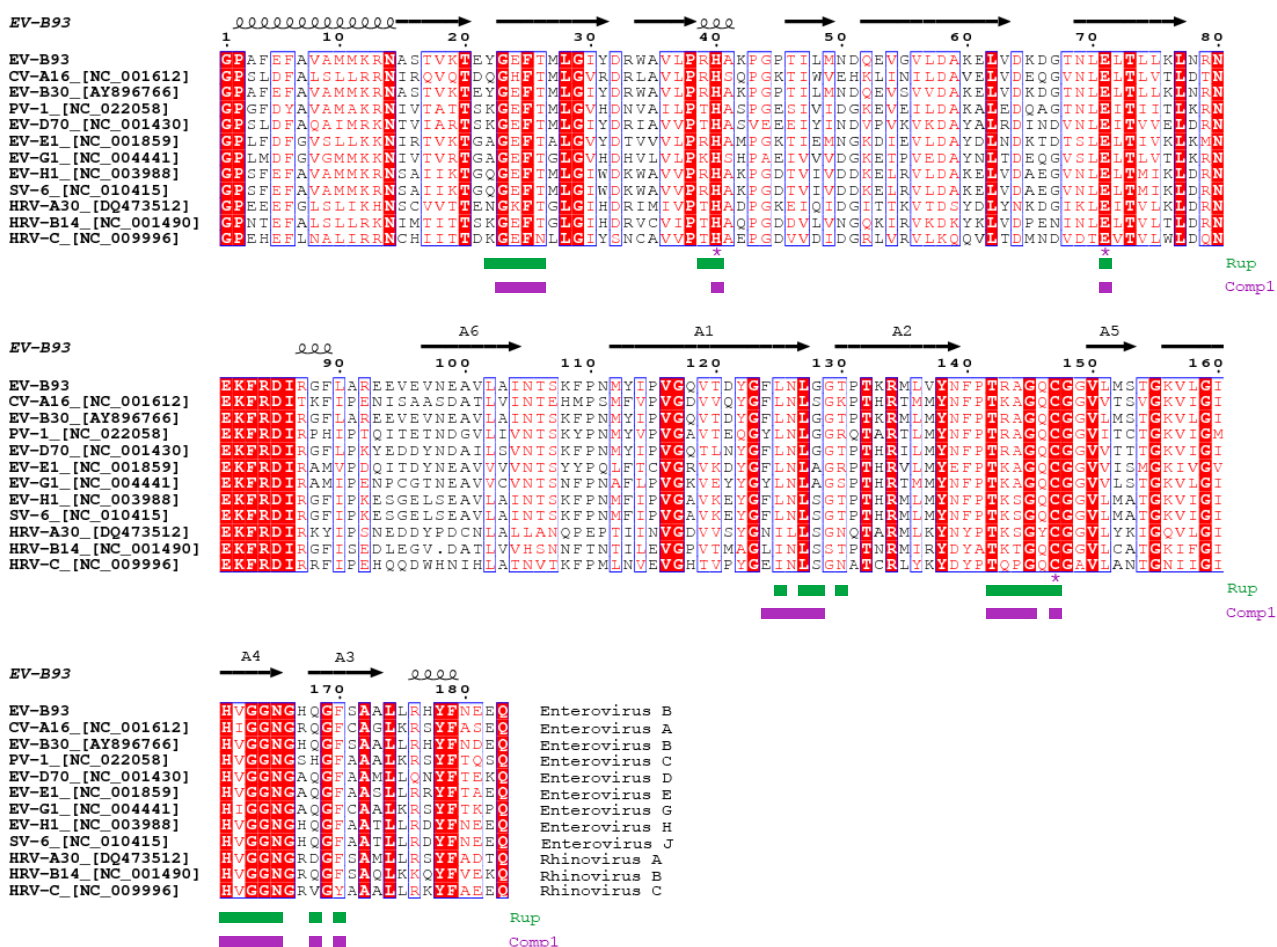


Figure 4.9 Multiple sequence alignment of 3C proteases for 11 enteroviruses representing the currently known species diversity of the *Enterovirus* genus (without Enterovirus F). The sequence names indicate the names and accession IDs of the viruses, and the corresponding species are indicated at the ends of the sequences. Invariant residues are white with a red background; conserved residues are shown in red font. The residues involved in the interactions are marked below as green and magenta bars for rupintrivir and compound 1, respectively. Catalytic residues Cys-147, His-40 and Glu-71 are indicated by violet asterisks.

A noticeable feature of the binding pocket is the conservation of EV-B93 3C protease residues interacting with rupintrivir and compound 1 within all 11 enterovirus species representing the entire genetic diversity of this genus (except Enterovirus F). Most of the residues are identical or physico-chemically similar (Figure 4.9). The results with EV-B93 3C protease indicate that these compounds may be valuable antivirals against other enteroviral species. The high level of conservation among EVs of the residues forming the 3C protease binding pockets for rupintrivir and compound 1, and the broad-spectrum antiviral activity of these compounds *in vitro* reinforce their potential as excellent candidates for developing potent antivirals against all

EVs [91]. The process of the optimization of rupintrivir and compound 1 should be continued, since in the studies to control natural rhinovirus infection by 3C protease inhibitors, both inhibitors showed unsatisfactory performance and were therefore excluded from further clinical development [127]. The analysis of multiple alignment also suggests that the level of conservation of the residues forming the substrate binding pocket can be useful in the process of designing novel antiviral compounds against new emerging enteroviruses.

### 4.3 Small covalent inhibitors of EV-B93 3C protease as a starting point for development of potent drugs

Despite the failure of rupintrivir and compound 1 in clinical trials, it is worth developing new covalent binders of enteroviral 3C proteases. Instead of optimizing already existing inhibitors, the new covalent low-molecular inhibitors may be selected through the screening of commercial or in-house designed libraries. The hits will bind with a high probability to the S1 sub-site of 3C protease due to its close proximity to the nucleophilic cysteine in the active site. When the binding is confirmed, the compound may serve as a starting point for further development. The compound may grow into adjacent pockets, resulting in a more potent inhibitor. The same path was followed by the collaborators from Freie Universität in Berlin. They were succeeded in selecting a series of low-molecular compounds able to inhibit 3C protease from CV-B3. Three molecules were sent to us for functional and structural studies with EV-B93 3C protease, a very close analog of CV-B3 (97 % identity).

#### 4.3.1 Biochemical characterization of NZN, NZO and DB5\_60 inhibitory activity

Chemical synthesis of three low-molecular-weight compounds was done by Daniel Becker from Freie Universität Berlin (Germany). They were designed to be irreversible covalent inhibitors of enteroviral 3C proteases. These inhibitors possess a P1/P1' – Michael-acceptor moiety, which is an  $\alpha,\beta$ -unsaturated ester (NZO), an  $\alpha,\beta$ -unsaturated amide (NZN) or a vinyl sulfonate ester (DB5\_60), which can covalently link to the Cys-147 in the active site. They may exist in two tautomeric forms: hydroxypyrazol and pyrazolone structure (Figure 4.10).

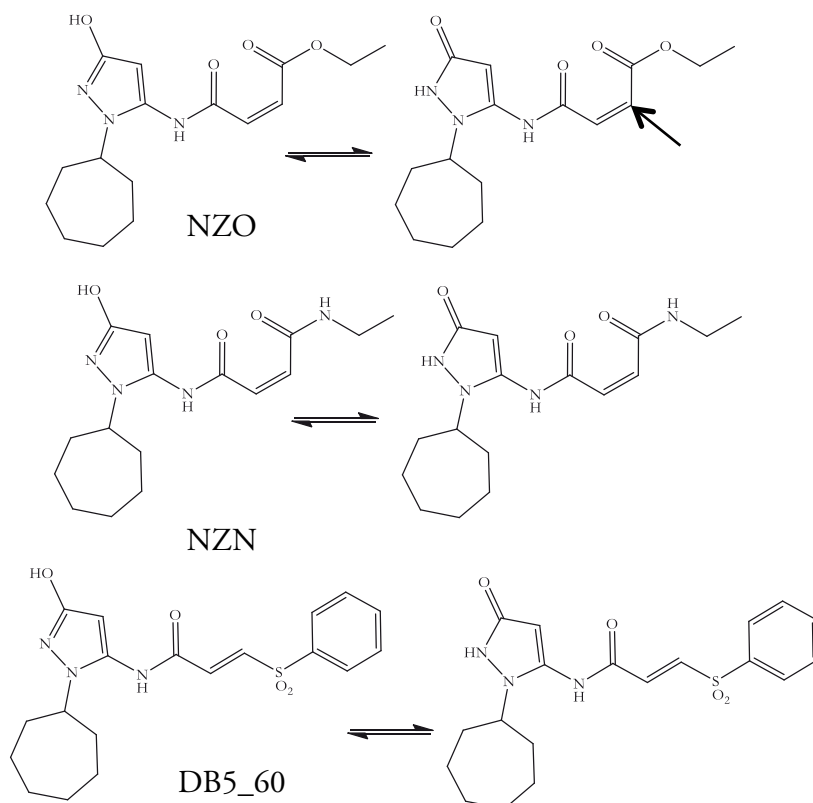


Figure 4.10 Chemical structures of three potentially irreversible inhibitors of EV-B93 3C protease. The compounds remain in the tautomeric equilibrium of hydroxypyrazol form (on the left side of each equilibration) and pyrazolone form (on the right side of each equilibration).

The stability of each compound was determined in two different buffers (30 mM Tris pH 8.0; 50 mM NaCl and 50 mM HEPES pH 7.4; 50 mM NaCl). NZO, NZN and DB5\_60 were diluted in each buffer and loaded on the reverse-phase HPLC column: immediately or after 3h-incubation in 37 °C. The chromatograms showed that the NZO compound was not stable in neither of the buffers while the NZN molecule appeared to be more stable in HEPES buffer. In the case of the sulfonate ester DB5\_60 the compound was stable in both conditions (Figure 4.11). Since all previous experiments were performed in Tris buffer, it was chosen for further activity tests.

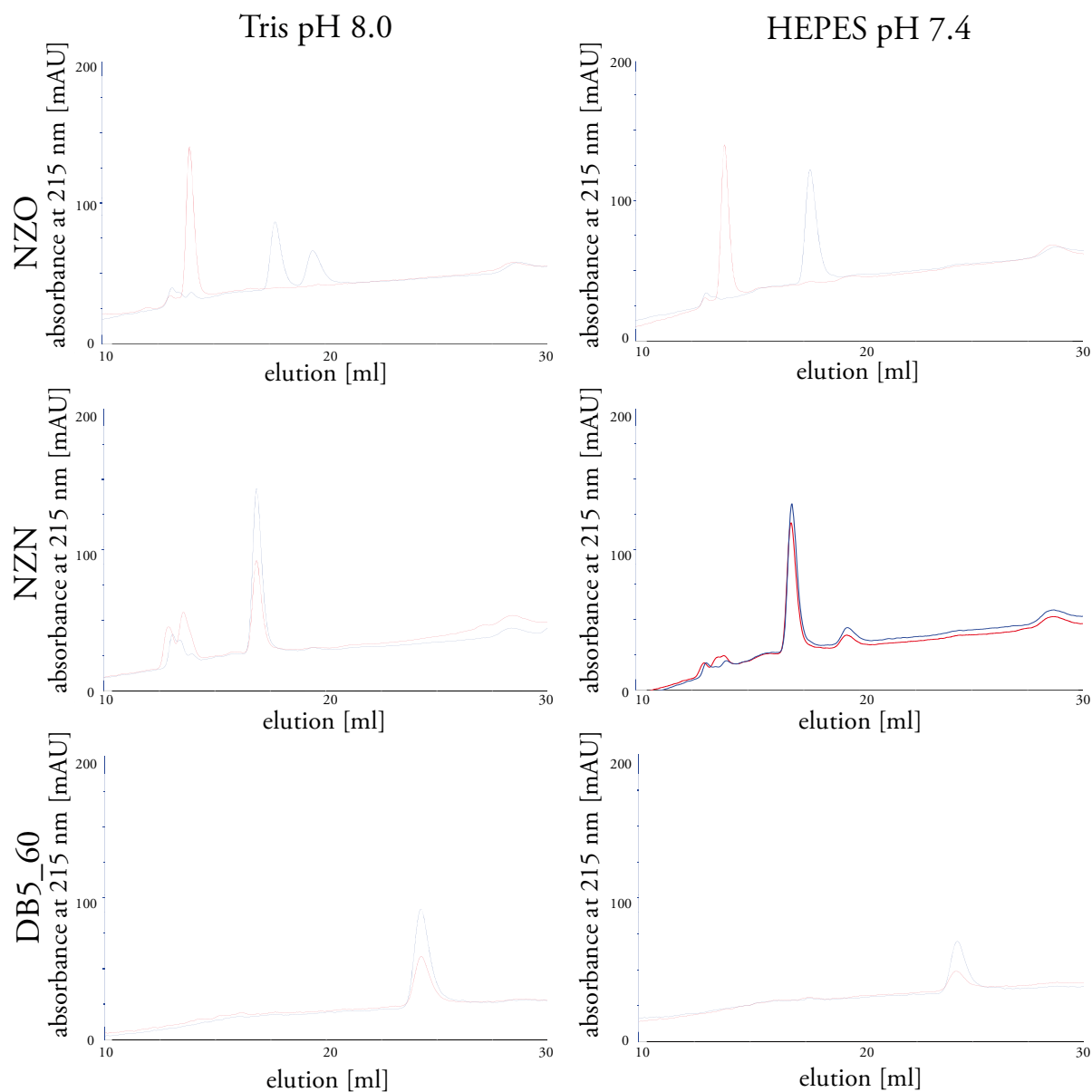


Figure 4.11 Stability of NZO, NZN and DB5\_60 compounds in Tris buffer pH 8.0 and HEPES buffer pH 7.4. Chromatograms of compounds before incubation are shown in blue and after 3h-incubation at 37 °C are shown in red. The concentration of each compound was 50  $\mu$ M.

In order to test the inhibitory activity of the NZO, NZN and DB5\_60 compounds, the *in vitro* proteolytic activity assays were performed. Native EV-B93 3C protease was incubated with three different inhibitor-to-protease molar ratios in the presence of the peptide substrate. For the compounds NZO and NZN both buffers were used while in the case of the DB5\_60

compound the reactions were performed only in Tris buffer. The obtained results are gathered in Table 4.6.

**Table 4.6 Cleavage efficiencies of substrate peptide by EV-B93 3C protease.**

buffer	inhibitor	inhibitor-to- protease ratio	relative activity RA [%]
30 mM Tris pH 8.0	-	0:1	100
50 mM NaCl	NZO	1:1	0
		1:3	0
		1:10	0
	NZN	1:1	~60
		1:3	~30
		1:10	0
	DB5:60	1:1	0
		1:3	0
		1:10	0
50 mM HEPES pH7.4	-	0:1	100
50 mM NaCl	NZO	1:1	8
		1:3	5
		1:10	0
	NZN	1:1	~90
		1:3	~80
		1:10	~60
	DB5_60	1:1	nd <sup>i</sup>
		1:3	nd
		1:10	nd

<sup>i</sup>nd – not determined

Despite the fact that the NZO compound was less stable than the NZN molecule, it showed better inhibitory activity in both buffers. It may be due to higher reactivity of  $\alpha,\beta$ -unsaturated ester moiety which on the other hand is more susceptible to hydrolysis when compared to  $\alpha,\beta$ -unsaturated amide functionality. The most stable compound DB5\_60 showed very good inhibitory activity. According to the literature [128] vinyl sulfones are 10-fold more reactive than  $\alpha,\beta$ -unsaturated ester-based inhibitors against human rhinovirus 3C protease. They may react not only with the cysteine active side but also modify nonspecifically amino acid side chains in proteins.

### 4.3.2 Crystallization screening and crystal optimization of EV-B93 3C protease in complex with NZO, NZN and DB5\_60

Native EV-B93 3C protease was incubated with a 12-fold molar excess of each compound (NZO, NZN and DB5\_60). After purification with the use of size exclusion chromatography, the protein samples were used for crystallization screening. Previous trials with lower amount of the NZO and NZN compounds did not succeed in producing protein-complex crystals. The DB5\_60 compound was insoluble in high concentration and caused the precipitation of the protein. For that reason higher amount of the protein was used to get enough protein for the crystallization purpose. Crystallization screenings resulted in many conditions with the needle-like crystals (for all three compounds) and several conditions with the rectangular-like crystals (for the NZO and NZN molecules). Both needle-like and rectangular-like crystals were used for further optimization (Figure 4.12).

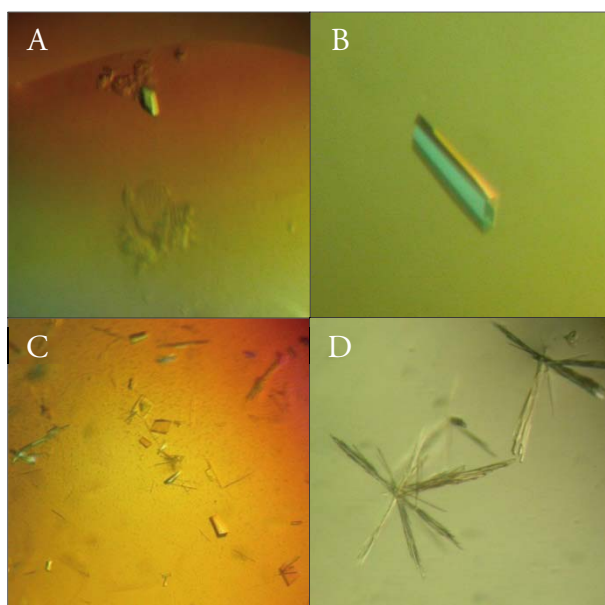
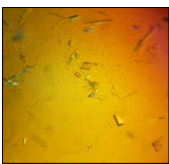




Figure 4.12 Crystallization of EV-B93 3C protease with NZN compound: initial condition (A) and after optimization (B); with NZO compound (C) and with DB5\_60 compound (D).

The crystals of EV-B93 3C protease-NZO complex were instable and difficult to optimize. Usually both types of the crystals (the needle-like and the rectangular-like) grew together which hampered the fishing process. The complex of NZN and EV-B93 3C protease gave well-defined large crystals, which were easy to handle. Precipitation of EV-B93 3C protease occurred in the presence of DB5\_60 most probably due to unspecific covalent modification of

the protein with highly hydrophobic moieties. The protein that did not precipitate after incubation with DB5\_60 is believed to stay unmodified, since it crystallized only in the needle-like form. The best crystal conditions were summarized in Table 4.7.

**Table 4.7 Crystallization conditions and corresponding crystal data for EV-B93 3C protease in complex with NZO, NZN and DB5\_60.**

	PAC	initial condition	optimized condition	crystal shape	resolution [Å]	space group	cell dimension
NZO	PAC3	25 % PEG3350	16 % PEG3350		1.86	P2 <sub>1</sub>	a=57.28 b=51.92 c=59.28 β=103.69
	G7	0.2 M NH <sub>4</sub> Ac	0.2 M NH <sub>4</sub> Ac				
		0.1 M Bis-Tris 6.5	0.1 M Bis-Tris 6.0				
NZN	PAC3	25 % PEG3350	20 % PEG3350		1.73	P2 <sub>1</sub>	a=57.46 b=52.33 c=62.27 β=103.73
	G7	0.2 M NH <sub>4</sub> Ac	0.2 M NH <sub>4</sub> Ac				
		0.1 M Bis-Tris 6.5	0.1 M Bis-Tris 6.0				
DB5_60	PAC4 C7	3.5 M sodium formate 0.1 M Bis-Tris propane pH 7.0	2.4 M sodium formate 0.1 M Bis-Tris propane pH 7.0		2.85	P2 <sub>1</sub>	a=39.03 b=57.30 c=84.84 β=94.43

#### 4.3.3 Structure determination of EV B93-3C protease in a complex with NZN

Many data sets were collected for the crystal complexes of EV-B93 3C protease and the NZO, NZN and DB5\_60 molecules. Only one crystal structure showed the presence of additional, however incomplete, electron density in the active side of only one protein molecule in the asymmetric unit. The structure of EV-B93 3C protease-NZN complex was solved at 1.73 Å using previously obtained structure of native EV-B93 3C protease as a search model (PDB code: 3Q3X). The complex was crystallizing in the monoclinic P2<sub>1</sub> space group with the cell parameters different from the ones obtained for the complexes with rupintrivir and compound 1. The structures were refined to a final R<sub>work</sub> of 16.1 % and R<sub>free</sub> of 20.1 % (see Table 4.8 for refinement statistics). The final model for the EV-B93 3C protease-NZN

complex contains all amino acids except the last two C-terminal (including His<sub>6</sub> tag) residues. The asymmetric unit has two protein molecules A and B, where in the chain B one molecule of NZN was defined. The refined structure of the EV-B93 3C protease-NZN complex contains 255 water molecules.

**Table 4.8 Data collection and refinement statistics for EV-B93 3C protease-NZN complex crystals.**

parameter	EV-B93 3C protease-NZN
resolution range [Å]	60.49-1.73 (1.77-1.73)
space group	P2 <sub>1</sub>
cell dimensions [Å, °]	a=57.46 b=52.33 c=62.27 β=103.73
no. of observed reflections	121071
no. of unique reflections	36869 (2654)
completeness [%]	97.86 (97.1)
Mean multiplicity	3.3 (3.3)
R <sub>merge</sub>	0.052 (0.724)
<I>/σ<I>	13.35 (2.05)
R <sub>work</sub>	0.161
R <sub>free</sub>	0.201
no. of protein atoms (non-H)	3101
no. of water molecules	255
no. of hetero compounds	1 NZN molecule
RMSD for bond length [Å]	0.02
RMSD for bond angles [°]	2.033
mean B value [Å <sup>2</sup> ]	31.6
mean B value for inhibitor molecules [Å <sup>2</sup> ]	67.2
Ramachandran analysis:	96.35/2.81/0.84
preferred/allowed/outliers [%]	

Values in parenthesis correspond to the outer resolution shell.

$R_{merge} = \sum_h \sum_i |I_{i,h} - \langle I_h \rangle| / \sum_h \sum_i I_{i,h}$ , where  $I_{i,h}$  is the  $i$ th-intensity measurement of reflection  $h$  and  $\langle I_h \rangle$  is the average intensity for multiple measurements.

$R_{work}$  and  $R_{free} = \sum_h ||F_o| - |F_c|| / \sum |F_o|$ .  $R_{free}$  was calculated for 5% of the reflections not used for refinement.

The quality of the electron density was sufficient to build one molecule of NZN in the chain B of the asymmetric unit. In the chain A the additional electron density in the active site was very



poor, therefore the placement of the NZN compound failed. As a result of a Michael addition, the electrophilic  $\beta$ -carbon of the NZN compound (Figure 4.10, arrow) was covalently linked to the nucleophilic Cys-147, forming a stable tetrahedral adduct. The NZN molecule is inserted between the solvent-exposed  $\beta$ -strand, A4, and the loop formed by the Thr-142, Arg-143 and Ala-144 and occupies the S1 and S1' sub-pockets of the protein (Figure 4.13 A and B). The P1' carbonyl oxygen of the ethyl amide is oriented towards the oxyanion hole and makes a hydrogen bond with the amide groups of the Cys-147, Gln-146 and Gly-145. The interaction between ethyl amide group of the NZN compound and the His-40 from the catalytic triad of the protein is mediated by a water molecule. The nitrogen atom from the P1-amide bond interacts with the Arg-143, Gly-164 and a water molecule that mediates the interaction with the His-161. Plausible hydrogen bonds were identified between the P1-heterocyclic ring and both the Gly-164 and Thr-142 as well as a water molecule that supports the interaction with the Ala-144. Additionally, delocalized electrons from the P1-heterocyclic ring make the  $\pi$ - $\pi$  interaction with the peptide bond of the Arg-143 – Ala-144 on the one side and the Gly-164 – Asn-165 on the other one. The cycloheptane moiety from the P1 part is solvent exposed, thus does not interact with the protein. The third amino acid from the catalytic triad Glu-71 is not involved in the binding of NZN.

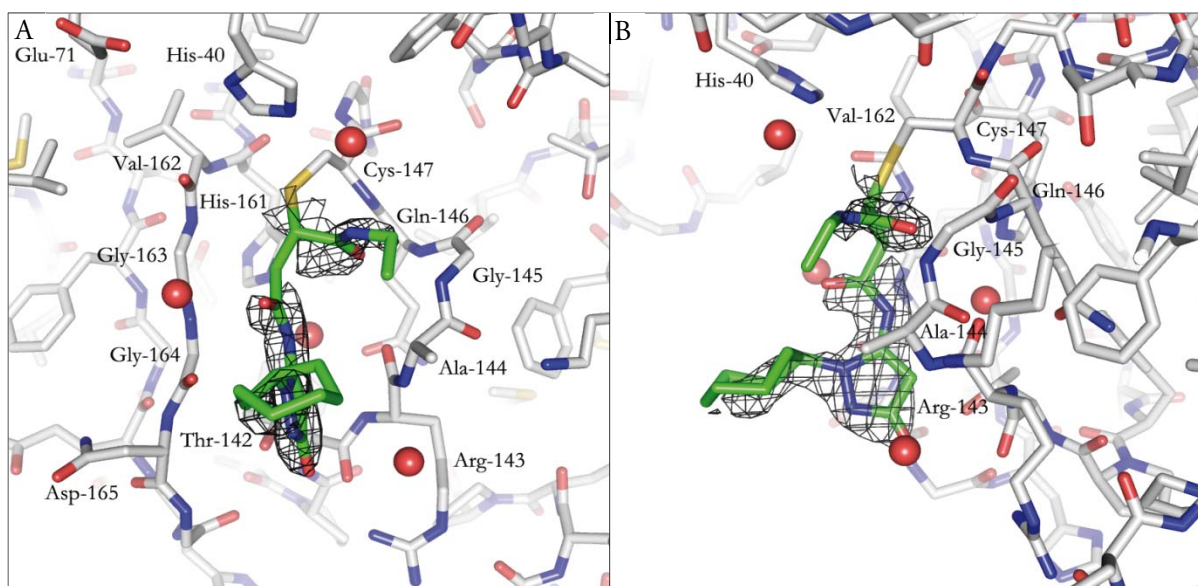


Figure 4.13 Refined structure of EV-B93 3C protease in complex with NZN. The inhibitor is represented as sticks, with its  $2F_o - F_c$ -weighted electron density contoured at  $1.0 \sigma$  and represented as black mesh. EV-B93 3C protease residues in close proximity to NZN compound are labeled. Water molecules are represented as red spheres. (A) top view, (B) side view.

The electron density corresponding to the NZN compound is not fully defined and suggests a weaker binding to the protein when comparing with rupintrivir and compound 1. It is not surprising since NZN is twice smaller molecule (598 Da and 320 Da for rupintrivir and NZN, respectively) and occupying only two sub-pockets of the active site, generates substantially fewer interactions with EV-B93 3C protease. Better defined electron density of the NZN compound is observed for the regions with more intramolecular interactions, i.e. the covalent bonding with the Cys-142, the interaction of the carbonyl oxygen with amides from the oxyanion hole and the heterocyclic ring with its hydrogen bonds and the  $\pi$ - $\pi$  interactions. On the other hand, the ethyl group from the NZN amide bond, two carbon atoms from the central part the molecule and the cycloheptane residue have the least visible electron density suggesting its high flexibility. Additionally, the interaction between the oxygen atom from the heterocyclic ring of the NZN compound and both the Thr-142 hydroxyl group from the side chain and the Thr-142 carbonyl oxygen from the main chain does not allow the prediction which tautomer hydroxypyrazol or pyrazolone is the most likely form to interact with the protein.

The NZN compound is small and can serve as a starting point for further modifications. There is not much space for additional alterations of the molecule in the S1 pocket. However, there is an empty space below the internal amide bond, which is occupied by a water molecule and mediates the interaction with the His-161 of EV-B93 3C protease. Introduction of a hydrogen-acceptor moiety in this position may result in an increase of the NZN affinity. The growth of the molecule towards the P2 pocket can be problematic, since the most suitable for the modification carbon atom ( $\beta$  carbon from the  $\alpha,\beta$ -unsaturated amide moiety) is a part of the reactive double bond. An additional substitution in this position may result in the reduction of the reactivity due to a steric hindrance. It is also possible to substitute the ethyl residue from the  $\alpha,\beta$ -unsaturated amide moiety by a more bulky group, thus enlarge the molecule towards the P1' sub-pocket and subsequently improve the NZN potency.

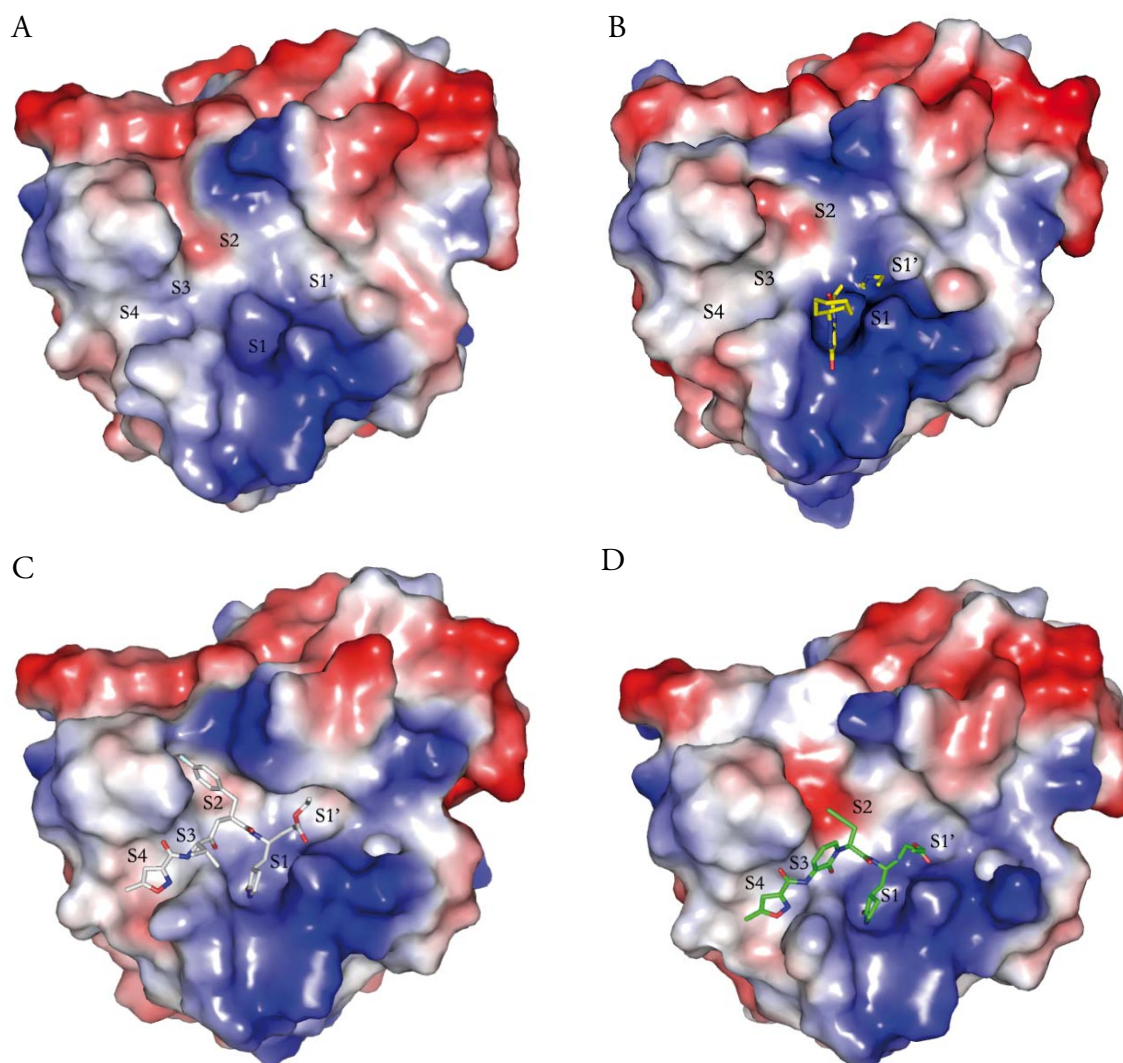


Figure 4.14 Proteolytic active site of EV-B93 3C protease alone (A) and in complex with NZN (B), rupintrivir (C) and compound 1 (D). Protein surface is colored according to the electrostatic potentials with red and blue indicating negative and positive charges, respectively. Inhibitors are represented as sticks: yellow for NZN, grey for rupintrivir and green for compound 1. The sub-pockets of the active sites are labeled.

The structure of the unliganded EV-B93 3C protease is similar to that of in complex with the covalent inhibitors. However, the S2 sub-pocket appeared to be more flexible which can be clearly noticed in the complex with rupintrivir (Figure 4.14 C). Upon binding of this antiviral, the S2 sub-pocket gets tighter, supporting its accommodation in the active site. In the case of the structures of the native protein and in the complex with NZN and compound 1, the S2 region remains open. This suggests its relative malleability towards different types of residues

that can be present in the position P2 (Figure 4.14 A, B and D). The same observation was also noticed in the case of 3C proteases from EV-A71 and CV-A16, where the S2 position is able to accommodate more variation in amino acid residues [129]. In contrast, the P1, P3 and P4 sub-pockets retain its shape within the studied structures, which may suggest more stringent structural requirements towards the size of accepted residues in the corresponding sub-pockets.

The electrostatic surface of EV-B93 3C protease active site is generally maintained upon binding of the low molecular weight NZN inhibitor, while after attachment of the larger compounds, rupintrivir or compound 1, it changes mainly in the S2 sub-pocket. The positively charged S1 cavity seems to be the most conserved region within the whole substrate-binding site, since its electrostatic surface does not change much between unbound and bound forms. The high conservation of the S1 region was previously observed [81], thus confirming the above-mentioned statement.

Covalent modification of EV-B93 3C protease by irreversible inhibitors led to the structures with a relatively higher resolution when comparing to the structures of the native protein. This happens due to tighter, more compact folding of the enzyme in bound state. For EV-B93 3C protease this relation is not so evident, but becomes clearer in the case of the other homologous 3C proteases within enteroviral species. For example EV-A71 3C protease alone gave the structure at 2.99 Å-resolution, while the same protein in complex with rupintrivir formed crystals that diffracted up to 1.0 Å [65]. Similar observations were described for EV-D68, CV-B3 and polio virus 3C proteases [64, 65, 95, 130].

#### 4.4 Discovery of non-covalent inhibitors of EV-B93 3C protease using fragment-screening methodology

Almost all known inhibitors of enteroviral 3C proteases, cited in the literature to date, possess electrophilic moiety that forms a covalent interaction with the cysteine in the active pocket. Despite its high potency and selectivity, there is a possibility of additional modifications of host proteins which may lead to cell toxicity. In order to avoid side reactions caused by reactive electrophilic groups, non-covalent inhibitors of enteroviral 3C proteases should be found. The aim of this study was to identify the first non-covalent modulators of EV-B93 3C protease, which will be able to efficiently inhibit its protease activity.

The use of fragments in a discovery of enzyme inhibitors usually involves four steps: preparation of a fragment library, screening of fragments, validation of hits using biophysical methods and elaboration of the best molecules. In this study the fragment library was already prepared and was used directly to perform the STD NMR screening followed by the competitive STD NMR screening with the use of covalent inhibitor rupintrivir. The best hits were validated with a thermal shift assay (TSA) and a surface plasmon resonance (SPR). The most promising compounds were then characterized biochemically in the proteolytic activity assay. In parallel crystallization experiments were carried out in order to obtain crystal structures of protein-ligand complexes. In the final step, the best compounds were optimized to give more potent inhibitors (Figure 4.15).

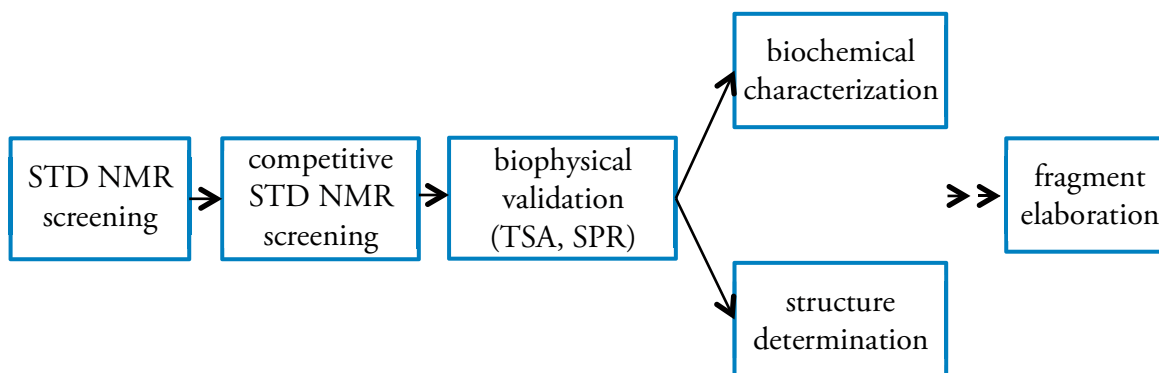


Figure 4.15 The workflow of the study.

#### 4.4.1 Library design

The fragment library suitable for the STD NMR screening was designed and prepared by Michael Goldflam in the laboratory of prof. Ernest Giralt (IRB Barcelona). The first and the second part of the library together consist of approximately 527 molecules selected from the initial set of over 1.2 million compounds and collected from different vendors. The chosen compounds were soluble in aqueous solutions, exhibited a molecular weight between 150 and 300 Da and possessed drug-like properties. The library was composed of structurally and chemically diverse molecules that were able to cover chemical space in an optimal manner.

#### 4.4.2. STD NMR-based fragment screening

Saturation-transfer NMR spectroscopy was introduced in 1999 by Meyer [131, 132] and has been used for many years to characterize interaction between protein and ligand in solution under physiological conditions. In a typical STD experiment, 1D  $^1\text{H}$  NMR spectra are recorded for a ligand in the presence of a small amount of target, with and without selective irradiation of protein. During selective irradiation, the magnetization spreads through the protein by intra-molecular cross-relaxation and it is transferred to the bound ligands during its residence time in the target [133]. In the case of longer relaxation rates for the hydrogen atoms of the small molecule than the dissociation rate constant ( $k_{\text{off}}$ ) of the complex (typically true for ligands with  $\mu\text{M}$  to  $\text{mM}$  dissociation constants), there will be an accumulation of the saturated ligand (even for the substoichiometric protein concentration). The selective saturation of the protein target can be obtained by irradiating the aliphatic region of the  $^1\text{H}$  NMR (typically between -1 and 2 ppm), usually well populated by methyl groups of the protein, but are not occupied by resonances from small-molecule, organic ligands. Subsequently, a difference spectrum is generated from two spectra that are recorded with and without pre-irradiation (Figure 4.16).



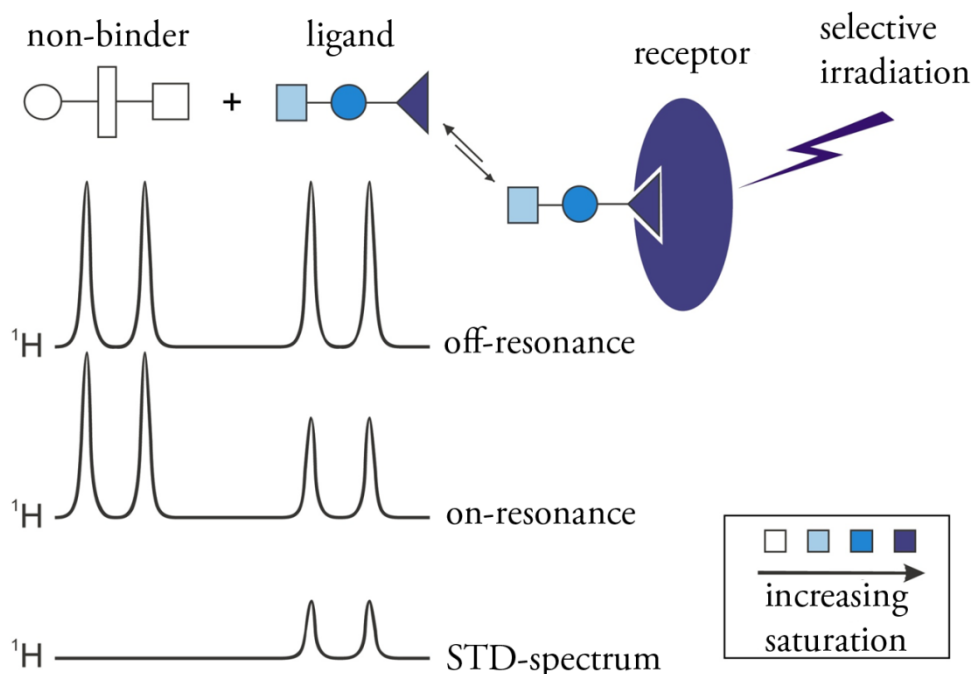
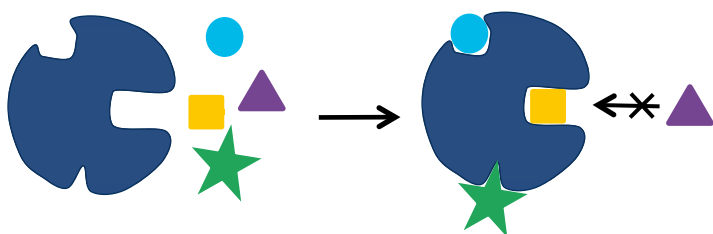


Figure 4.16 Schematic representation of the STD-NMR experiment.

For large ligands it is possible to predict binding epitope. This is because the molecule that interacts directly with the protein, receive the highest degree of saturation which can be identified as a stronger STD signal [24, 134].

The fragment library contained initially 402 molecules, and was further expanded with additional 125 fragments reaching finally 527 compounds. Both parts of the library were screened separately following the same steps. In the first step, the compounds that bound to EV-B93 3C protease were identified on the basis of the visible STD signals. At this point all the fragments that interact with all binding sites of the protein were selected (Figure 4.17, top). The experiment was repeated for all positive hits to confirm their binding to EV-B93 3C protease. Subsequently, the positive hits were taken to perform the competitive STD NMR screening in the presence of irreversible covalent binder rupintrivir. Rupintrivir was chosen due to its excellent affinity to EV-B93 3C protease as well as ability to occupy completely the proteolytic active site. During this step, only molecules that interact with the enzyme (outside of proteolytic active pocket) showed the STD signals, and were therefore excluded from further studies. The compounds that possess affinity to the proteolytic active side could not compete with rupintrivir, therefore did not bind to EV-B93 3C protease (Figure 4.17, bottom). No STD signals were recorded for these molecules.

## STD NMR screening



## STD NMR competitive screening

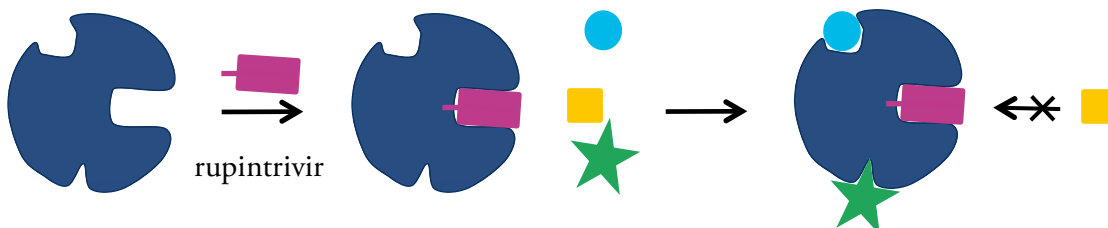


Figure 4.17 Schematic representation of the STD NMR fragment screening and the STD NMR competitive screening. Navy blue shape represents EV-B93 3C protease; blue, yellow, green and violet shapes correspond to the low-molecular compounds coming from the fragment library and pink shape represents covalent inhibitor rupintrivir.

The screening of the first part of the library (402 compounds) resulted in the selection of 38 molecules that interacted with EV-B93 3C protease, among which 24 gave strong STD signals. When repeating the experiment in the presence of rupintrivir, 11 compounds showed good competition (no STD signal was observed), 9 average competition (reduced STD signals were observed) and 4 no competition (STD signals did not change their intensity). In the case of the second part of the library (125 compounds) 13 molecules showed good STD signal with good competition in the presence of rupintrivir. The overall hit rate for both primary and competitive screening of EV-B93 3C was 4.6 %, since only 24 out of 527 molecules appeared to interact relatively strongly with the protease active site.

The aim of the screening process was to identify the molecules that not only bind to EV-B93 3C protease but more specifically interact with its proteolytic active site. Various limitations should be considered when analyzing the screening data. Firstly, the STD NMR technique is able to detect only low or medium binding affinities. Compounds with high affinity may not



be detected showing false negative results. However, this kind of molecules should not be present in the library due to the small size of the fragments. The second problem is related to the false positives that can arise by two means: by direct irradiation of ligand signals or by the presence of aggregates. Irradiation of the ligand may occur when the frequency of the saturation pulse is set too close to the aliphatic signals of the molecule. Thus, one should be very careful not to confuse those spectra. Repeating the spectrum in the absence of the protein may help to identify the presence of the aggregates. If the STD signals are still observed, they originate from ligand aggregates.

The overall hit rate for EV-B93 3C protease was 4.6 % which is in accordance with the hit rates for fragment based screening mentioned in the literature. The reported hit rates vary from 0.01 to 0.94 % for multiple screening campaigns [135] to 10 % for FABP4 [136]. The extreme example was described for norovirus-like particle where the observed hit rate reached 61 %, possibly due to the size of the target [137].

Small, simple molecules have lower affinities than HTS compounds, but when screened at high concentration, they result in higher hit rate, which is a good indicator for a protein druggability [135]. The fragment screening of EV-B93 3C protease gave quite low hit rate which may suggest poor druggability of the protein. Indeed, the crystal structure described in the section 4.1.4 revealed no deeply embedded cavity, but rather shallow binding pocket widely dispersed over the protein surface.

#### 4.4.3 Biophysical validation of the selected hits – TSA

The ligand-induced thermodynamic stabilization method commonly known as thermofluor or thermal shift assay was introduced by 3DP (currently part of Johnson&Johnson) [138, 139]. In this technique, the stabilization manifests as an increase in thermal stability upon binding of a ligand. The experiment is performed with a real time PCR machine (qPCR) and environmentally sensitive fluorescent dyes such as 1,8-ANS, 2,6-TNS or Sypro Orange, which bind to hydrophobic regions exposed on the protein during unfolding process. The fluorescence is quenched in an aqueous solution, but starts to be visible in the hydrophobic environment. Fluorescence increases gradually when more hydrophobic residues of the protein become exposed during heating process and this way melting curve can be recorded. Melting temperature,  $T_m$ , is defined as the temperature at which 50 % of the protein is unfolded. Comparison of this value with melting temperature of the protein in an absence of a ligand,  $T_0$ , allows the determination of the degree of stabilization. Using this method, the validation of 38 hits coming from the screening of the first part of the library was done. High concentration of each compound (1 mM) was used due to expected low affinities of the selected fragments to

EV-B93 3C protease. The melting curve raw data were plotted as fluorescence versus temperature (Figure 4.18 A). In order to visualize the  $T_m$  more clearly, the negative first derivatives were plotted, making the  $T_m$ s appear as peaks (Figure 4.18 B). No significant thermal shifts were observed for any of the tested compounds. The protein alone showed melting temperature equal to 54.5 °C, while in the presence of 38 hits the  $T_m$  values ranged from 54.5 to 55.0 °C, which fell within the measurement error. Due to weak response of EV-B93 3C protease in the thermal shift assay, the most promising hits from the second part of the library were not evaluated by this method.

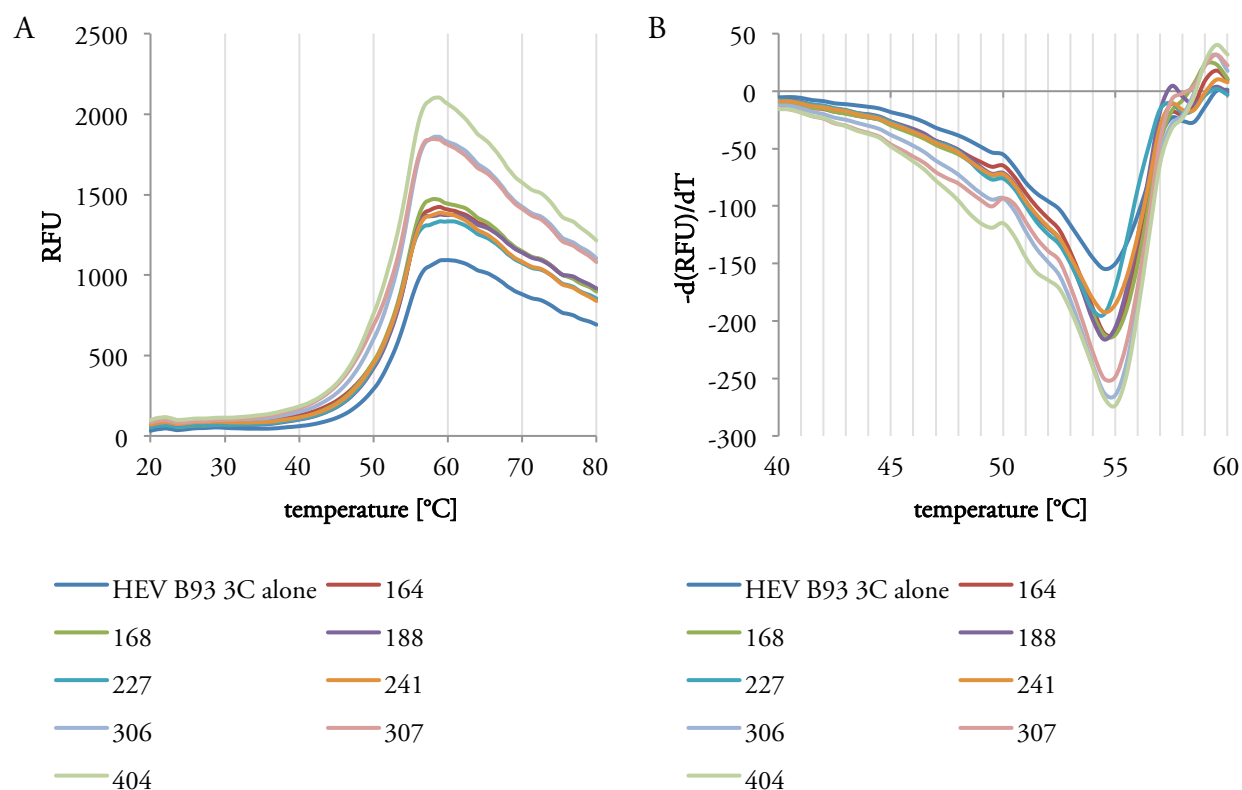


Figure 4.18 Melting curves for EV-B93 3C protease alone and in the presence of 8 random hits from the first part of the screening. The fluorescence response (relative fluorescence units, RFU) versus time (A) and the negative first derivative of the change in fluorescence versus time (B).

Thermal shift assay is not as sensitive as STD NMR, thus may not be able to detect weak affinities. On the other hand, the fragments themselves may not stabilize the protein enough to change its melting temperature, which can result in many false negatives during fragment screening or hit validation. Both aforementioned statements are probably true for EV-B93 3C protease, since none of the hits showed stronger ( $T_m - T_0 > 0.5$  °C) stabilization of the protein.

However, there may be the third cause of an inability for further stabilization of the protein – the high melting temperature value. Indeed, EV-B93 3C protease has relatively high melting temperature equal to 54.5 °C. It was described in the literature that thermostability of proteins depends on many factors among which are: surface contribution of polar atoms (nitrogen and oxygen atoms) able to form hydrogen bonding with water molecules, better hydrophobic internal packing, protein density and volume of the cavities present in the protein [140]. It was already mentioned that EV-B93 3C protease is a very compact enzyme lacking deep cavities. Moreover, it is very soluble in aqueous solutions due to the presence of a polar surface and a deeply hidden hydrophobic core. The above suggestion was proven in the thermal shift assay performed in the presence of rupintrivir, potent inhibitor of EV-B93 3C protease, which served as a positive control. Rupintrivir was also unable to further stabilize the protein.

Thermofluor was not the best method for the hit validation in the case of EV-B93 3C protease, however, its performance varies for different protein targets [141]. For example, thermal shift assay was proven to be the most appropriate fragment screening method for the tumor suppressor p53, which reactivates upon binding of small organic molecules due to thermal stabilization [142].

#### 4.4.4 Biophysical validation of the selected hits – SPR

Surface plasmon resonance (SPR) technology has emerged as a powerful technique to investigate the interaction between low-molecular-weight molecules and target proteins. The first commercial Biacore instrument was developed by Pharmacia Biosensor in 1990 [143]. SPR occurs when plane-polarized light hits a metal film (usually silver or gold) under conditions of total internal reflection. The SPR angle, which is the angle of the incident light that results in the lowest intensity of reflected light at a constant wavelength, depends mainly on the properties of the metal film and the refractive index of the medium that is close to that film. By monitoring alterations in the refractive index, SPR is able to measure changes in the mass of dissolved material in the aqueous layer close to the metal film which allows the interaction of proteins with other molecules or ligands to be monitored in real-time [144]. This allows for determination of the kinetic parameters and equilibrium constants, i.e. on-rate, off-rate, dissociation constants as well as thermodynamic parameters. The signal of SPR depends linearly on the mass of binding ligand to the immobilized target. The resulting sensogram includes association and dissociation phase (Figure 4.19).

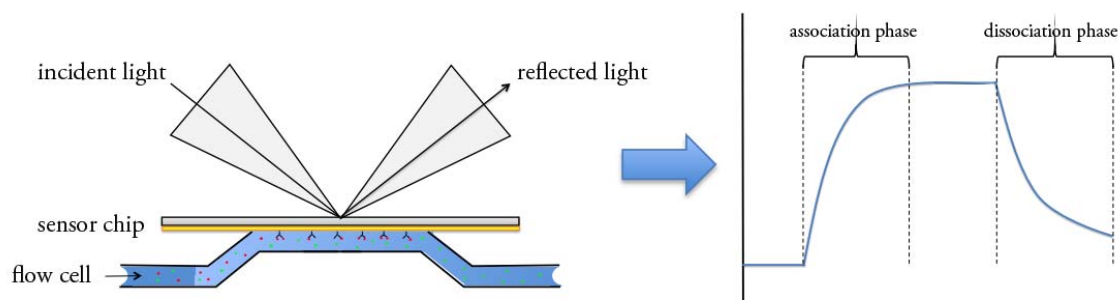


Figure 4.19 Experimental set up of SPR and a sensogram.

51 compounds coming from the STD NMR screening were tested in SPR to order the hits according to their  $K_d$  values. The first trial was done using amine coupling for the immobilization of EV-B93 3C protease via its free primary amine groups (lysine residues). Regrettably, the maximum binding response  $R_{max}$  for the screened molecules was very low around 5 RU while the expected values should oscillate between 15-26 RU (response unit). This may have occurred due to the random orientations of the bounded protein and the lack of homogeneity, which could result in the partial blockage of the protein active site. To overcome this problem indirect capture immobilization was applied. This method involves high-affinity capture of His-tagged protein onto covalently immobilized anti-His antibodies. The benefit of this capture approach is the creation of a homogenous surface, since all protein molecules are orientated through a common site – His-tag. This time the obtained results seemed to be reasonable and repeatable (Table 4.9).

**Table 4.9 Results of validation by SPR**

no.	name	molecular mass [Da]	$K_d$ [ $\mu$ M]
1	92	241	522.4
2	144	192	nd
3	160	178	113.4
4	164	210	270.9
5	168	216	167.7
6	169	256	297.2
7	179	174	135.2
8	188	233	297.8
9	211	222	23.0
10	227	203	nd
11	232	172	150.7
12	239	169	298.3
13	241	217	12.1

14	243	202	89.1
<b>15</b>	<b>259</b>	<b>157</b>	<b>91.1</b>
16	290	190	243.2
17	306	292	168.2
18	307	230	168.2
19	318	174	363.6
20	404	159	279.0
21	406	150	141.3
22	412	164	290.2
23	427	173	275.6
24	432	195	352.8
25	438	193	1096.0
26	441	157	163.4
27	442	186	263.6
28	443	166	272.3
29	445	213	477.4
<b>30</b>	<b>459</b>	<b>178</b>	<b>90.6</b>
31	471	215	263.1
<b>32</b>	<b>476</b>	<b>267</b>	<b>89.0</b>
33	484	162	293.3
34	487	271	219.9
35	494	178	220.4
36	542	176	237.9
37	544	193	177.3
38	550	179	252.1
39	578	193	261.5
40	606	232	253.0
41	629	218	185.0
42	631	197	252.0
43	635	215	366.3
44	644	279	836.5
45	656	248	353.3
46	668	207	nd
<b>47</b>	<b>675</b>	<b>246</b>	<b>51.2</b>
48	686	227	146.9
49	688	233	1544.0
50	717	271	nd
51	723	210	135.0

nd – not determined; compounds and corresponding Kd values < 100  $\mu\text{M}$  are indicated in bold blue; compounds and corresponding Kd value  $\approx$  100  $\mu\text{M}$  are indicated in blue

The affinity determination ( $K_d$ ) of the fragments was assessed on the basis of the binding isotherm obtained by the screening of 5 different concentrations of a ligand (0, 0.125, 0.250, 0.5 and 1 mM). The estimation of the kinetic constants was impossible due to the fast kinetics of the fragments. 6 compounds showed  $K_d$  lower or very close to 100  $\mu\text{M}$  (Table 4.10). In three cases the  $K_d$  was impossible to be measured because of the solubility problem or insufficient amount of compound, which left after the TSA.

**Table 4.10 Chemical structures of the most potent hits in the SPR experiment.**

name	$K_d$ [ $\mu\text{M}$ ]	chemical structure
241	12.1	
211	23.0	
675	51.2	
476	89.0	
459	90.6	
160	113.4	

The analysis of the selected fragment hits showed a number of structural similarities. All the compounds have at least one aromatic or heteroaromatic ring. 4 out of 6 molecules possess two aromatic (one five-member and one six-member) rings separated by the different-length spacers (only covalent bond for compound 241 and 160, the one-atom spacer for compound 211 and the two-atom spacer for compound 476). 3 fragments have the 4-methylphenyl moiety (241, 211 and 459) while one possesses the 4-methoxyphenyl (476) residue. 5 molecules have the heterocyclic ring with one (675, 160), two (211, 476) or three nitrogen atoms (241).

#### 4.4.5 *In vitro* proteolytic assay

6 most promising compounds with the highest affinity were characterized in the *in vitro* proteolytic assay with EV-B93 3C protease. Only one compound 241 with the best  $K_d$  value showed inhibitory effect (RA-relative activity = 76 %) at high concentration 350  $\mu\text{M}$  (protein compound ratio: 1:200) (Figure 4.20). Thus, the compound 241 was chosen for further elaboration.

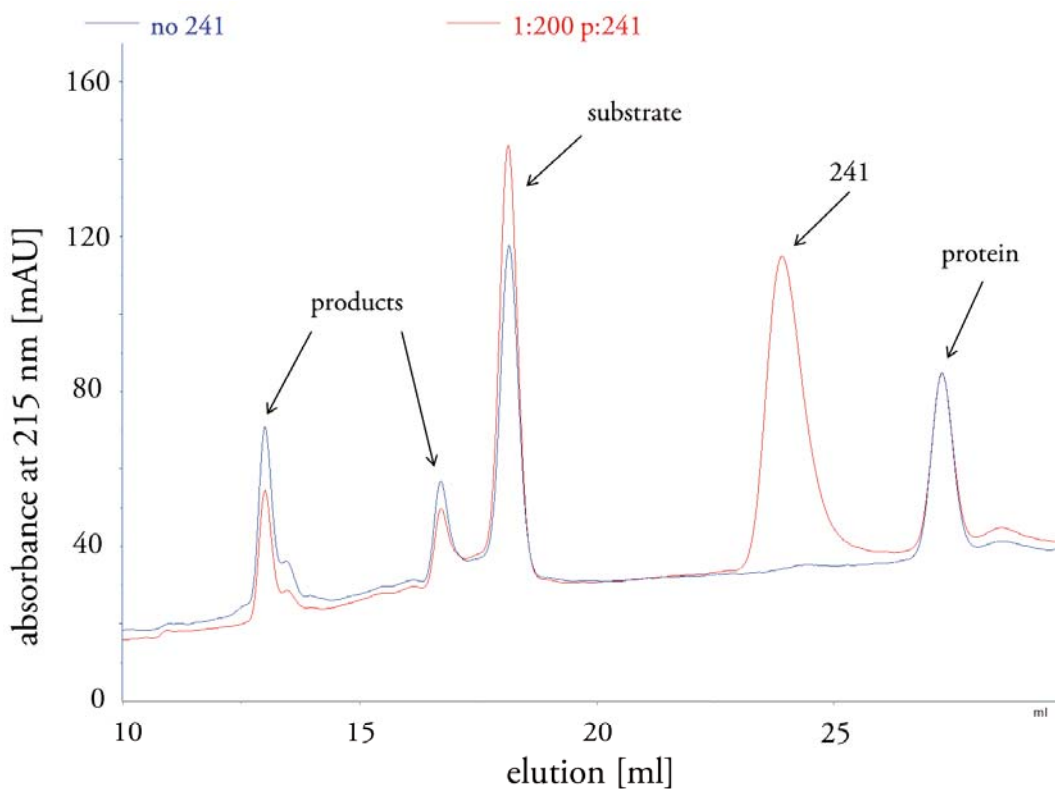
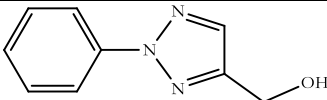
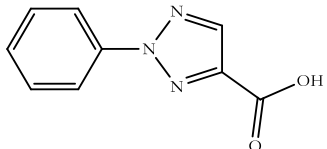
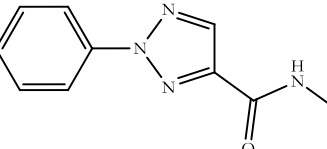
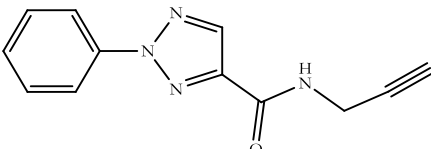
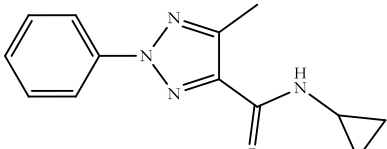
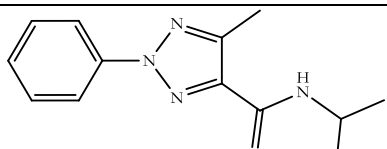
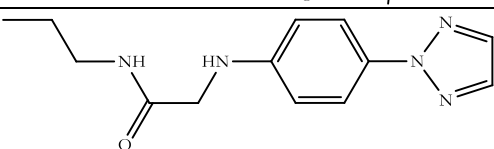


Figure 4.20 *In vitro* proteolytic activity assay of EV-B93 3C protease in the absence and presence of the small molecule 241 at high concentration 350  $\mu\text{M}$ .

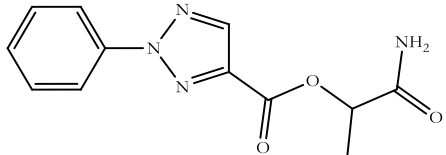
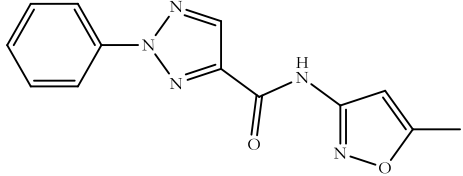
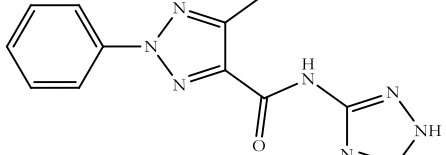
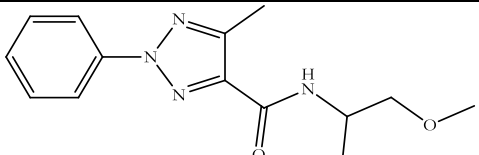
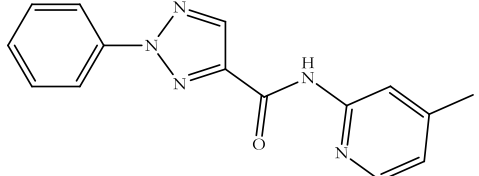
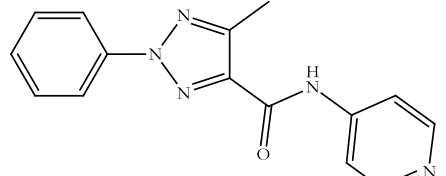
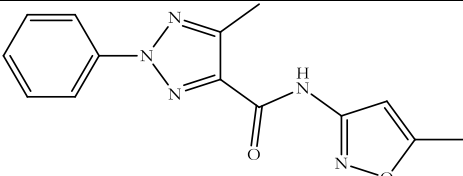
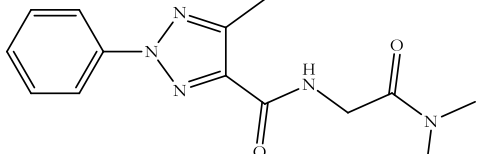
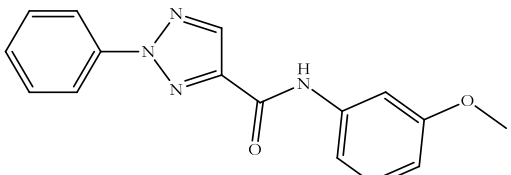
## 4.4.6 Elaboration of compound 241

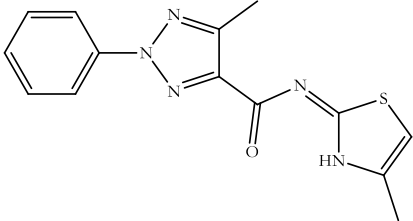
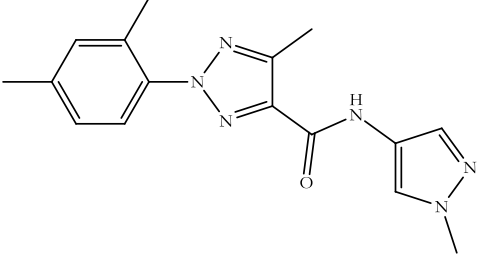
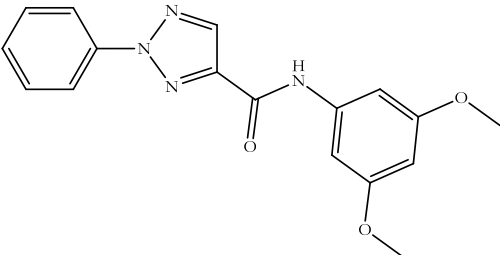
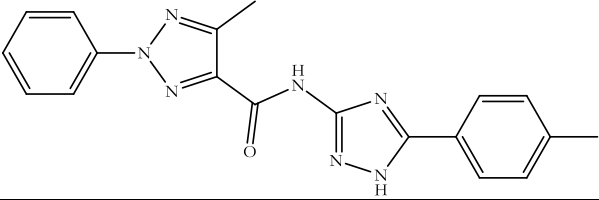
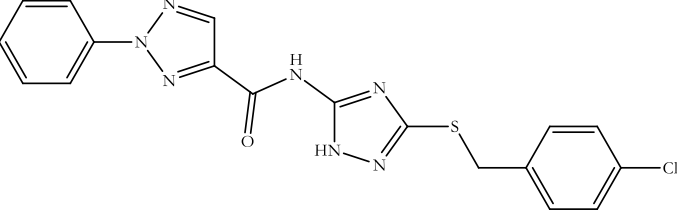
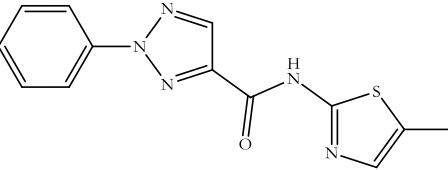
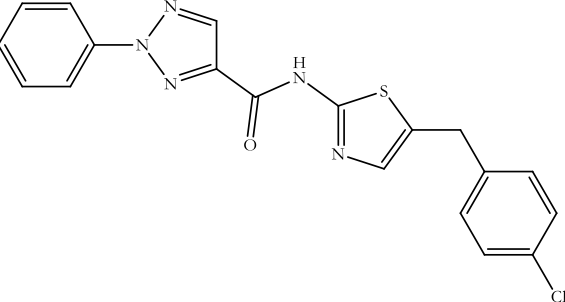
44 analogs of the best hit 241 were selected based on the sub-structural similarity searching (done by Daniel Álvarez from prof. Xavier Barril's laboratory, the University of Barcelona), and purchased from different vendors. All the compounds were tested in the *in vitro* proteolytic assay at high concentration 350  $\mu$ M (protein compound ratio: 1:200). The chemical structures of compounds and their inhibitory activities are listed in Table 4.11.

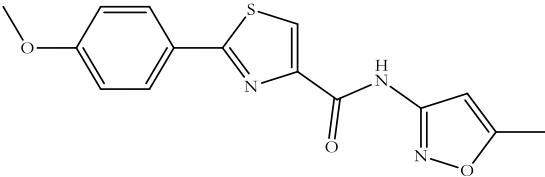
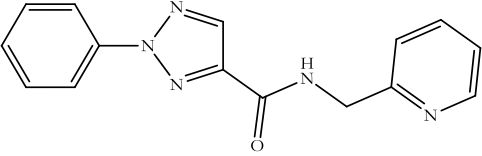
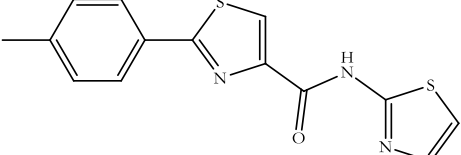
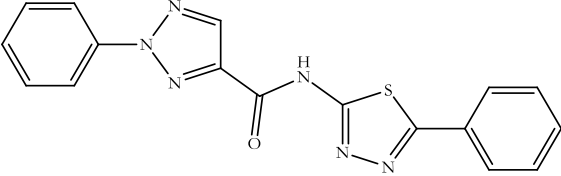
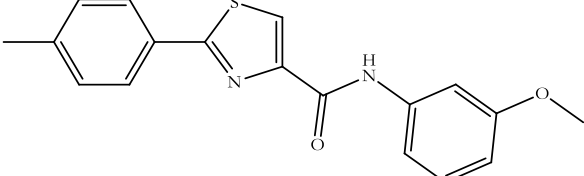
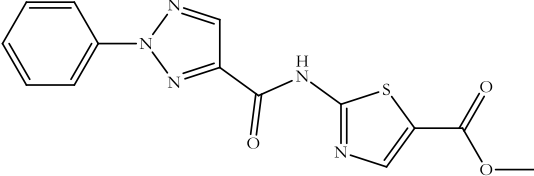
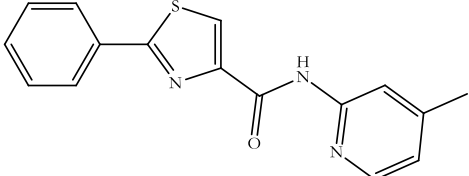
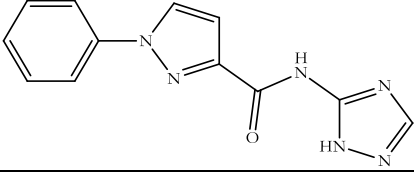
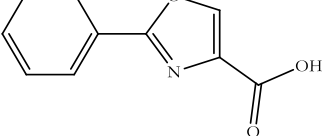
**Table 4.11** Analogs of compound 241 and their inhibitory activity against EV-B93 3C protease.

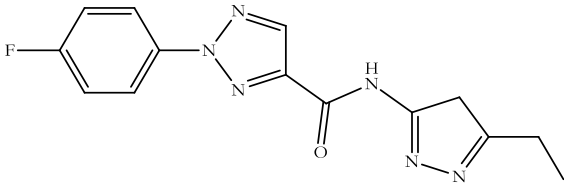
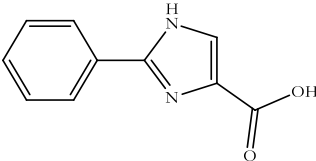
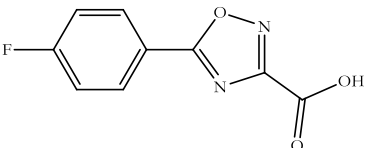
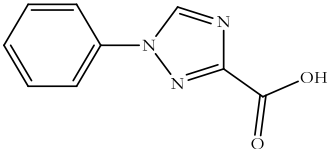
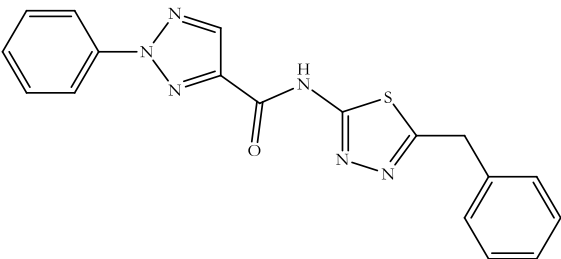
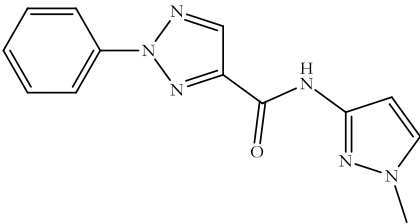
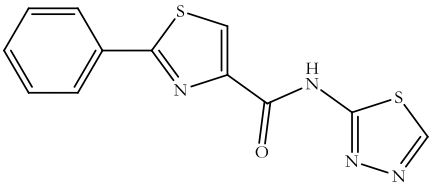
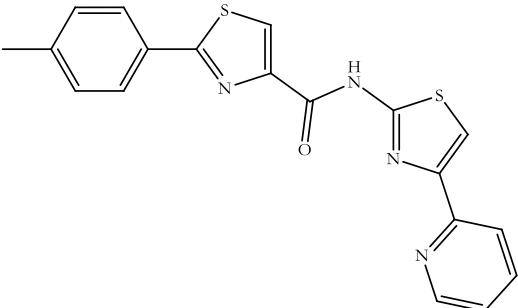
no.	name	molecular mass [Da]	chemical structure	RA <sup>i</sup> [%]
1	EN1	175		5.5 <sup>ii</sup>
2	EN2	189		0.0
3	EN3	202		100
4	EN4	226		100
5	EN5	242		100
6	EN6	244		100
7	EN7	259		100

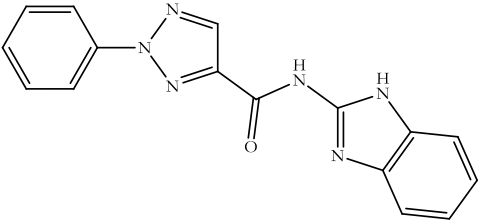
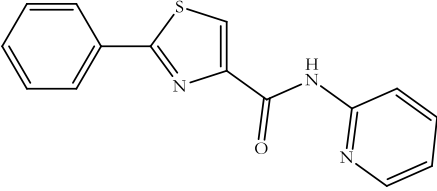
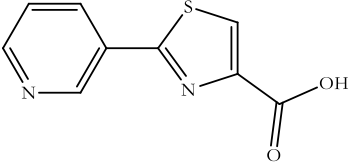
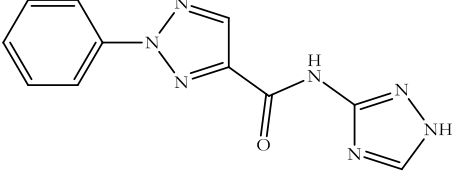


8	EN8	260		98.6
9	EN9	269		3.0
10	EN10	269		64.7 <sup>ii</sup> <sub>i</sub>
11	EN11	274		98.3
12	EN12	279		0.0
13	EN13	279		99.7
14	EN14	283		93.6
15	EN15	287		100
16	EN16	249		51.4

17	EN17	299		80.3
18	EN18	310		84.6
19	EN19	324		82
20	EN20	359		53.3
21	D1	411		76.7
22	D2	285		54.0
23	D3	396		65.8

24	D4	315		94.1
25	D5	279		0.0
26	D6	301		82.7
27	D7	348		68.3
28	D8	324		96.5
29	D9	329		38.6
30	D10	295		87.2
31	D12	254		72.5
32	D13	189		92.7

33	D14	300		74.6
34	D15	188		67.5
35	D16	208		76.1
36	D17	189		0.0
37	A1	362		50.0
38	A2	268		98.4
39	A3	288		96.6
40	A4	379		0.0

<b>41</b>	<b>A5</b>	<b>304</b>		<b>65.7</b>
<b>42</b>	A6	281		73.5
<b>43</b>	A7	206		82.1
<b>44</b>	<b>FCH</b>	<b>255</b>		<b>69.0</b>

<sup>i</sup>RA – relative activity of EV-B93 3C protease; the molecules for which RA < 6 % are highlighted in red; the molecules for which 30 % < RA < 70% are highlighted in blue.

7 compounds out of 44 were able to inhibit EV-B93 3C protease effectively (RA < 6 %) (Figure 4.21 A). For those compounds, the IC<sub>50</sub> values were estimated based on the corresponding dose-response curves (Figure 4.21 B). The values ranged from 115 μM for the compound D17 to 5 μM for the compound A4. The best 4 molecules: A4, EN2, EN12 and D17 were taken for further evaluation.

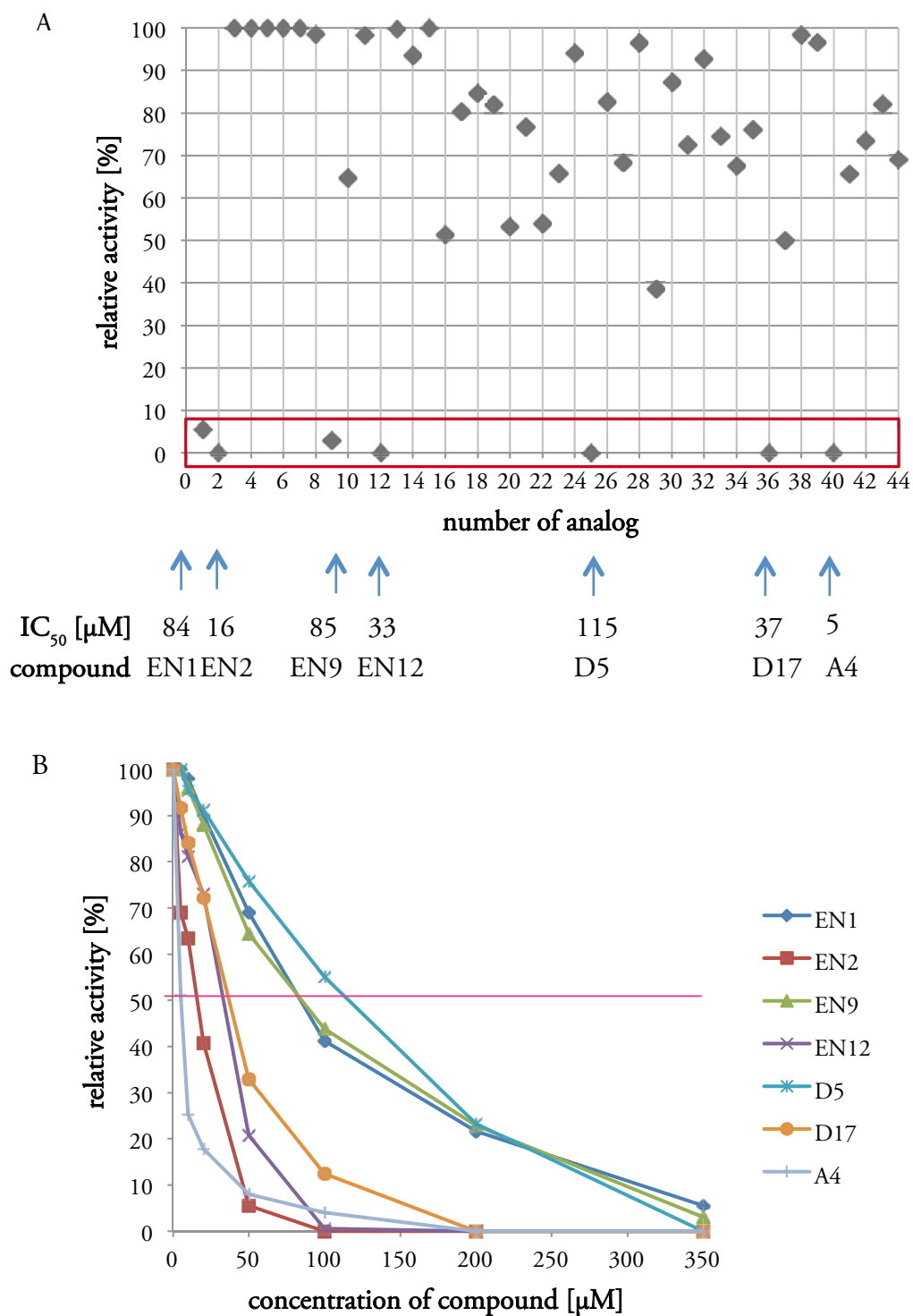


Figure 4.21 Relative activity of EV-B93 3C protease in the presence of 44 analogs of compound 241 at high concentration (350 μM) (A). The IC<sub>50</sub>s for 7 best compounds were estimated based on the corresponding dose-response curves (B).

Next, structure-activity relationship (SAR) of 44 analogs was analyzed in order to improve the potency of the most promising compounds. Several observations supported by the additional experiments have been made and are as follows:

1. The triazole ring B is necessary for the activity.

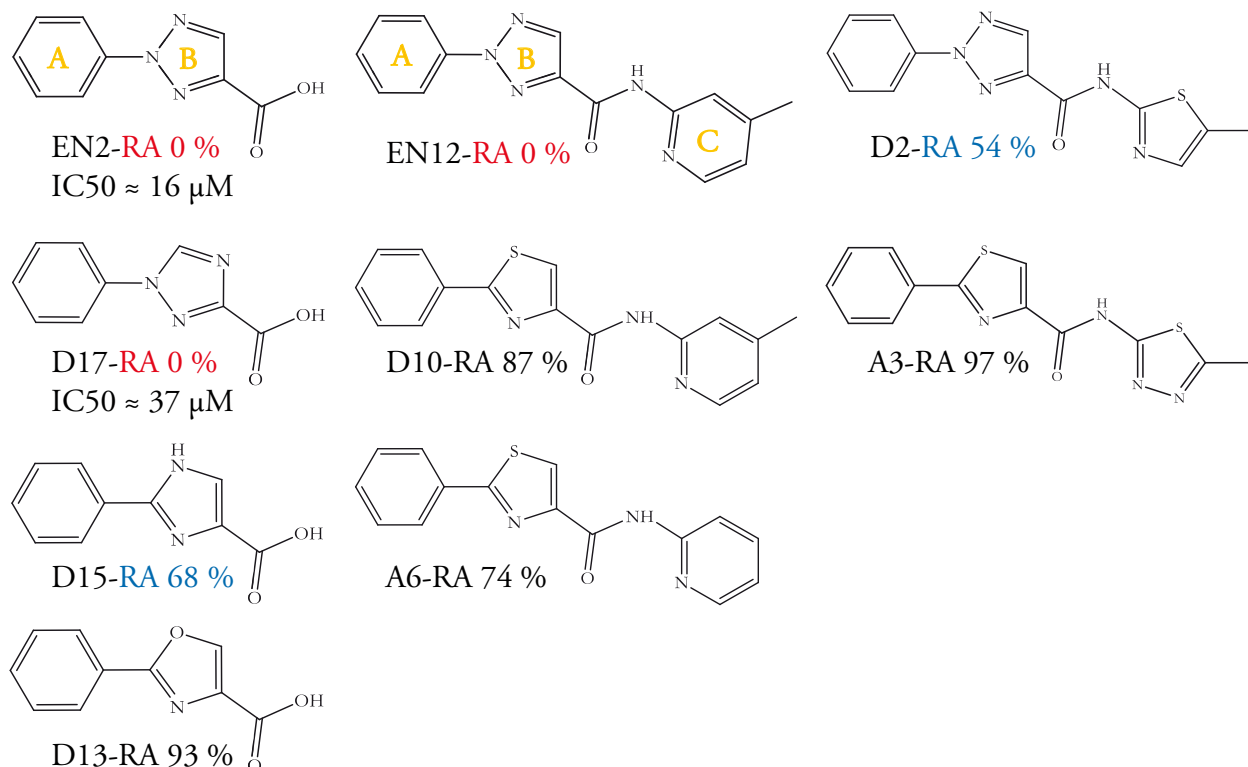


Figure 4.22 SAR comparison of the selected compounds based on the presence of the triazole ring B. Relative activity (RA) is shown for each compound. IC<sub>50</sub> values are shown for the compounds EN2 and D17.

In the case of compounds with 2 (A and B) or 3 (A, B and C) rings, the triazole ring B is essential for their potency (Figure 4.22, compounds EN2 and EN12). Any changes in the order of the nitrogen atoms (compound D17) or reduction of their number resulted in the activity loss (compounds D15 and D13). For the molecules with 4 rings this tendency is not so evident (compound A4 in Figure 4.26).

2. For the compounds bearing two aromatic rings (A and B) and in some cases even three aromatic rings (A, B and C) the presence of the methyl group in the triazole ring B causes a considerable activity loss (Figure 4.23).

3. The presence of the methyl group in the *para* position of the phenyl ring A affects the inhibitory activity in the case of the molecules with two aromatic rings A and B (Figure 4.23).

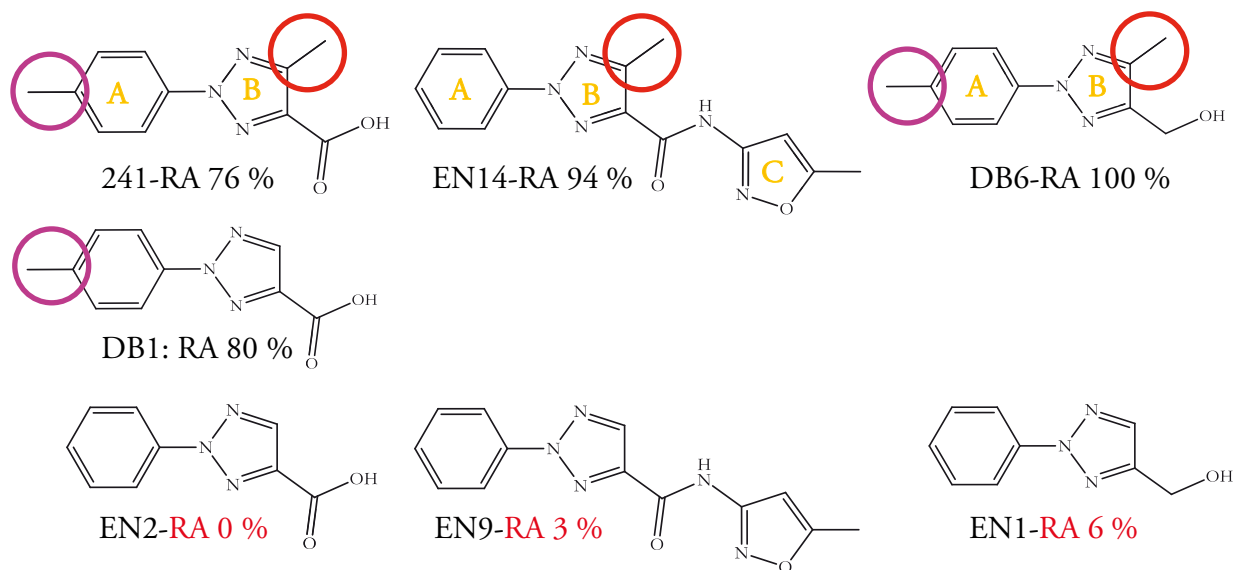


Figure 4.23 SAR comparison of the selected compounds based on the presence of the methyl groups in the ring A and B. Relative activity (RA) is shown for each compound.

The statement 2 is generally true for the 2- or 3-ring molecules, while the statement 3 was supported by the additional experiments performed with the compounds with 2 aromatic rings DB1 and DB6 (synthesized by Daniel Becker from the Freie Universität in Berlin). The compound 241 selected in the STD NMR screening is a weak inhibitor of EV-B93 3C protease (RA  $\approx$  76 %). The elimination of only one methyl group (from the triazole ring B) did not change its potency (Figure 4.23, compound DB1 with RA  $\approx$  80 %). However, the removal of the second methyl group (from the phenyl ring A) resulted in the substantial increase in the potency (compound EN2 with RA  $\approx$  0 %). The same observation was made for the compounds DB6 and EN1. Deleting both methyl groups from the ring A and B of the compound DB6 (RA  $\approx$  100 %) gave the molecule EN 1 with a good inhibitory activity (RA  $\approx$  6 %). The presence of only one methyl group in the 3-ring compound EN14 was not favorable for its activity (RA  $\approx$  94 %), since its elimination resulted in the potent inhibitor EN9 (RA  $\approx$  3 %). Not enough data were gathered to support the same hypothesis for the 4-ring analogs.



4. The aliphatic substitutions in the triazole ring B lead to the activity loss.

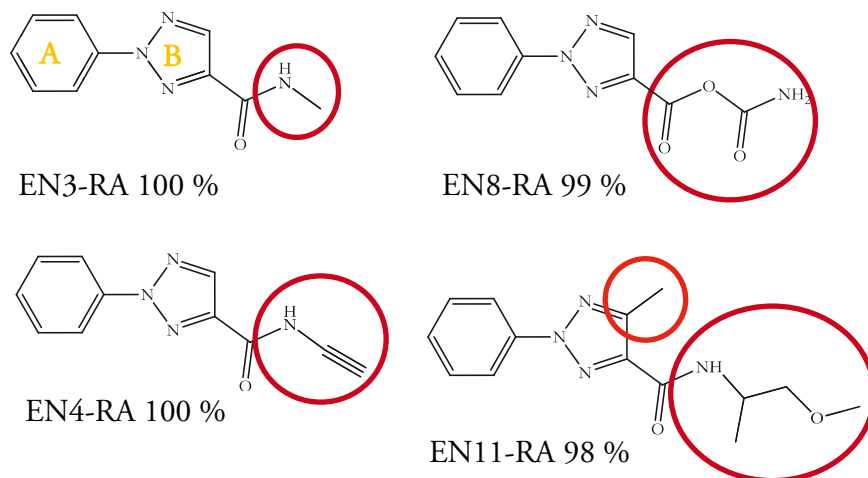


Figure 4.24 SAR comparison of the selected compounds based on the presence of the aliphatic substitutions in the ring B. Relative activity (RA) is shown for each compound.

When the amide moiety in the position of the ring C contains an aliphatic group, the compound loses completely its activity (Figure 4.24).

5. The additional third aromatic ring C with appropriate heteroatoms and groups enhances the potency of the compound.

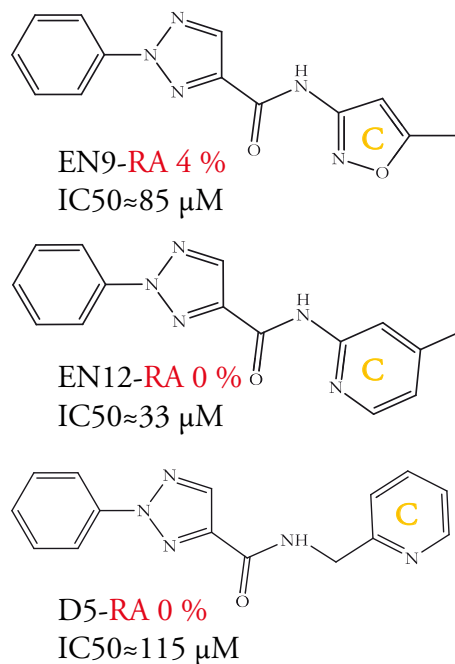


Figure 4.25 SAR comparison of the selected compounds based on the presence of the ring C with appropriate heteroatoms. Relative activity (RA) and IC<sub>50</sub> values are shown for each compound.

The most active molecules with 3 aromatic rings contains one nitrogen atom in *para* position of the ring C and preferably no larger than methyl substituents (Figure 4.25, compounds EN9 and EN12). The amide bond should be linked with the ring C by a covalent bond. Any additional atoms in a linker lead to the decrease of the activity (compound D5).

6. The number and position of the nitrogen atoms in the ring C affect the inhibitory activity. The *ortho* position for the nitrogen atom is favorable.

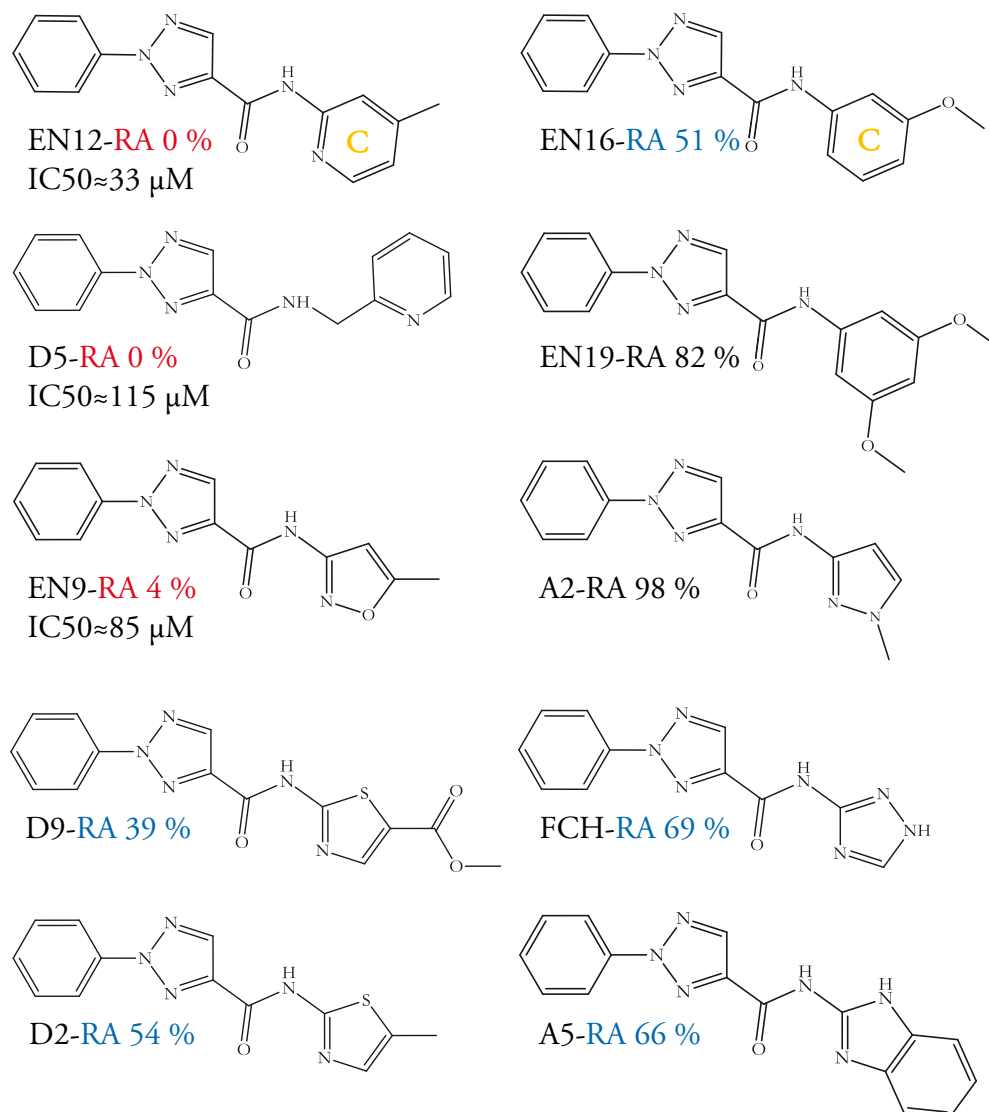


Figure 4.26 SAR comparison of the selected compounds based on the presence and number of the nitrogen atoms in the ring C. Relative activity (RA) is shown for each compound. IC<sub>50</sub> values are shown for the compounds EN9, EN12 and D5.

One nitrogen atom in the *ortho* position of the ring C seems to be optimal for the good inhibitory activity of a compound (Figure 4.26, compounds EN12, D5 and EN9). Any additional heteroatoms cause the decrease in the activity.

7. The additional fourth aromatic ring D can have some positive effect on the inhibitory activity.

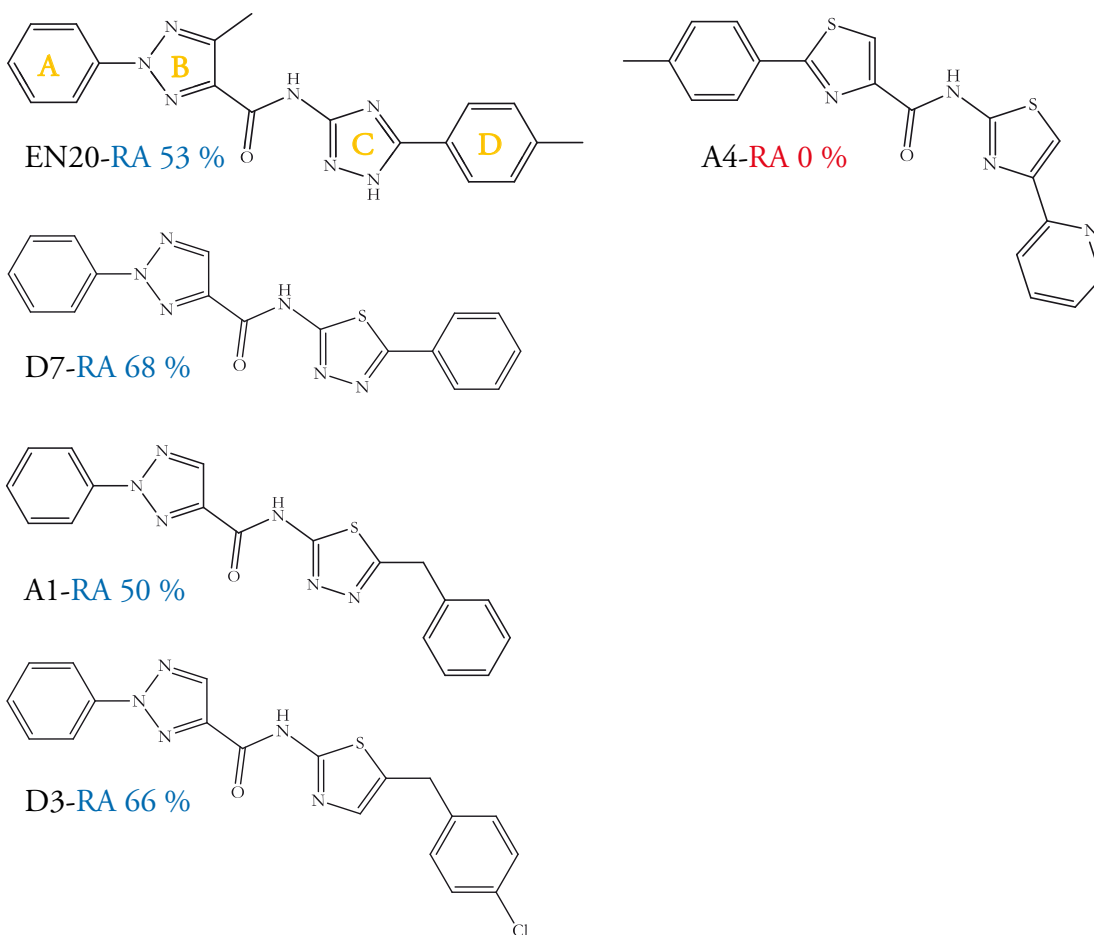


Figure 4.27 SAR comparison of the selected compounds based on the presence of the aromatic ring D. Relative activity (RA) is shown for each compound.

The introduction of the ring D may lead to a potent inhibitor (Figure 4.27, compound A4). The compound bearing the ring D is less sensitive to other unfavorable substitutions, i.e. methyl in the triazole ring B (compound EN20) or in the phenyl ring A (compound A4), additional heteroatoms in the ring C (compounds EN20, D7, A1, D3 and A4) or even lack of the triazole ring B (compound A4).

#### 4.4.7 Chemical optimization of the best compound A4

Among all purchased analogs, the compound A4 showed the best inhibitory activity against EV B93-3C protease with the  $IC_{50} \approx 5 \mu M$  (protein inhibitor ratio 1:2.6). The molecule obeys a couple of rules listed in section 4.1.6 and can be used as a starting point for further modification in order to improve its potency. The compound A4 possesses 4 aromatic rings and does not have a methyl group in the ring B. Moreover, its ring C has a nitrogen atom in the *ortho* position. However, there are some unfavorable features that may be modified to improve the potency of the A4 analogs. The compound lacks the presence of the triazole group, which is substituted by the thiazole ring. Additionally, the phenyl ring A has a methyl group in the *para* position (Figure 4.28).

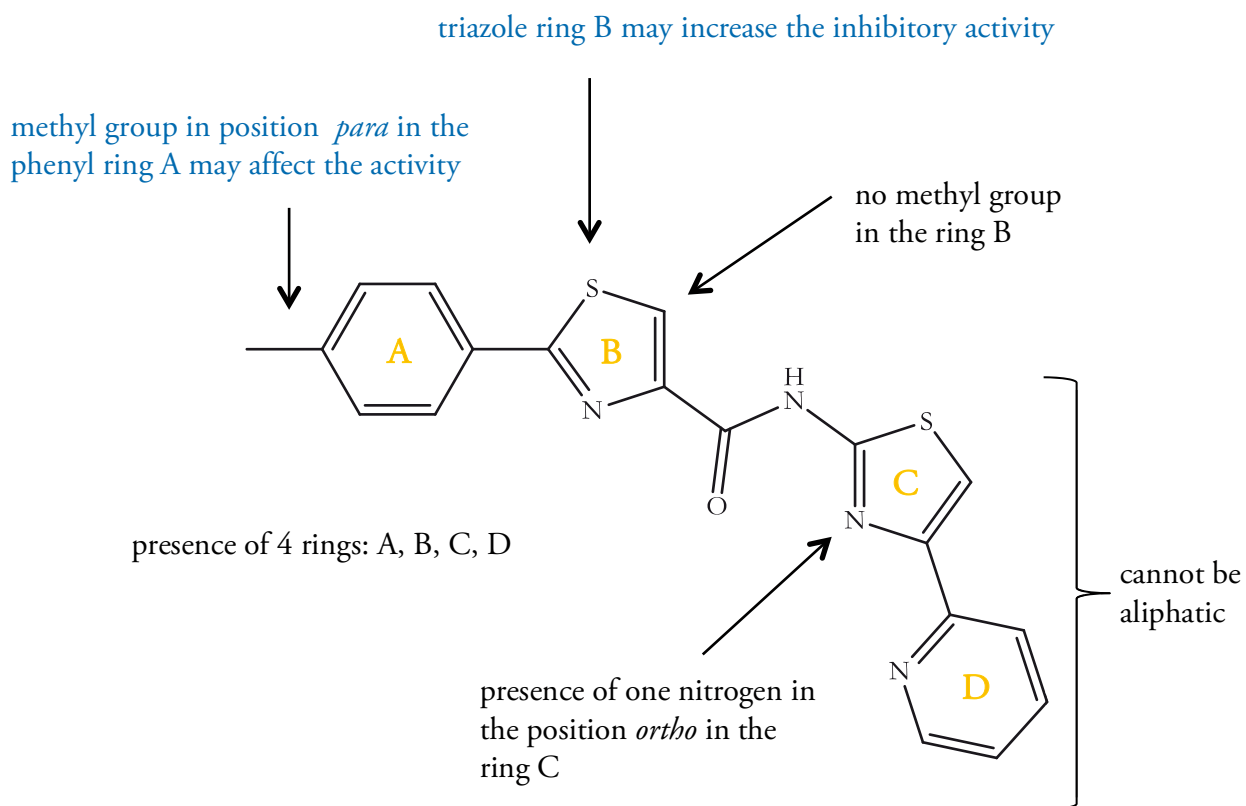


Figure 4.28 Compound A4 as an inhibitor. Chemical features that may affect the inhibitory properties are marked in blue.

Two analogs of the compound A4 were proposed (A4\_B and A4\_C) in order to improve its inhibitory activity against EV-B93 3C protease. Both compounds have the triazole ring B. The

compound A4\_B kept the methyl group in the *para* position of the phenyl ring A, while this substitution was removed in the case of the compound A4\_C (Figure 4.29 A). The chemical synthesis was done by Daniel Becker from the Freie Universität in Berlin.

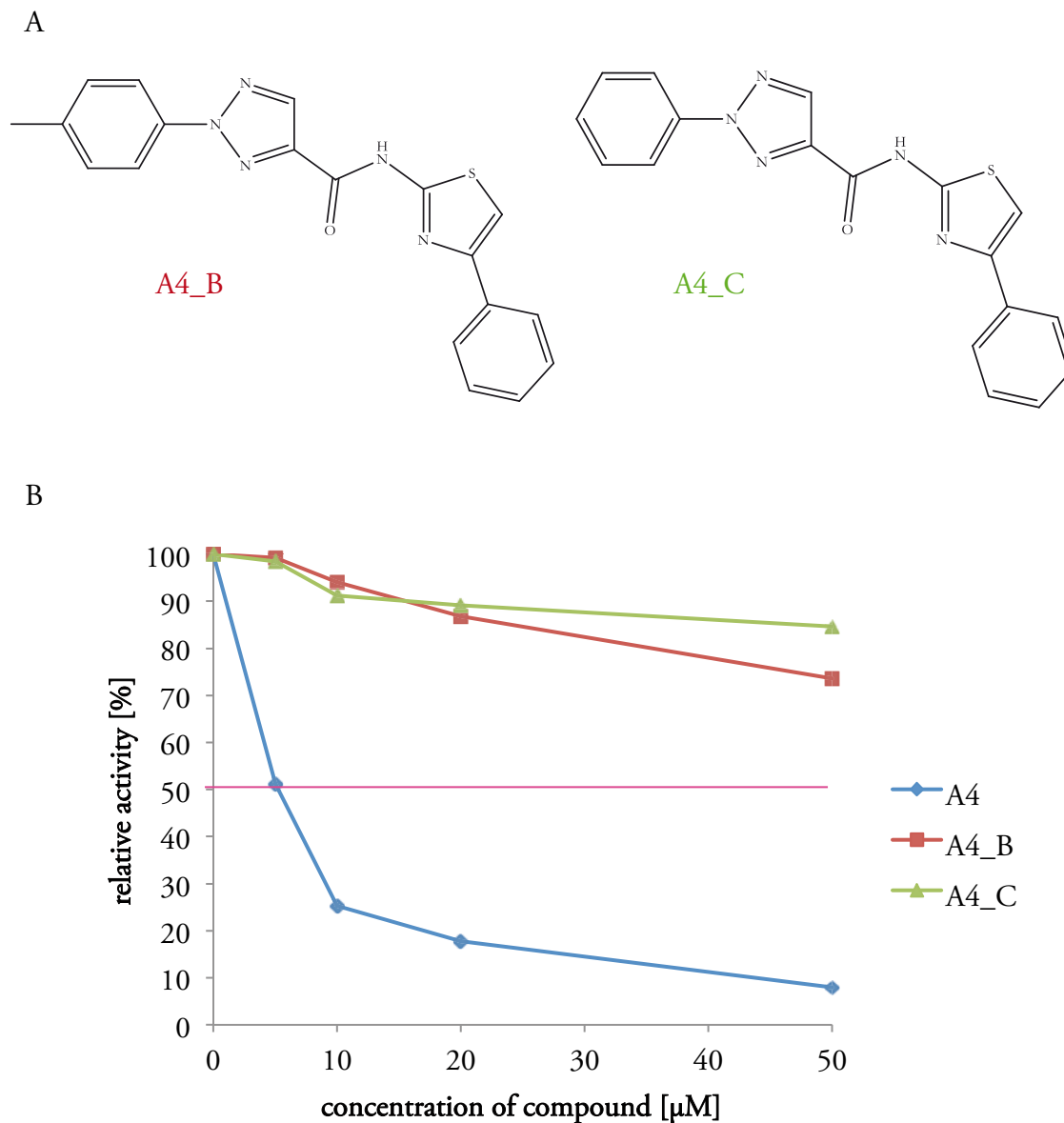


Figure 4.29 Chemical structures of two analogs of A4: A4\_B and A4\_C (A). Comparison of the activity of EV-B93 3C protease in the presence of different concentrations of A4, A4\_B and A4\_C (B).

The inhibitory potency of A4\_B and A4\_C towards EV-B93 3C protease was verified in the *in vitro* proteolytic activity assay. Surprisingly, both compounds demonstrated similar but

considerably weaker inhibition against the protein than the molecule A4 (Figure 4.29 B). This result suggested that another feature present in the compound A4 is much more important for the activity than the triazole ring. The careful structure analysis of A4 and its analogs revealed that nitrogen atom in the *para* position from the ring D must be responsible for the enormous inhibitory activity of the A4 molecule. The appropriate analog (A4\_D) was commercially available (Figure 4.30 A). Activity evaluation proved that the A4\_D analog is similarly active towards EV-B93 3C protease ( $IC_{50} \approx 5 \mu M$ ) when compared to A4, but is more efficient during inhibition (at 20  $\mu M$  compound A4\_D is able to inactivate the protease almost completely while 20 % of the activity is still observed in the case of the molecule A4) (Figure 4.30 B).

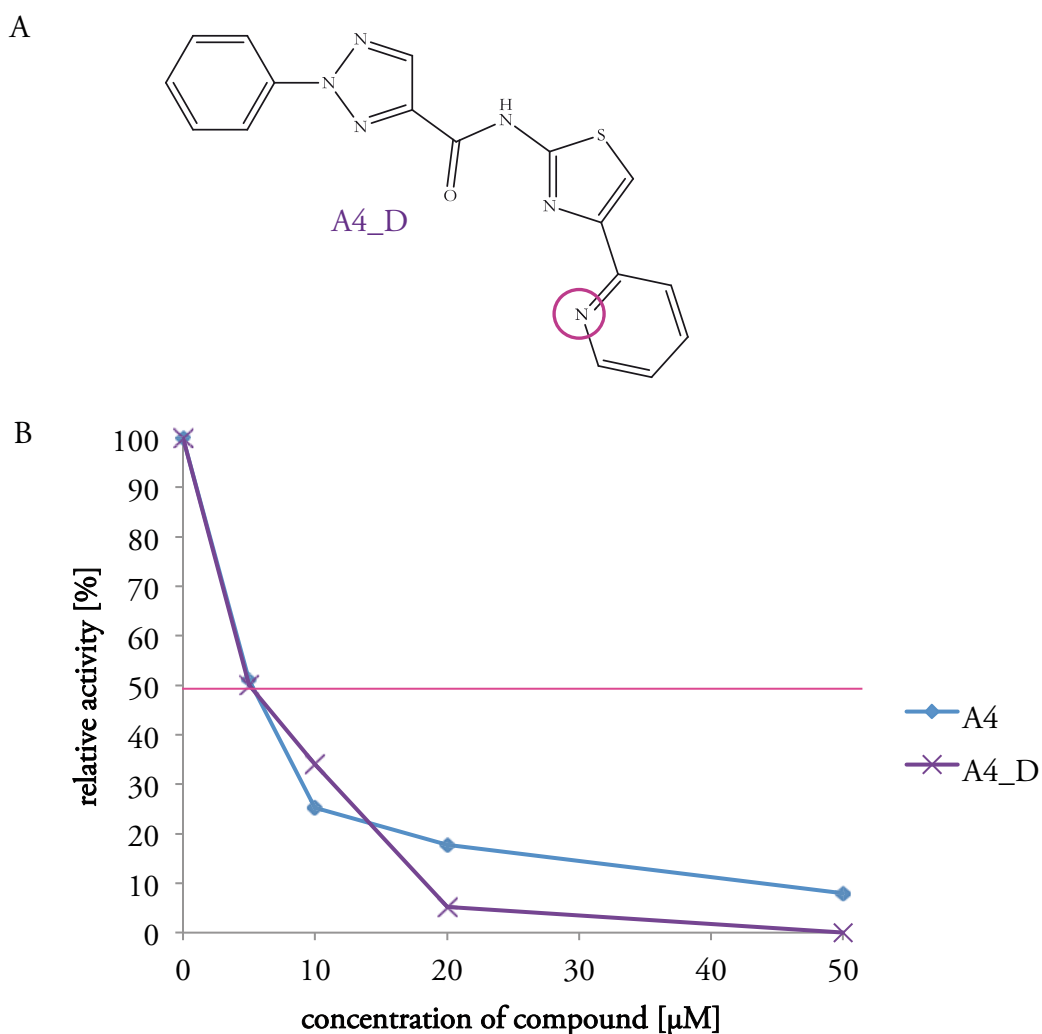


Figure 4.30 Chemical structure of the third analog of A4: A4\_D (A). Comparison of the activity of EV-B93 3C protease in the presence of different concentrations of A4 and A4\_D (B). Important nitrogen atom is encircled in pink.

Looking at the molecule A4\_D it is easy to notice that it is structurally symmetric. After the cleavage of the peptide bond, it is possible to get two similar compounds with 5 and 6-member rings each (Figure 4.31, compounds A and B). Compound A is equivalent to compound EN2, while compound B, called PA4, was purchased from a commercial vendor. Both compounds were tested in the proteolytic activity assay with EV-B93 3C protease.

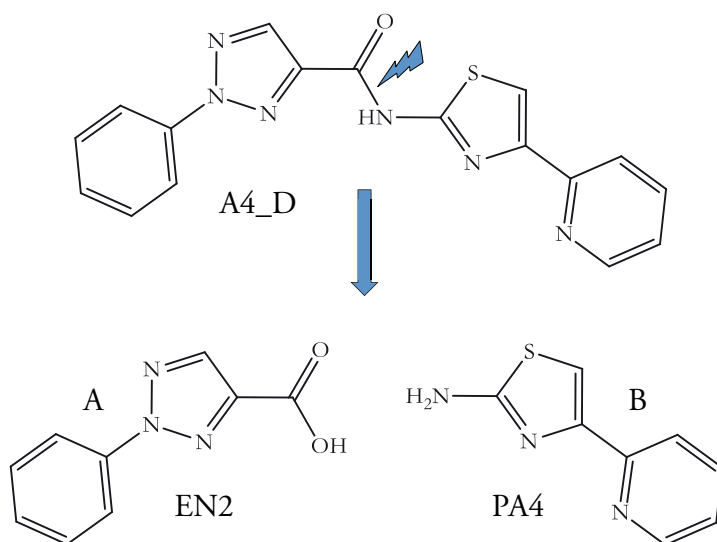


Figure 4.31 Peptide bond cleavage in the compound A4\_D gives two compounds A (EN2) and B (PA4).

The molecule A showed good inhibitory activity against EV-B93 3C protease. The presence of the nitrogen atom in the molecules A4 or A4\_D is crucial for their potency, which may suggest that the second molecule B alone can be active as well. To prove this hypothesis the compound PA4 was purchased and its inhibition towards EV-B93 3C protease was evaluated in the *in vitro* proteolytic assay (Figure 4.32).



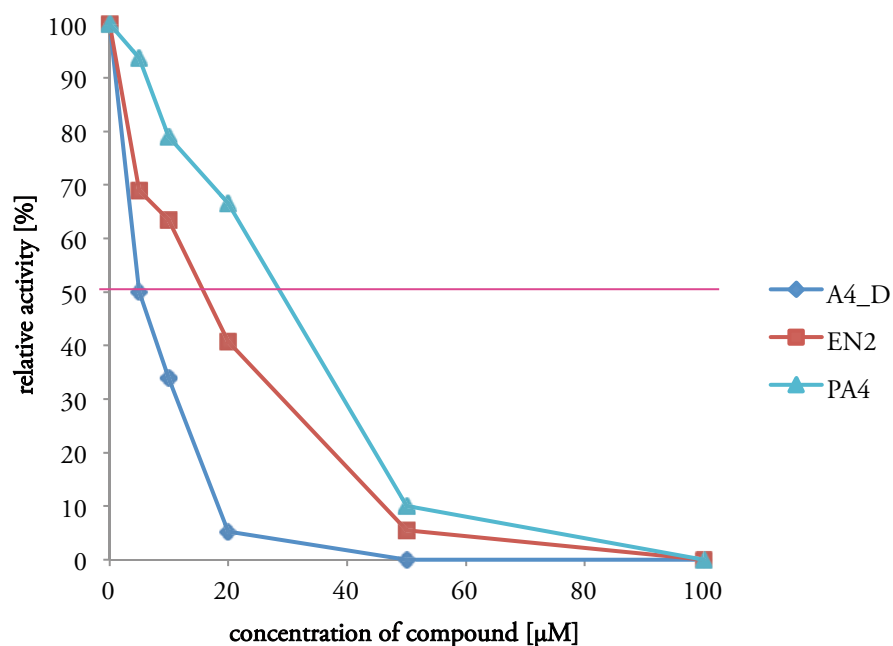


Figure 4.32 Comparison of the activity of EV-B93 3C protease in the presence of different concentrations of A4, EN2 and PA4.

The compound PA4 exhibited considerable good inhibitory activity ( $IC_{50} \approx 29 \mu\text{M}$ ) against EV-B93 3C protease, however its potency was not as good as of the compound EN2 or A4. This experiment shows that the nitrogen atom in the pyridine ring D is extremely important for the activity of the molecule A4/A4\_D. The compound PA4 was very active even without the triazole ring which is crucial for the activity of the molecules containing only two aromatic rings.

Ligand efficiency ( $LE = \frac{pIC_{50}}{N_{non-hydrogen\ atoms}}$ ) can provide a useful comparison between compounds with different MWs and the  $IC_{50}$  values, and can be applied for the monitoring of the ligand optimization process. In general, the LEs of starting fragments are greater than those of larger, more elaborated structures. However, the ligand efficiency values should not decrease much and at least stay constant when altering a lead molecule. In the case of the compounds EN2 and PA4, the LE values are relatively high (0.34 and 0.38, respectively). The combination of these two compounds resulted in the molecule A4\_D with the LE equal to 0.21. This suggests that the optimization step would be desirable for the improvement of its potency by reaching the LE value closer to 0.3. This can be a challenging goal to achieve, since no information of the binding mode is known. Using only chemical structures of the compound A4 or A4\_D it is impossible to predict their binding sides in the active pocket of EV-B93 3C protease without further structural or bioinformatics studies (see sections 4.4.8 and 4.4.11).

#### 4.4.8 Co-crystallization/soaking trials

The co-crystallization as well as soaking trials were performed at the each step of the study. In the case of the hits selected in the screening, the co-crystallization trials were done in the presence of the 20-fold excess of a compound. For more potent analogs (i.e. A4, EN2, EN12 and A4\_D) the crystallization screening were performed in the 5, 10 or 20-fold excess of a ligand. Additionally, for the molecules A4 and A4\_D, the crystallization condition was performed in the equimolar presence of the peptide substrate, and was followed by the seeding with the EV-B93-compound 1 complex crystals. The procedure was aimed at aiding the protein to crystallize in the crystal form matching the one obtained for the covalent complexes (Figure 4.33 C). The soaking of the EV-B93 3C protease crystals was done with all promising ligands but was not always possible to fish the crystal due to its dissolution or heavy precipitation of a soaking compound. The diffraction of more than 200 crystals led to more than 100 datasets that unfortunately did not show any additional electron density in the active site that could be unambiguously assigned to any of the ligands.

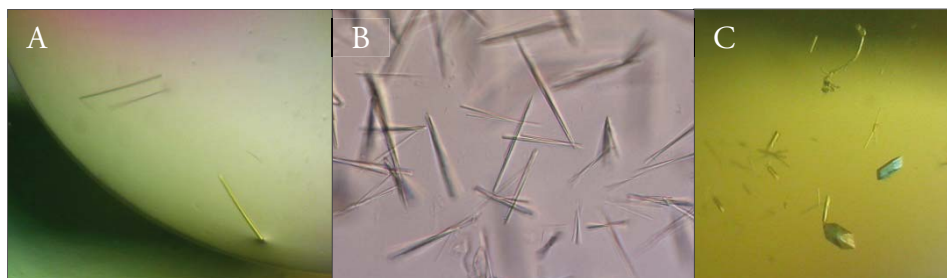


Figure 4.33 Different crystals obtained during crystallization trials with non-covalent ligands. EV-B93 3C protease crystals: native (A), methylated (B) and after seeding with EV-B93-compound 1 complex crystals (C).

In order to change the physicochemical properties of the protein and enhance the formation of the protein-ligand complex, EV-B93 3C protease was methylated. This method was successfully applied to obtain only one known crystal structure of the non-covalent protein-ligand complex with HRV 3C protease [114]. However, the result was not reproducible even by the authors of the paper. EV-B93 3C protease was subjected to a reductive methylation. According to the MALDI-TOF spectra the protein was modified by 24 methyl groups (the mass changed from 21379 Da to 21715 Da), which may attach not only to the lysine side chains, but also to the N-terminal end and the cysteine active site. The methylation of the

cysteine was supposed to be removed by the incubation with DTT, but it was not successful in the case of EV-B93 3C protease, since the modified protein showed no activity in the *in vitro* proteolytic assay. The crystallization screening of the methylated protein resulted in several, but much less crystallization conditions containing the small needle-like crystals when compared to native EV-B93 3C protease (Figure 4.34 B). The optimization process was much more difficult and did not yield any high quality crystals appropriate for soaking trials. Only one crystal was suitable for X-ray diffraction, but it did not reveal any additional electron density in the proteolytic active site.

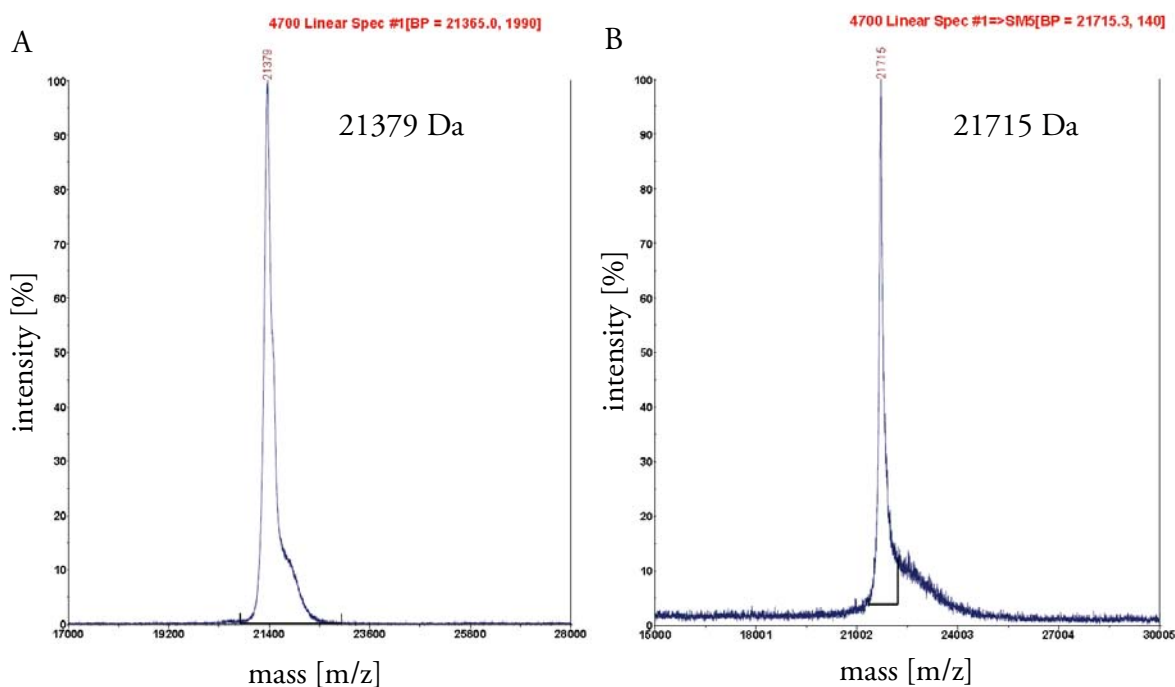


Figure 4.34 MALDI-TOF spectra of native EV-B93 3C protease (A) and after reductive methylation (B).

The inability of obtaining the stable protein-non-covalent ligand complex with EV-B93 3C protease was possibly due to the nature of the specificity pockets for the substrate (or inhibitor) recognition. The active sites of the serine and cysteine proteases are usually shallow, and the binding determinants are widely dispersed over large surface areas. In this situation the bounded molecule interacts only partially with the protein, and possesses a few effective residues exposed to the solvent. The situation is opposite in the case of deep and narrow active sites where substrates/inhibitors are able to interact more efficiently with the protein, thus being easier to obtain protein-ligand complex structures. In this case, even for low-molecular

weight compounds it is possible to get high-affinity complexes with a target protein. The situation is much less favourable for molecules interacting with widely solvent-exposed active sites of serine and cysteine proteases, where a ligand occupancy value is reduced, which results in low or none electron density.

Another possible explanation of the inability to determine the non-covalent EV-B93 3C protease-ligand complex structure is that interactions between the protein molecules in the crystal could be stronger than between the protein and a small compound. Thus, conformational changes necessary for the ligand binding cannot occur and the protein crystallizes alone. The same situation may occur during soaking experiments. The already formed protein-protein interactions in the crystal can be too strong to be disrupted upon binding of a low molecular weight molecule. A non-favorable conformation of the protein in the crystal state may not allow the compound to form stable non-covalent interaction with the active site. The situation is opposite with covalent inhibitors, which are bounded strongly enough to the protein and are able to alter its conformation.

Finally, the non-covalently bound ligands may exhibit greater thermal motion or conformational disorder than the surrounding protein, leading to a poor and insufficient electron density, preventing the identification of a molecule in the active site of an enzyme.

#### 4.4.9 Homologous 3C proteases as targets for the best compounds

Due to the good inhibitory activity of the compounds A4, EN2, EN12 and D17 against EV-B93 3C protease in the *in vitro* proteolytic assay, their potency was also tested towards the homologous picornaviral 3C proteases from HRV-A49, EV-D68, AiV, PSV and ERBV-1 (Table 4.11).

**Table 4.11 Picornaviral 3C proteases tested with the best analogs of compound 241**

Genus	Species	Serotype
<i>Enterovirus</i>	Rhinovirus A	HRV-A49
<i>Enterovirus</i>	Enterovirus D	EV-D68
<i>Kobuvirus</i>	Aichivirus A	AiV
<i>Sapelovirus</i>	Porcine sapelovirus	PSV
<i>Erbovirus</i>	Equine rhinitis B virus	ERBV-1

In the first step, the proteolytic activity of the homologous proteins was compared in the biochemical assay. 3C proteases from HRV-A49, EV-D68, AiV and PSV were able to cleave the peptide containing the Q↓G sequence, while ERBV-1 3C protease processed the same peptide with a very low yield < 5 %. For this protein a new peptide was ordered with the Q↓S sequence that mimics 2C/3A and 3A/3B cleavage site in the polyprotein of ERBV-1. The differences in the cleavage efficiency were observed. The most effective proteins appeared to be HRV-A49 and EV-D68 3C proteases, which processed the peptide completely during 3 h in 37 °C. In the case of the protease from PSV the cleavage progress was a little bit smaller (96 %), while for AiV and ERBV-1 3C proteases the cleavage efficiency was similar to those observed for EV-B93 3C protease and corresponds to 60 % (Figure 4.35).

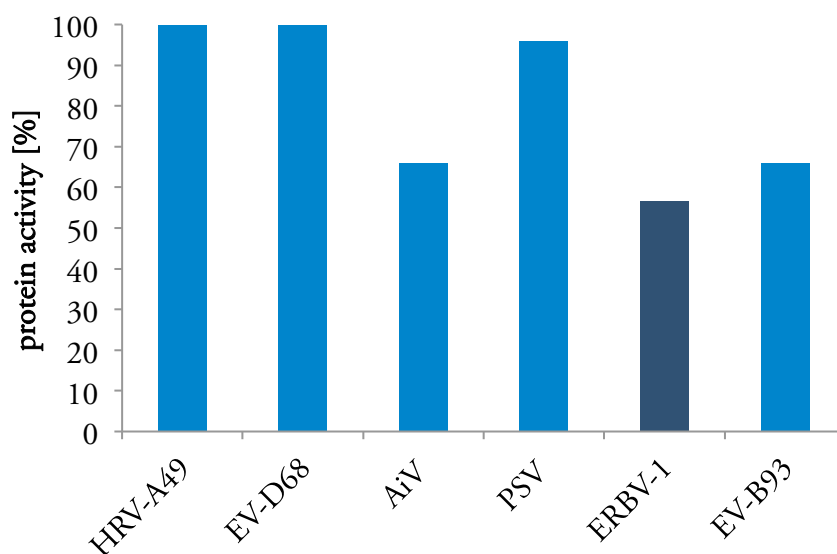
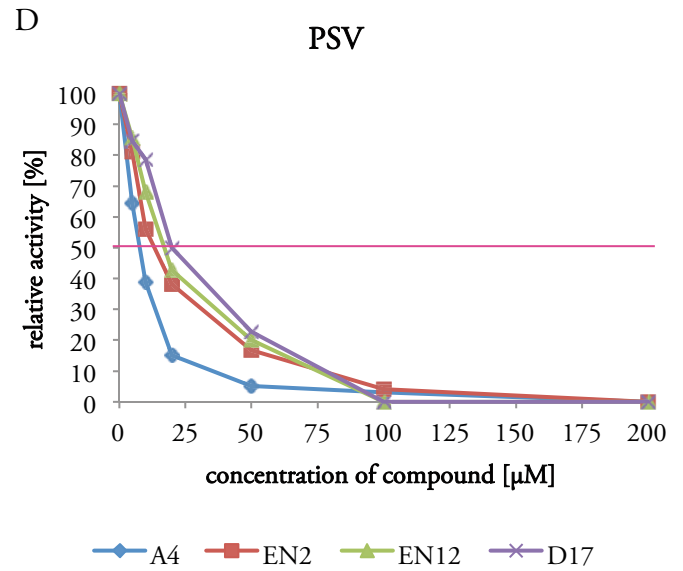
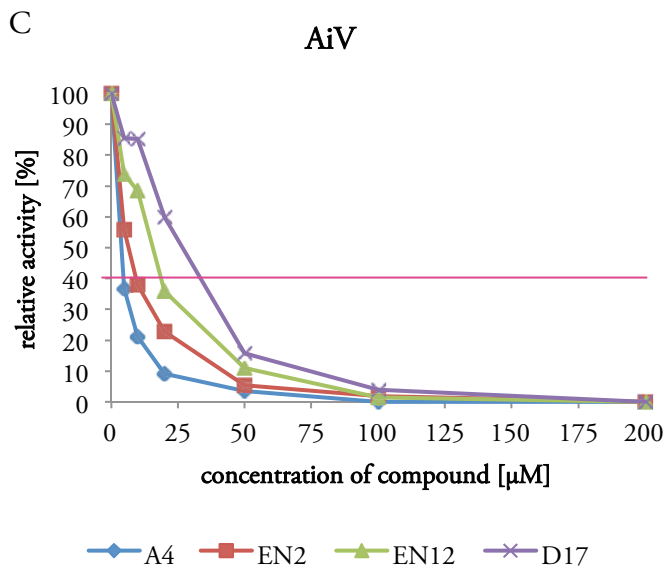
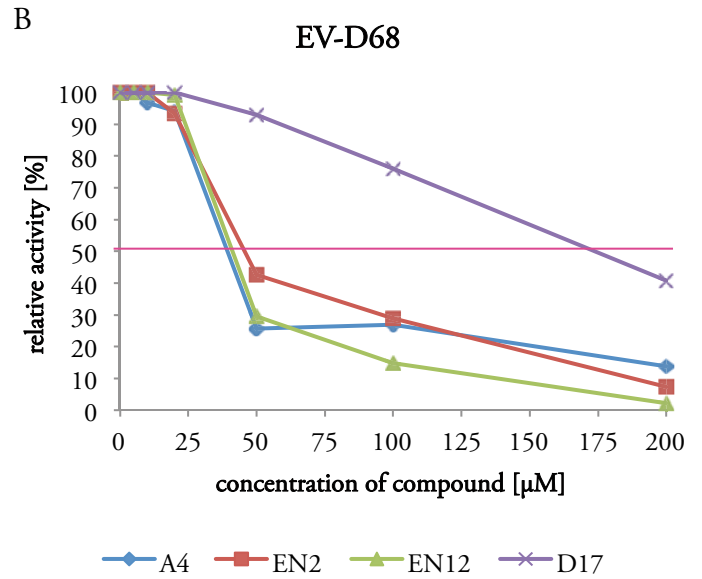
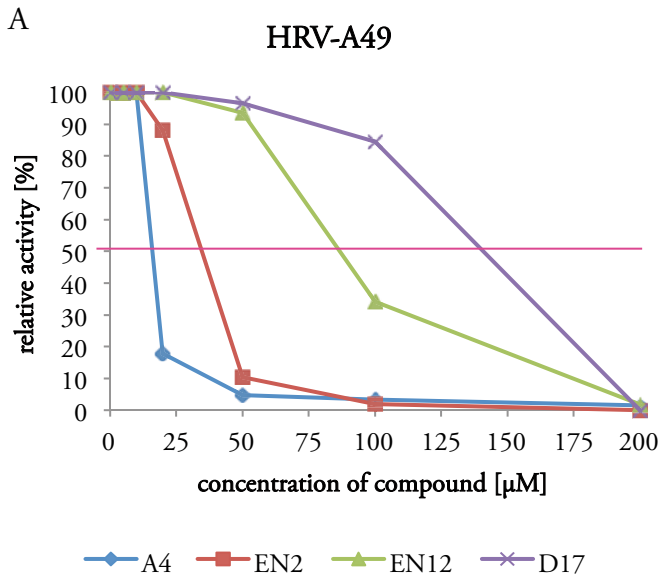


Figure 4.35 Cleavage efficiencies of 6 tested 3C proteases from HRV-A49, EV-D68, AiV, PSV, ERBV-1 and EV-B93 observed after 3 h-incubation at 37 °C. Blue and navy blue bars corresponding to peptides with the Q↓G and the Q↓S cleavage site, respectively.

In the *in vitro* assay all the compounds demonstrated inhibitory activity against homologous proteases. The potency of the analogs A4, EN2, EN12 and D17 was decreasing in the same order as for EV-B93 3C protease, i.e. the best compound was A4, and then EN2, EN12 and the least potent was D17 (Figure 4.36). The highest activity of the molecules was observed against AiV, PSV and ERBV-1 3C proteases and was similar for EV-B93 3C protease. HRV-A49 and EV-D68 3C proteases appeared to be much worse targets for A4, EN2, EN12 and D17. A4 demonstrated the best potency with IC<sub>50</sub>: 4, 8 and 5 μM for AiV, PSV and ERBV-1 3C protease, respectively. All the IC<sub>50</sub> values are gathered in Table 4.12.



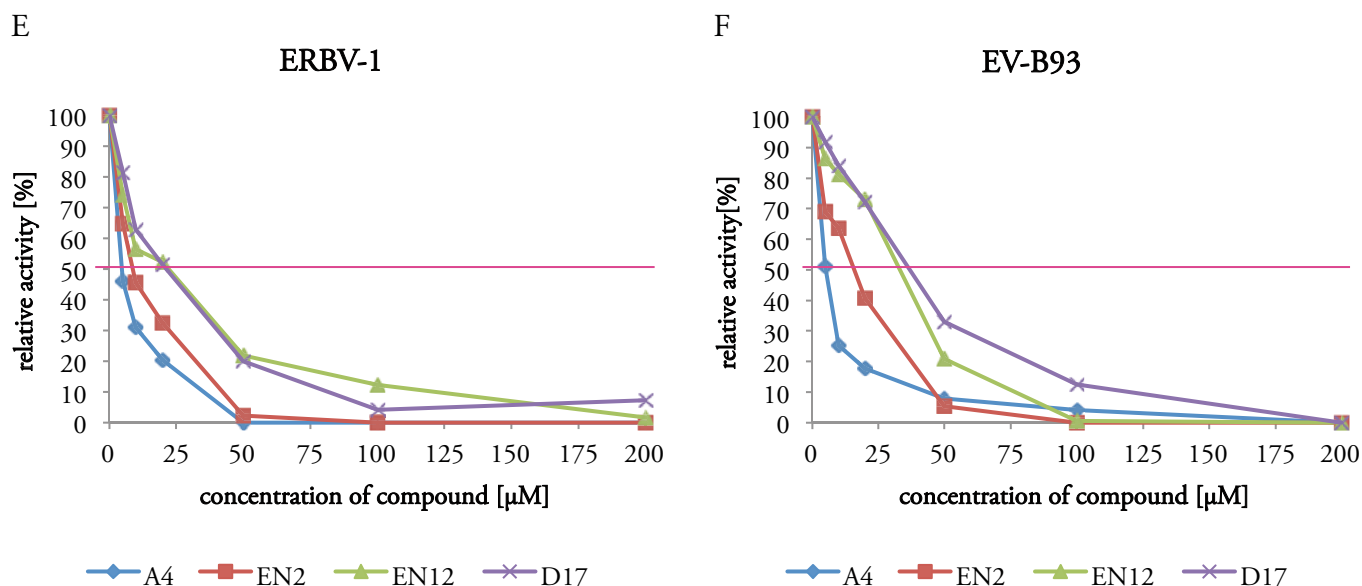


Figure 4.36 Relative activity of the homologous 3C proteases from HRV-A49 (A), EV-D68 (B), AiV (C), PSV (D), ERBV-1 (E) and EV-B93 (F) in the presence of different concentrations of A4, EN2, EN12 and D17.

**Table 4.12 IC<sub>50</sub> values [µM] of A4, EN2, EN12 and D17 for HRV-A49, EV-D68, AiV, ERBV-1 and EV-B93 3C proteases estimated on the corresponding dose-response curves (Figure 4.33)**

	HRV-A49	EV-D68	AiV	PSV	ERBV-1	EV-B93
<b>A4</b>	16	39	4	8	5	5
<b>EN2</b>	35	45	7	13	9	16
<b>EN12</b>	86	41	16	17	20	33
<b>D17</b>	140	173	27	20	20	37

Since the compound A4 was the most potent towards homologous 3C proteases, the analogs of A4 (A4\_B, A4\_C and A4\_D) were also tested. Firstly, due to the possible weak potency of A4\_B and A4\_C, the activity of the molecules was determined at one, high concentration of 350 µM (Figure 4.37). At this concentration, the inhibition of the potent compounds should reach 100 % (relative activity (RA) of the protease should be 0 %).

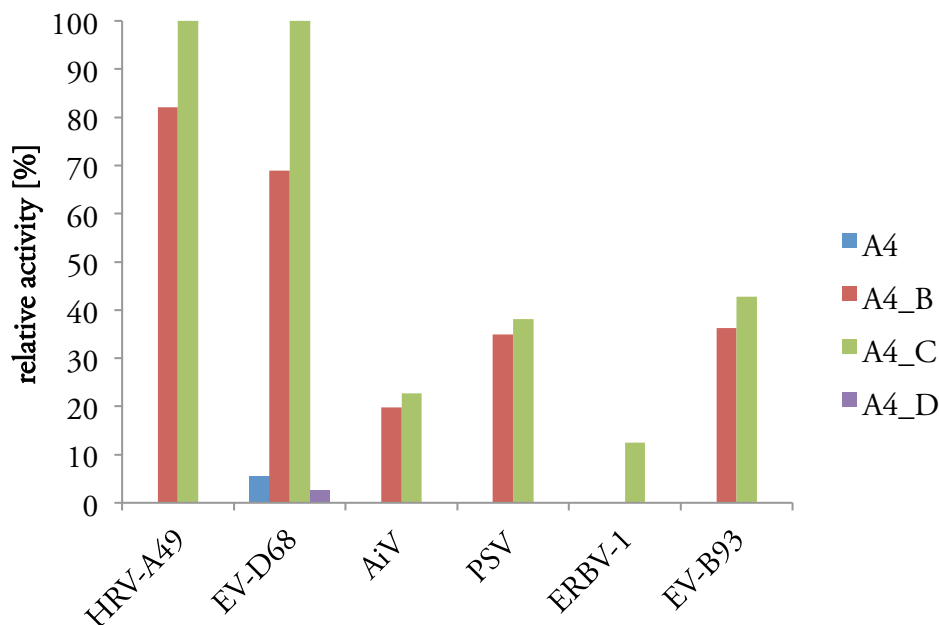
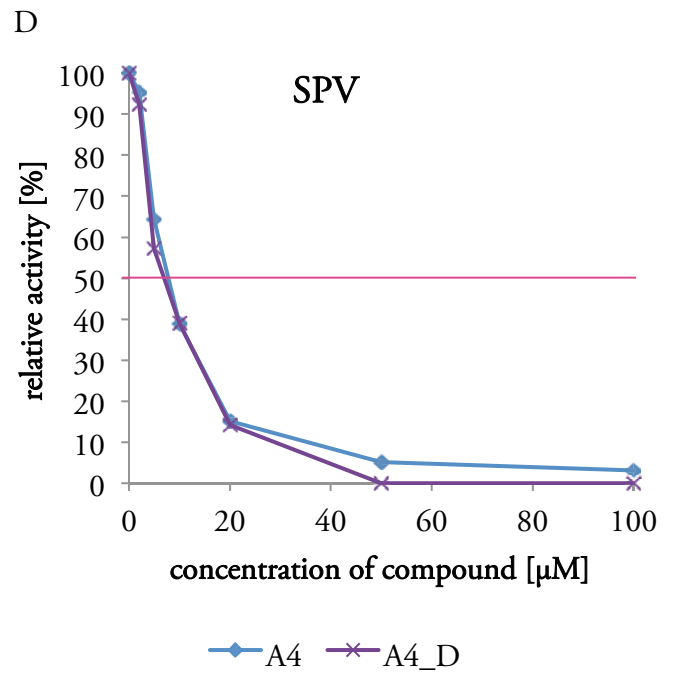
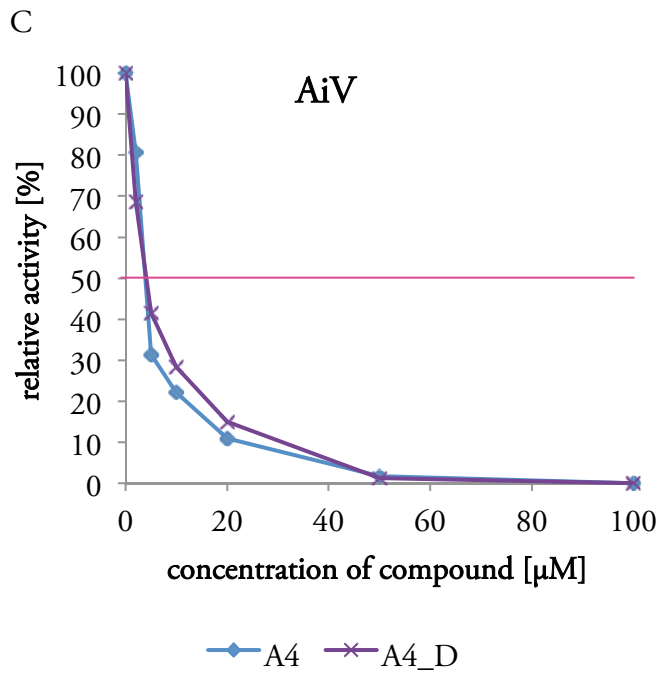
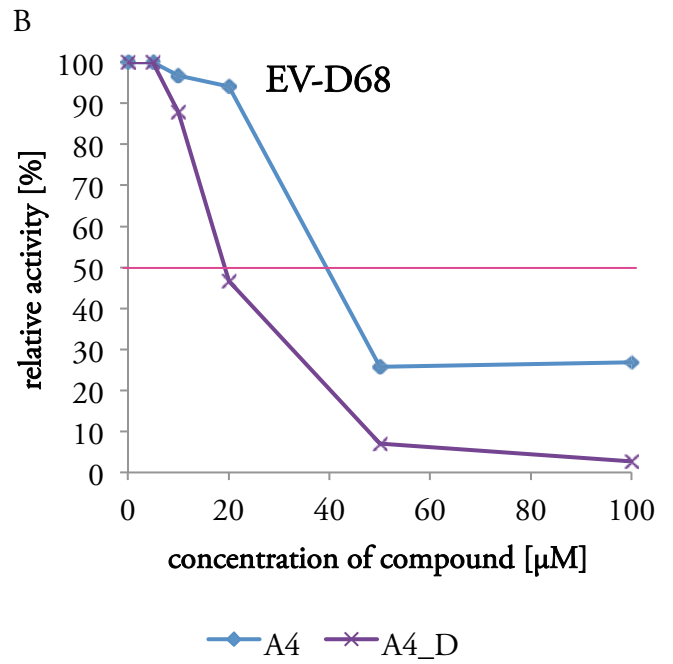
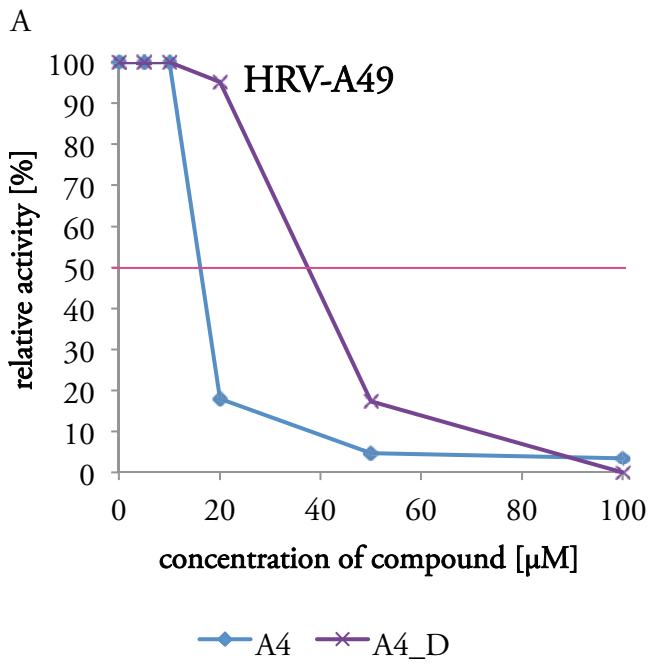


Figure 4.37 Comparison of the inhibitory activity of A4 and its analogs A4\_B, A4\_C and A4\_D at 350  $\mu\text{M}$  concentration towards 3C proteases from HRV-A49, EV-D68, AiV, PSV, ERBV-1 and EV-B93.

The analogs A4\_B and A4\_C were the weakest inhibitors towards homologous 3C proteases and the results were similar to those observed for EV-B93 3C protease. For all tested proteins the molecule A4\_C was slightly worse than A4\_B. This proves that for the inhibitors containing 4 aromatic rings, the methyl group in the phenyl ring A is favorable for the activity, which is opposite for the compounds with 3 or 2 aromatic rings. Only in the case of ERBV-1 3C protease, the molecule A4\_B was able to inhibit completely the protein activity at 350  $\mu\text{M}$ . The compound A4\_D demonstrated a very good inhibitory activity, since it inhibited almost all proteases (for 3C protease from EV-D68 the relative activity was very low 2.6 %). In order to estimate the  $\text{IC}_{50}$  values for A4\_D in the presence of the homologous 3C proteases, dose-response curves were plotted. The values were compared with those obtained for the compound A4 (Figure 4.35 A, B, C, D, E). The same curve was prepared for the compound A4\_B and ERBV-1 3C protease (Figure 4.38 F).





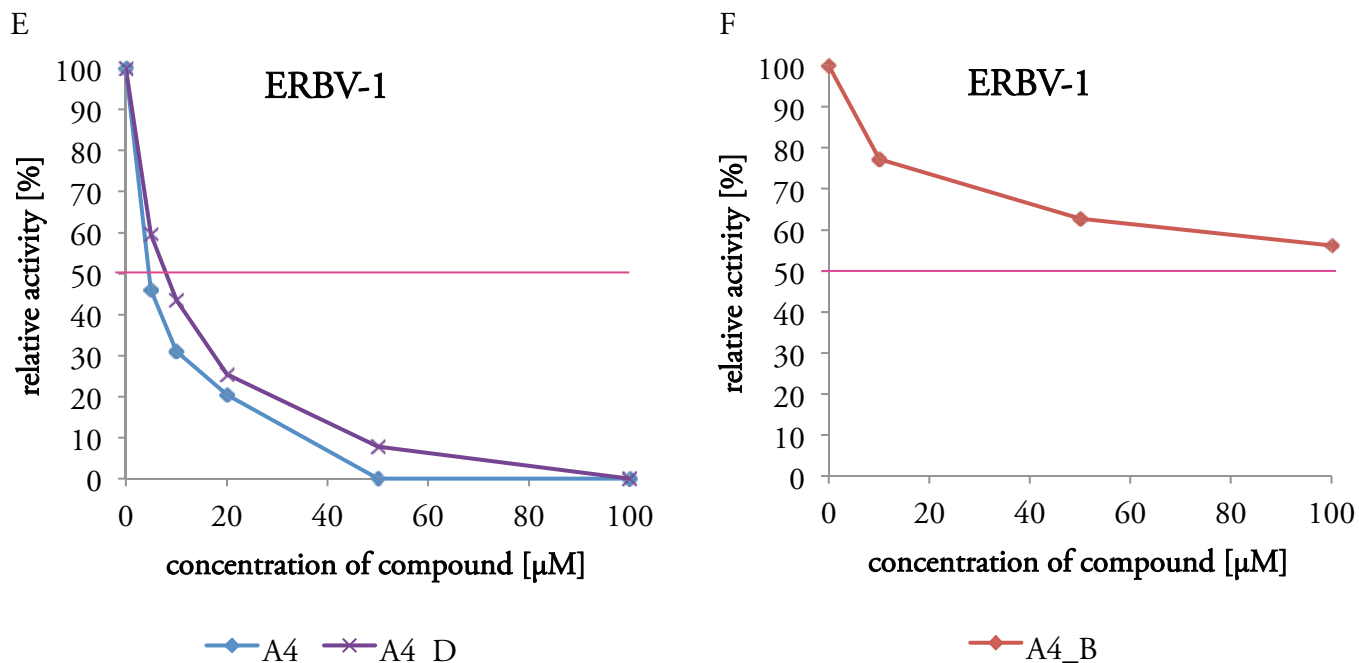


Figure 4.38 Dose-response curves of A4 and A4\_D against homologous 3C proteases from HRV-A49 (A), EV-D68 (B), AiV (C), PSV (D) and ERBV-1 (E). Dose-response curve of A4\_B against ERBV-1 3C protease (F).

The compounds A4 and A4\_D showed the same or similar activity towards 3C proteases from AiV, PSV and ERBV-1. However, in the case of the proteins against which the compounds A4 and A4\_D had poor inhibitory performance, the differences in the potency were more noticeable. For HRV-A49 3C protease, the analog A4\_D appeared to be weaker inhibitor than A4 (IC<sub>50</sub>s were ca. 37  $\mu\text{M}$  vs. 16  $\mu\text{M}$ , respectively). The opposite effect was found for EV-D68 3C protease, where molecule A4\_D showed better potency than A4 (IC<sub>50</sub>s were ca. 19  $\mu\text{M}$  vs. 39  $\mu\text{M}$ , respectively). The compound A4\_B was very active against ERBV-1 3C protease at 350  $\mu\text{M}$ , but was not able to reduce the activity of the protein by 50 % at the concentration lower than 100  $\mu\text{M}$ .

EV-B93, EV-A49, EV-D68, AiV, PSV and ERBV-1 belong to a different picornaviral genus. EV-B93, EV-A49, EV-D68 are part of the *Enterovirus* genus, while AiV, PSV and ERBV-1 are classified within the *Kobuvirus*, *Sapelovirus* and *Erbovirus* genus, respectively. Multiple sequence alignment of the 3C proteases from these viruses revealed that EV-B93 3C protease shares the highest sequence similarity with EV-D68, PSV and EV-A49 3C proteases (66.67 %, 52.46 % and 48.63 %, respectively). However, when comparing to AiV and ERBV-1 3C proteases, the sequence identity is very low (14.75 % and 10.93 % respectively) (Table 4.12).

**Table 4.12 Sequence similarity of 3C proteases from EV-B93, EV-A49, EV-D68, AiV, PSV and ERBV-1.**

	<b>EV-B93</b>	<b>EV-A49</b>	<b>EV-D68</b>	<b>AiV</b>	<b>PSV</b>	<b>ERBV</b>
<b>EV-B93</b>	-	48.63 %	66.67 %	14.75 %	52.46 %	10.93 %
<b>EV-A49</b>	48.63 %	-	48.63 %	19.67 %	45.36 %	13.66 %
<b>EV-D68</b>	66.67 %	48.63 %	-	13.66 %	46.45 %	19.13 %
<b>AiV</b>	14.75 %	19.67 %	13.66 %	-	18.03 %	13.61 %
<b>PSV</b>	52.46 %	45.36 %	46.45 %	18.03 %	-	17.49 %
<b>ERBV</b>	10.93 %	13.66 %	19.13 %	13.61 %	17.49 %	-

Multiple sequence alignment was performed separately between EV-B93 3C protease and more similar 3C proteases from EV-A49, EV-D68 and PSV, and between EV-B93 3C protease and more distant AiV and ERBV 3C proteases (Figure 4.39 A and B, respectively). In the case of the first alignment, most of the residues present in the proteolytic active pocket are conserved. For the second alignment the sequence similarity was much less noticeable. Only a few amino acids were identical or physico-chemically similar, including one residue from the catalytic triad. The Glu-71 in EV-B93 3C protease corresponds to Asp-90 in 3C protease from ERBV-1. However, the majority of the invariant residues form the proteolytic active site.

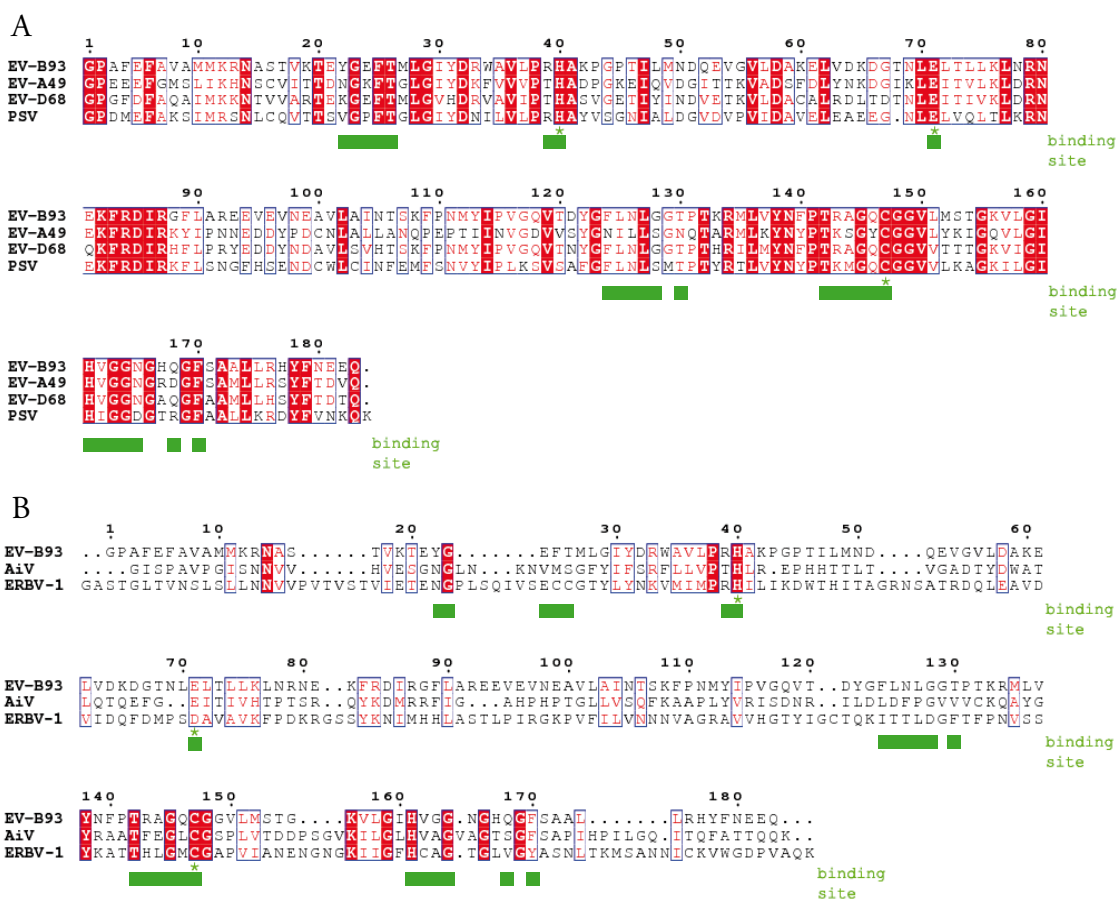


Figure 4.39 Multiple sequence alignment of 3C proteases from EV-B93, EV-A45, EV-D68 and PSV (A) and from EV-B93, AiV and ERBV-1 (B). Invariant residues are white with red background; conserved residues are shown in red font. The residues of proteolytic binding site are marked below as green bars. Catalytic residues Cys-147, His-40 and Glu-71/Asp-90 are indicated by green asterisks.

Despite the high degree of the sequence similarity among the enteroviral 3C proteases, the inhibitory activities of the compounds discovered within this study were very different. For instance the molecule A4 exhibited the IC<sub>50</sub> ranging from 5  $\mu$ M for EV-B93 3C protease to 16 and 39  $\mu$ M for EV-A49 and EV-D68 3C proteases, respectively. In the case of the non-enteroviral 3C proteases, the active molecules (A4, EN2, EN12 and D17) displayed higher potency with the IC<sub>50</sub>s close or better to the ones obtained for EV-B93 3C protease (see the values in Table 4.12). The observed differences may be related not only to the affinity of a compound to the proteolytic active site but also to the peptide substrate specificity. 3C proteases from EV-A49 and EV-D68 were able to process the peptide substrate completely during 3 h in 37 °C, which did not occur for other homologous proteins including EV-B93 3C

protease (Figure 4.35). Moreover, it is worth mentioning that the cleavage specificity of ERBV-1 3C protease is different from the other tested enzymes, since it recognizes the Q↓S, not the Q↓G sequence. This may influence the affinity of its possible inhibitors, for example the analog of A4 – A4\_B was able to reduce the proteolytic activity of ERBV-1 3C protease entirely at 350 μM, while for the homologous 3C proteases only some activity was observed (see Figure 4.36).

#### 4.4.10 Cell-based assays with the best compounds

Fragment screening and optimization of the best hit resulted in the potent non-covalent inhibitors that displayed good inhibitory activity against isolated 3C proteases not only from EV-B93 but also from the homologous picornaviral 3C proteases. Performing cell-based assays were aimed at investigating whether the inhibitors are able to enter the cells, are stable in the cytoplasmic environment, can inhibit the 3C protease in this condition, or if they are toxic to cells. Two different assays were applied in parallel: the cell-based proteolytic assay and the antiviral assay.

##### 4.4.10.1 Cell-based proteolytic assay with the best compounds

The first method allows the verification if the compounds are able to inhibit 3C proteases from CV-B3 or EV-A71 in living cells. The assay, developed by our collaborators from the Utrecht University in The Netherlands, is based on the induction of expression of a firefly luciferase reporter by a chimeric transcription factor [123]. The transcription factor consists of a part of the CV-B3 or EV-A71 polyprotein (15 C-terminal amino acids of 3A, 3B, 3C<sub>pro</sub>, and 15 N-terminal acids of 3D) flanked by GAL4 binding domain (GAL4BD) and VP16 activation domain (VP16AD). Active protease cleaves the chimeric protein at the 3C protease cleavage sites (arrows). If the protease is catalytically inactive, binding of GAL4BD to the GAL4 sequences in the reporter plasmid recruits VP16AD to the transcription start site, resulting in induction of Firefly luciferase expression (Figure 4.40). The use of the inhibitor SG85 or CA (Cys147A mutant of CV-B3) served as a positive control, while native protease without any addition of the inhibitor served as a negative control.

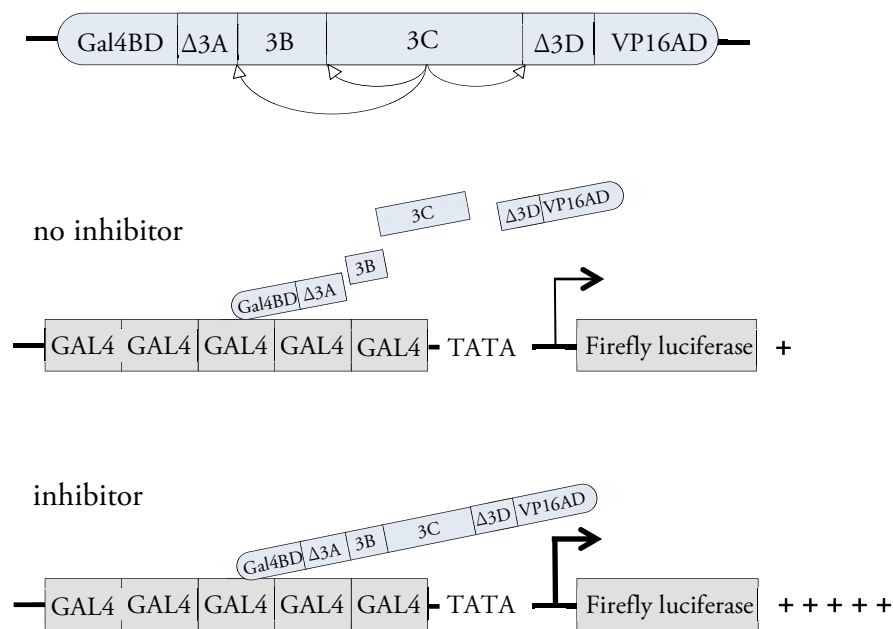
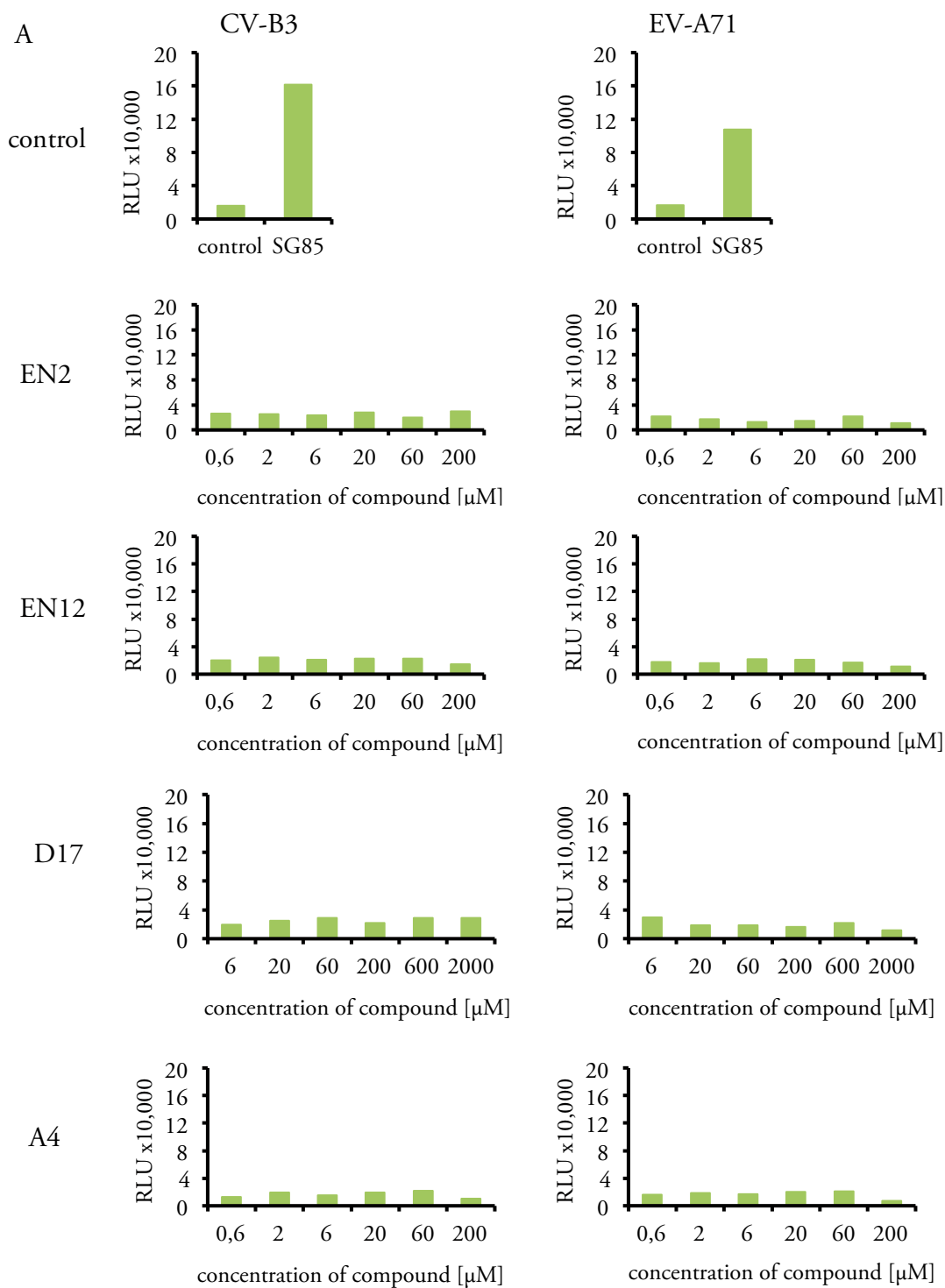


Figure 4.40 Schematic representation of the cell-based proteolytic assay (adapted from [123]).

All the compounds EN2, EN12, D17, A4 and its analogs A4\_B, A4\_C and A4\_D were tested in the cell-based proteolytic assay. For the analogs EN2, EN12 and A4 the maximum concentration applied was 200  $\mu$ M, for the compound D17 it was 2 mM, while for the compounds A4\_B, A4\_C and A4\_D the highest concentration tested was equal to 30  $\mu$ M. These values depended on the solubility or cytotoxicity of the compounds (see section 4.4.10.2). The inhibitory activity was checked towards CV-B3 and EV-A71 3C proteases (Figure 4.41 A and B).





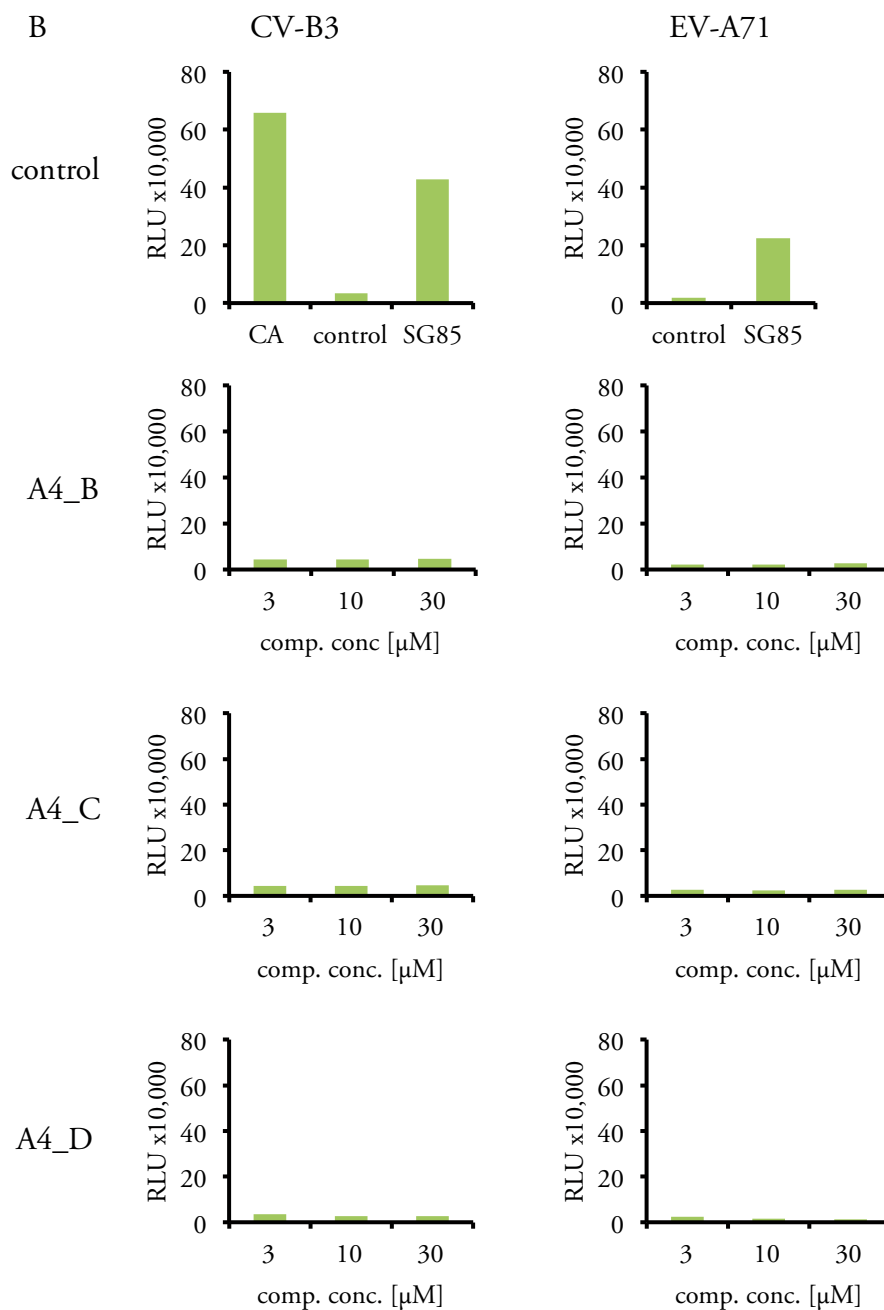


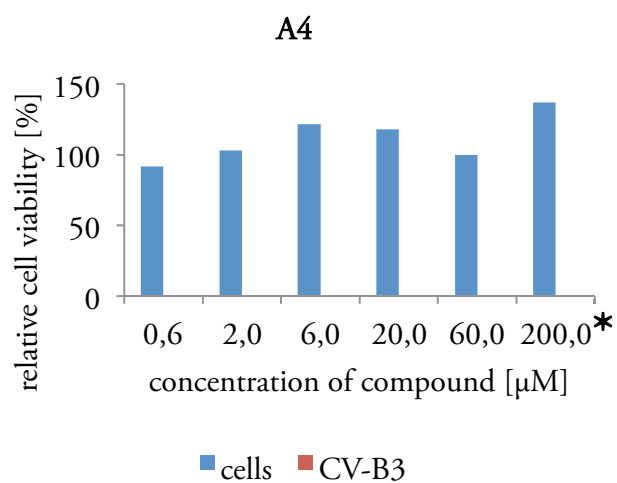
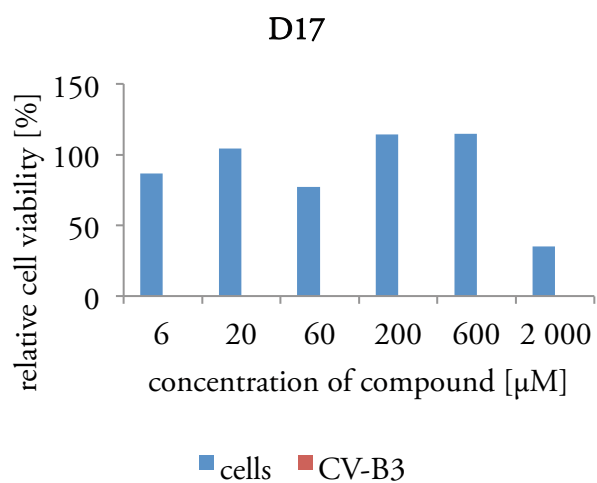
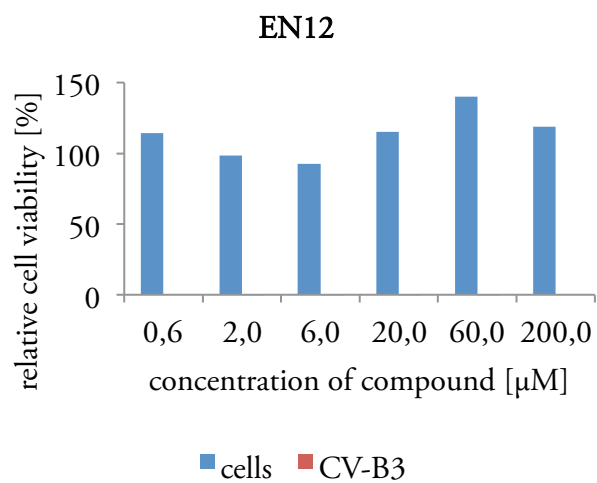
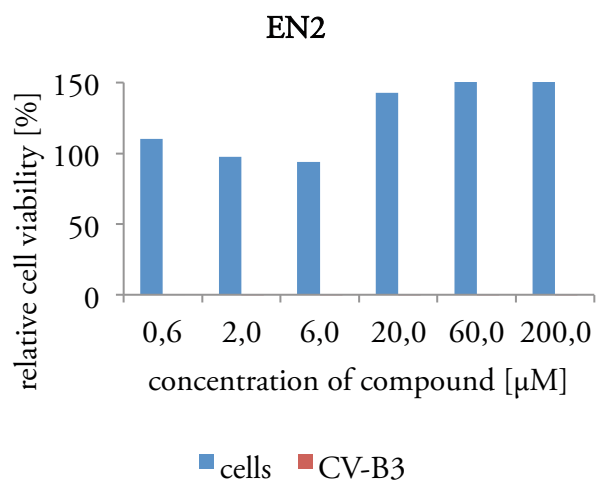
Figure 4.41 Results obtained in the cell-based proteolytic assay for the compounds EN2, EN12, D17 and A4 (A) and for the compounds A4\_B, A4\_C and A4\_D (B). Induction of the Firefly luciferase expression was measured in RLU (Relative Light Unit). SG85 (inhibitor of 3C protease) used in the concentration 50  $\mu\text{M}$  and CA (Cys147A mutant of CV-B3) served as positive controls.

Unfortunately, no inhibition could be detected for any of the compounds. Most probably the molecules were not able to enter the cells or were not stable in the applied conditions. Another hypothesis could be related to the insufficient solubility of the inhibitors that were unable to reach minimal necessary concentration. The sequence similarity between 3C protease from CV-B3 (used in the assay) and EV-B93 is 97 %. Only 6 amino acids are different and they are not placed in the proteolytic active side, thus suggesting that the lack of inhibitory activity of compounds is not caused by the differences in the proteases that were used in the assay.

#### 4.4.10.2 Cytopathic effect (CPE)-based antiviral assay

In parallel another cell-based method was used in order to check if the compounds are able to extend the lifespan of the virus-infected cells. The *in vitro* cytopathic effect (CPE)-based assay is a one-step procedure for evaluating antiviral efficacy of a drug. Virus-susceptible cells (BGM cells) were treated with a test compound and then infected with a virus (CV-B3 – genus: Enterovirus, species: Enterovirus B or ERAV genus: Aphovirus, species: Equine rhinitis A virus). After a 3 day-post-inoculation period, untreated cells exhibited full CPE. Cell viability was assessed using MTT assay reagent and results were confirmed by microscopy. In this assay, the inhibitors of virus replication were detected by an increase in the viability of the virus infected cells. Cells not infected by the virus, but treated with the inhibitors in different concentrations served as a toxicity control (Figure 4.42 A and B).

A



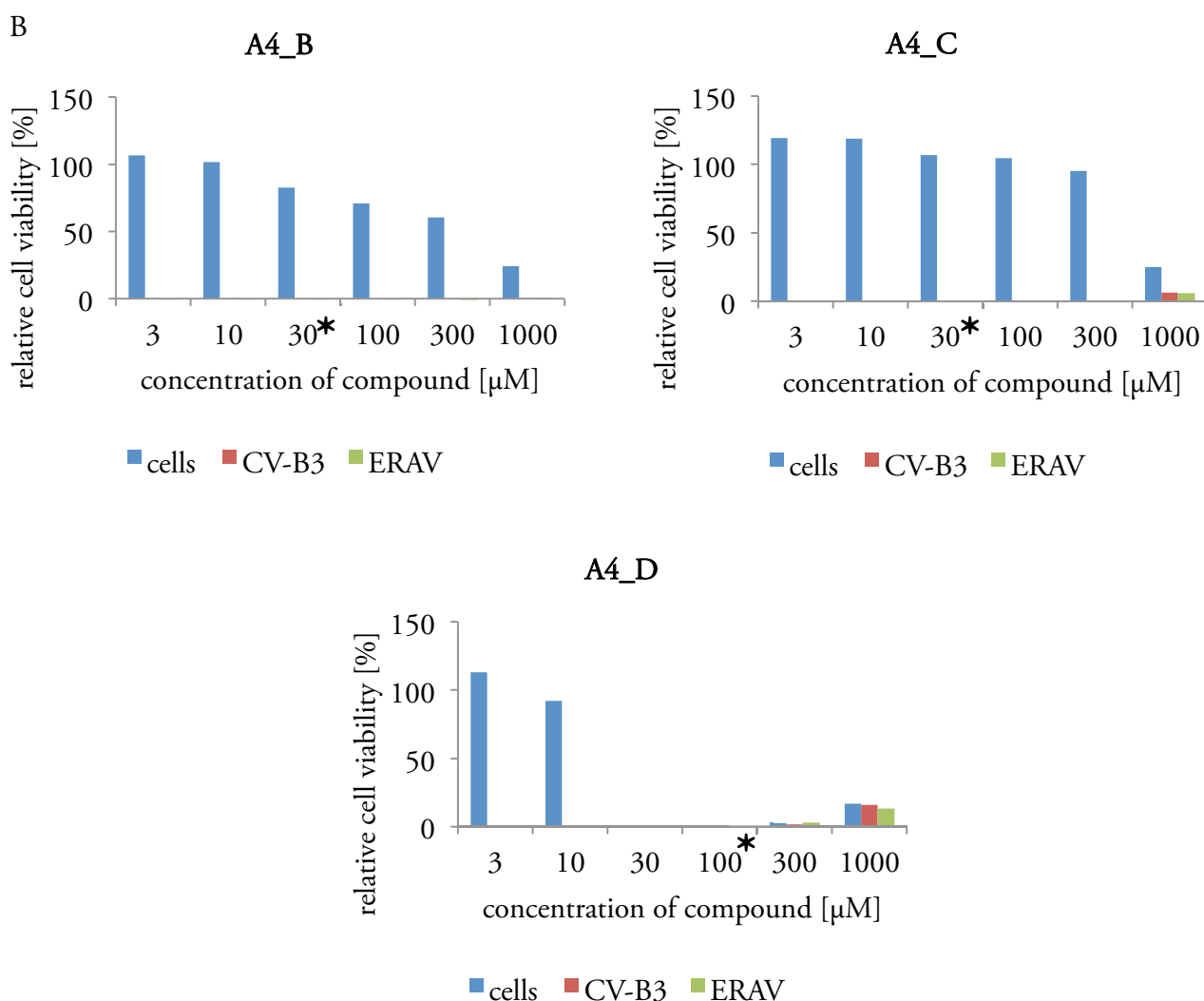


Figure 4.42 Results obtained in the CPE-based antiviral assay for EN2, EN12, D17 and A4 (A) and for A4\_B, A4\_C, A4\_D (B). Blue data series corresponds to the uninfected cells treated with different concentrations of the compounds. Red and orange data series correspond to CV-B3- and ERAV-infected cells, respectively, treated with different concentrations of the compounds. Asterisk indicates the lowest concentration with compound precipitate. Relative cell viability was adjusted to pretreated conditions.

Compounds A4, EN2, EN12 and D17 were tested in the cells infected only with CV-B3, while the molecules A4\_B, A4\_C and A4\_D were evaluated in the cells infected by both CV-B3 and ERAV. The results clearly show that none of the compounds exhibited antiviral activity. A4, EN2 and EN12 were not cytotoxic at 200  $\mu\text{M}$  (the highest concentration tested), while D17 showed cytotoxicity at 2 mM (the highest concentration tested). Furthermore, in the case of the analogs of A4, the compound A4\_D showed the highest cytotoxicity at 30  $\mu\text{M}$ . The molecules A4\_B and A4\_C exhibited strong cytotoxicity at 1 mM and were the least

soluble compounds giving some precipitation at 30  $\mu$ M. The results are in agreement with the ones obtained previously in the cell-based proteolytic assay.

Lipophilicity expressed as clogP is especially important drug-like property, the control of which is important for ultimate success of the drug development. ClogP influences on drug potency, pharmacokinetics and toxicity since it governs the cell membrane permeability and affinity to usually hydrophobic protein binding sites. If lipophilicity is too high, there is an increased likelihood of binding to multiple targets that results in toxicity as well as poor solubility.

Lack of inhibitory activity and toxicity of the compounds EN2 and D17 may be caused by relatively low clogP values equal to 2.35 and 1.40, respectively. Polarity and the presence of ionizable carboxylic group influence the cell membrane permeability and make the molecules unable to reach their target. On the other hand, the compounds EN12 and A4 seem to be sufficiently lipophilic (clogP equal to 2.9 and 4.3, respectively) to enter the cells, however no observed inhibitory effect and poor solubility suggest their inability to reach proper inhibitory concentration outside and inside the cells. Finally, the compound A4\_D showed moderate lipophilicity (clogP 3.2) and the strongest toxicity among all tested compounds, which may indicate that it was successful in crossing the cell membranes. Unfortunately, the compound was not specific enough towards the 3C protease, which resulted in cell toxicity. However, further chemical optimization of the molecule A4\_D modulating its affinity and selectivity may lead to a potent non-covalent inhibitor of the enteroviral 3C proteases. Lack of experimental data on the binding mode of A4\_D compelled to perform bioinformatics simulations that may finally help to infer the requirements for its association to the protein and suggest potency-improving modifications.

#### 4.4.11 Bioinformatics simulations of A4\_D binding mode

In parallel to the crystallization trials, bioinformatics calculations were conducted by Daniel Álvarez from the laboratory of prof. Xavier Barril to elucidate the binding mode of the active compounds, including EN2, EN12, D17, A4 and A4\_D. Unfortunately, the docking simulations failed, revealing no clear reliable binding pose. For smaller molecules (e.g. EN2), multiple conformations were possible, while in the case of A4 or A4\_D the calculated positions locate the compound partially outside of the active pocket. The biochemical analysis confirmed that both parts of A4\_D (EN2 and PA4) are essential for the potency of the molecule. EN2 as well as PA4 are able to bind independently to EV-B93 3C protease, but with lower affinities than the final A4\_D compound (see section 4.4.3). The reason why the docking was not successful for the best compound A4\_D may be the fact that the molecule does not fit to the active site of the protein in the conformation present in the crystal structure. The binding

simulations were calculated with the use of the previously solved crystal structure of the unbound EV-B93 3C protease. The conformation of the enzyme may differ significantly from the one indispensable for creating interaction with A4\_D, thus making this protein model unsuitable for docking purposes. The results described in the literature mentioned the presence of a highly flexible surface loop in the homologous enteroviral (EV-A71, PV, HRV, CV-B3) 3C proteases, important for recognizing the P2-P4 region of the peptide substrates or inhibitors [64, 92, 145, 146]. Presumably, some conformational changes should occur in the structure of EV-B93 3C protease to make enough room for larger and not flexible A4\_D molecule.

For this reason, modeling of the EV-B93 3C protease-A4\_D complex was carried out using molecular dynamics (MD) simulations. Firstly, the simulations were performed with the protein alone in a mixture of pyridine/water to obtain the information on the conformational flexibility of the protease active site (Figure 4.43). Pyridine was chosen to force larger conformational changes of the protein. By applying this strategy, it was possible to look for the cavity openings induced by a pyridine ring and investigate which regions of the protein are interacting more preferably with a pyridine molecule.

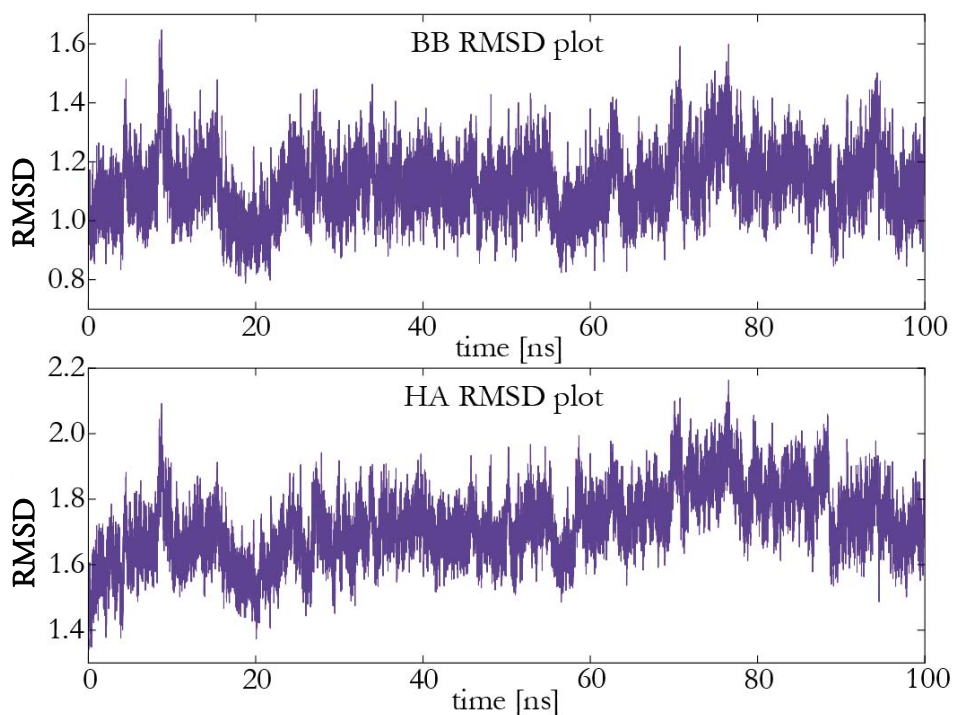


Figure 4.43 The RMSD values of the EV-B93 3C protease backbone (A), and the backbone and side chains (B) during 100 ns of MD simulations in the pyridine/water mixture.

Monitoring the root mean square deviation (RMSD) values of the protein can give insights into its structural conformation throughout the simulation, providing an indication of the

stability of the protein and whether the simulation has equilibrated. The RMSD values for EV-B93 3C protease did not fluctuate much, thus showing that the protein is stable compact, and not prone to unfolding process. However, in the presence of the pyridine/water mixture, the flexible loop formed by the Leu-127, Gly-128, Gly-129 and Thr-130 separating the S2 from the S3/S4 sub-pocket slightly opened forcing the changes in the conformation of the Phe-170, His-40, Tyr-25, and Lys-42, thereby making the active pocket more accessible for a ligand. The aforementioned protein conformation obtained in the first step of the MD simulations and the best pose of the compound A4\_D from the docking studies were subjected to the subsequent MD simulations. The RMSD values were monitored to control changes and stability of the binding position of the molecule A4\_D during the simulation process (Figure 4.44).

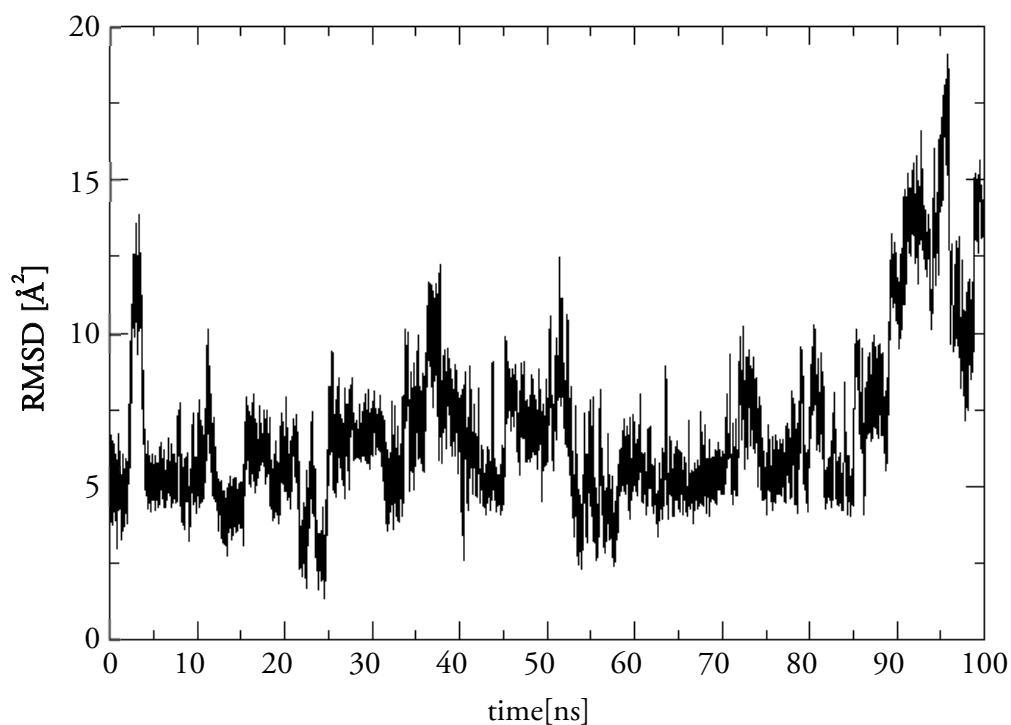


Figure 4.44 The RMSD values of A4\_D in the presence of EV-B93 3C protease during 100 ns of MD simulations.

Figure 4.44 shows that not until 90 ns of MD simulations the ligand was able to change significantly its position and found stable conformation within the protease active pocket. The differences in the ligand binding poses throughout the simulations are represented in the Figure 4.45.

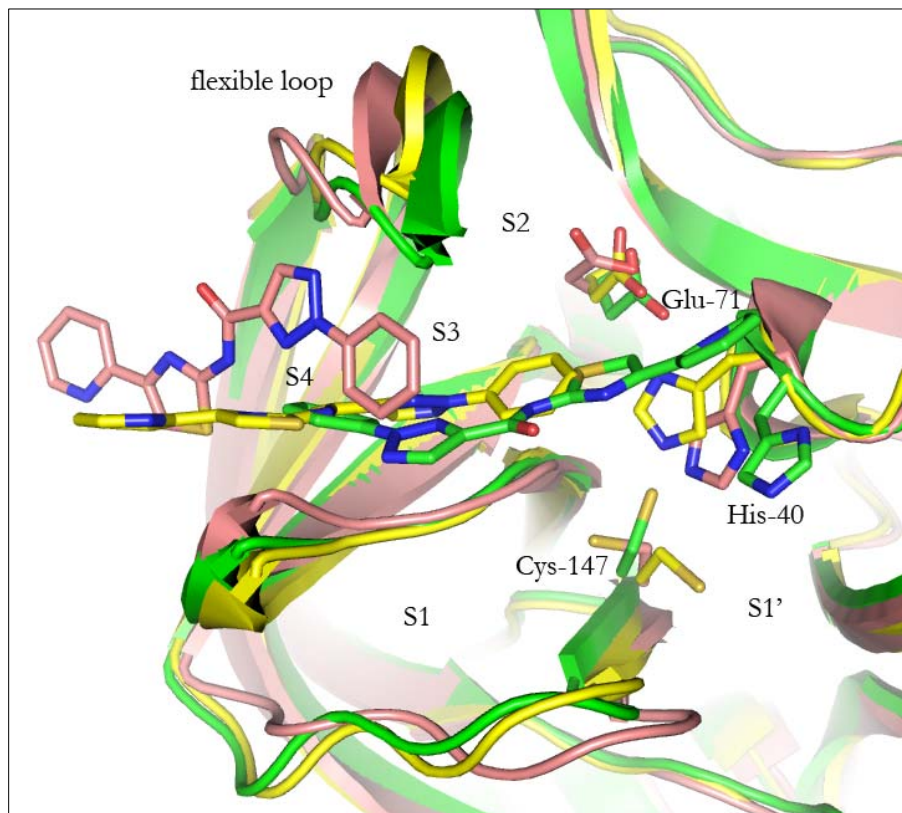


Figure 4.45 Binding poses of A4\_D throughout the MD simulations. Initial (0 ns), intermediate (50 ns) and final (100 ns) pose is represented as pink, yellow and green sticks, respectively. The residues from the catalytic triad are labeled.

The initial binding conformation of A4\_D does not differ much from the intermediate pose. Only after the final binding mode is completely flipped, it allows the stacking interaction of pyridine ring with His-40 and water-mediated interaction with the Glu-71, two conserved amino acids from the catalytic triad (Figure 4.46 A). This interaction may explain the importance of the nitrogen atom in the ring D, whose absence would render the interactions with the Glu-71 impossible.



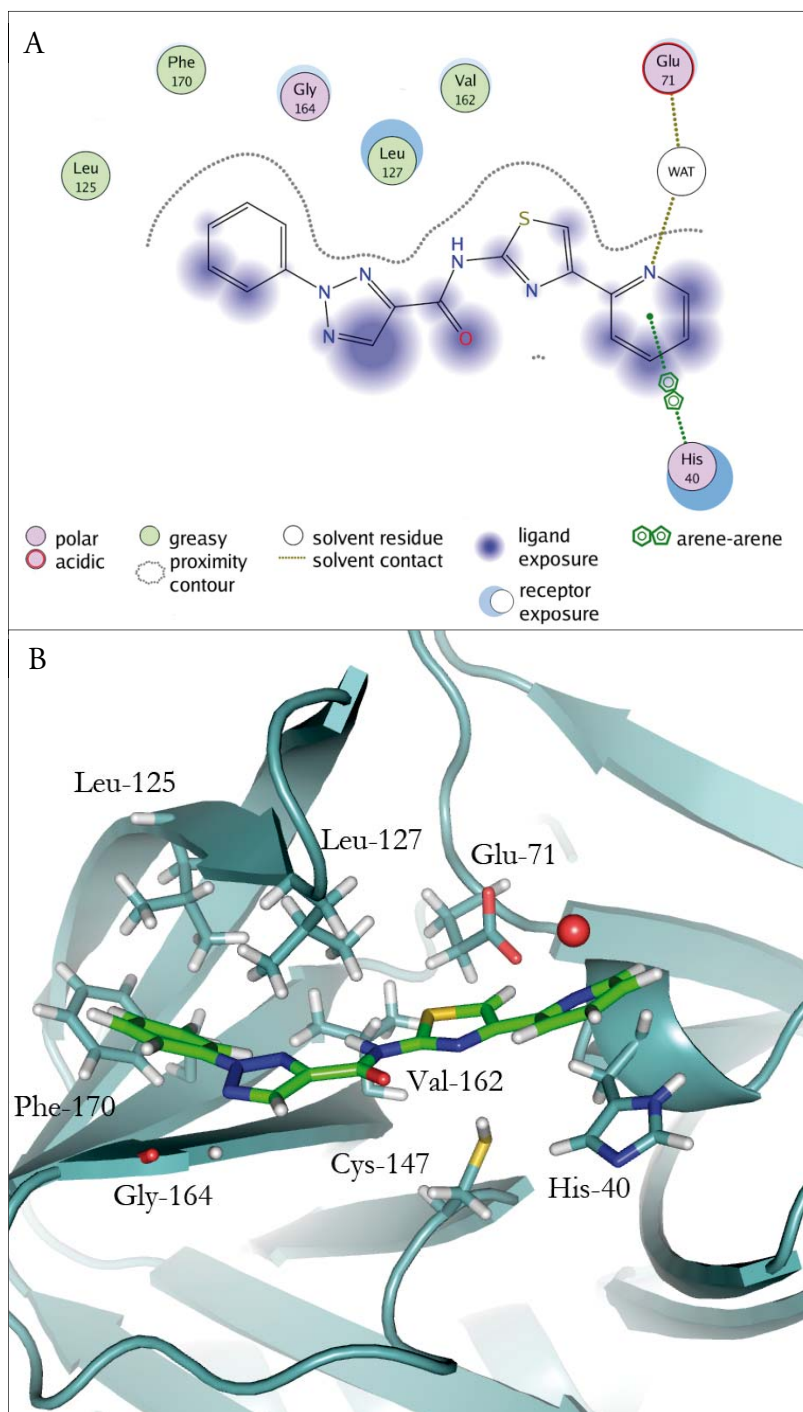


Figure 4.46 EV-B93 3C protease-A4\_D interaction diagram (A), and 3D EV-B93 3C protease -A4\_D interaction model (B). The inhibitor is represented as green sticks. Water molecule is represented as red sphere. Amino acid side chains interacting with A4\_D are represented as celadon sticks and are labeled.

Compound A4\_D occupies the S4, S3 and partially the S2 sub-pocket, changing the shape of the last considerably. Analyzing the protein-ligand interaction model (which is one of the many possible, since the protein and the inhibitor molecule oscillate continuously), it is noticeable that the most frequent interactions are hydrophobic (Figure 4.46 B). They are formed by the phenyl ring A with the Leu-125, and Phe-170 as well as the sulphur from the ring C with the Val-162 and Leu-127. Compound A4\_D is, to a large degree, solvent exposed, therefore, not many residues may be subjected to affinity-improving modifications. However, there are some possibilities. Firstly, the nitrogen atom in the ring D should be substituted by a hydroxyl group, which may interact directly with the Glu-71. Additionally, the nitrogen atom from the peptide bond should be replaced by a carbon atom since it is buried in the molecule and is not able to form any interaction with the protein. Moreover, its proximity to the carbonyl oxygen from the amide bond may be unfavorable in terms of the binding stability of the whole molecule.

Results obtained in the simulations helped to explain the probable inability of the compound A4\_D to interact with EV-B93 3C protease in a crystalline state. Cysteine and serine proteases contain an active cavity composed of several sub-pockets, to which excessively extended molecules are not able to bind without its previous rearrangement. The compound A4\_D can serve as an example of such a molecule since it has a relatively large and rigid structure (Figure 4.47).

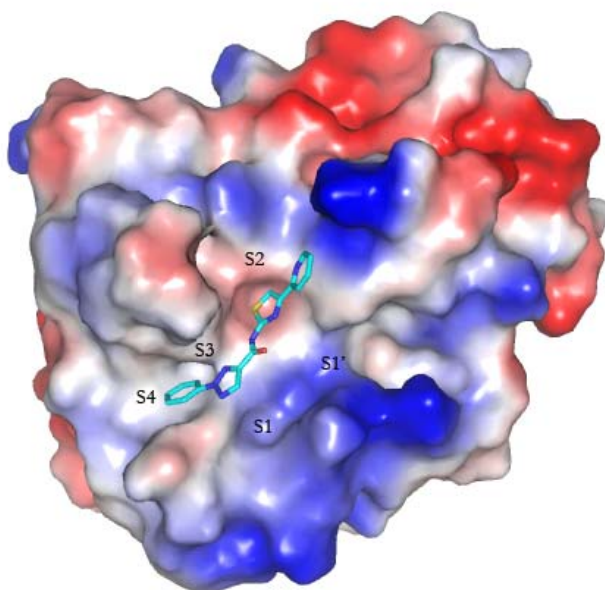


Figure 4.47 Model of the proteolytic active site of EV-B93 3C protease in complex with A4\_D. Protein surface is colored according to the electrostatic potentials with red and blue indicating negative and positive charges, respectively. Inhibitor is represented as turquoise sticks. The sub-pockets of the active site are labeled.

The experimental data revealed its lack of interaction with the crystallized EV-B93 3C protease and suggests that the protein conformation, which is indispensable for the binding of A4\_D, is not favorable in the crystal structure. In order to improve the potency of A4\_D, the precise binding site should be determined experimentally. Since all methods used by us failed in achieving this goal, we propose to use an additional approach to describe the EV-B93 3C protease-A4\_D interactions:  $^{15}\text{N}/^1\text{H}$  HSQC (Heteronuclear Single Quantum Coherence). This technique is considered one of the best assays for detecting protein-ligand interactions [147]. The HSQC is able to provide detailed information on which amino acids from a protein interacts with an inhibitor. Because of its time consumption it was not possible to perform this assay in the time frame of this thesis.

# 5. Conclusions



1. An *in vitro* assay was established to test the proteolytic activity of EV-B93 3C protease.
2. Crystallization screening of EV-B93 3C protease resulted in needle-like crystals that diffract up to 1.57 Å. The structure revealed that the protein adapts a chymotrypsin-like fold similarly to other picornavirus 3C proteases and possesses His-40, Glu-71 and Cys-147 as a catalytic triad.
3. Rupintrivir and compound 1 efficiently inhibited EV-B93 3C protease in the *in vitro* proteolytic assay.
4. Preparation of the complex of EV-B93 3C protease with rupintrivir and compound 1 and its crystallization resulted in rhombohedron-like crystals that diffracted up to 1.50 and 1.32 Å, respectively.
5. NZO, NZN and DB5\_60 showed good inhibitory activity against EV-B93 3C protease in the *in vitro* proteolytic assay.
6. The 1.73 Å-resolution crystal structure of EV-B93 3C protease in complex with NZN compound was determined and shows the molecule to be attached in the S1 sub-pocket of the proteolytic active site.
7. The STD NMR-based screening of EV-B93 3C protease resulted in the selection of 51 molecules that were subsequently validated. Compound 241 displayed the highest affinity and the highest inhibitory activity *in vitro*.
8. 44 analogs of compound 241 were purchased and tested in the *in vitro* proteolytic assay. 4 molecules exhibited strong inhibitory activity against EV-B93 3C protease.
9. The best molecule A4 was subjected to chemical optimization and resulted in compound A4\_D with a similar IC<sub>50</sub> value, but better efficiency.
10. Selected inhibitors were taken for further evaluation with the homologous 3C proteases from EV-A49, EV-D68, AiV, PSV and ERBV-1. The compounds displayed similar or better inhibition towards 3C proteases from AiV, PSV and ERBV-1.
11. Bioinformatics simulations suggested a binding pose for compound A4\_D. These calculations indicated that a partial reorganization of the active cavity of EV-B93 3C protease is required for ligand binding.



# 6. Bibliography





1. Orloff, J., et al., *The future of drug development: advancing clinical trial design*. Nat Rev Drug Discov, 2009. **8**(12): p. 949-57.
2. Shuker, S.B., et al., *Discovering high-affinity ligands for proteins: SAR by NMR*. Science, 1996. **274**(5292): p. 1531-4.
3. Hajduk, P.J., R.P. Meadows, and S.W. Fesik, *Discovering high-affinity ligands for proteins*. Science, 1997. **278**(5337): p. 497,499.
4. de Kloe, G.E., et al., *Transforming fragments into candidates: small becomes big in medicinal chemistry*. Drug Discov Today, 2009. **14**(13-14): p. 630-46.
5. Boehm, H.J., et al., *Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening*. J Med Chem, 2000. **43**(14): p. 2664-74.
6. Edwards, P.D., et al., *Application of fragment-based lead generation to the discovery of novel, cyclic amidine beta-secretase inhibitors with nanomolar potency, cellular activity, and high ligand efficiency*. J Med Chem, 2007. **50**(24): p. 5912-25.
7. Hopkins, A.L., C.R. Groom, and A. Alex, *Ligand efficiency: a useful metric for lead selection*. Drug Discov Today, 2004. **9**(10): p. 430-1.
8. Gribbon, P. and A. Sewing, *High-throughput drug discovery: what can we expect from HTS?* Drug Discov Today, 2005. **10**(1): p. 17-22.
9. Schuffenhauer, A., et al., *Library design for fragment based screening*. Curr Top Med Chem, 2005. **5**(8): p. 751-62.
10. Davies, T. and I. Tickle, *Fragment Screening Using X-Ray Crystallography*, in *Fragment-Based Drug Discovery and X-Ray Crystallography*, T.G. Davies and M. Hyvönen, Editors. 2012, Springer Berlin Heidelberg. p. 33-59.
11. Congreve, M., et al., *A 'rule of three' for fragment-based lead discovery?* Drug Discov Today, 2003. **8**(19): p. 876-7.
12. Lipinski, C.A., et al., *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings*. Adv Drug Deliv Rev, 2001. **46**(1-3): p. 3-26.
13. Clark, D.E. and S.D. Pickett, *Computational methods for the prediction of 'drug-likeness'*. Drug Discov Today, 2000. **5**(2): p. 49-58.
14. Hughes, J.D., et al., *Physicochemical drug properties associated with in vivo toxicological outcomes*. Bioorg Med Chem Lett, 2008. **18**(17): p. 4872-5.
15. Veber, D.F., et al., *Molecular properties that influence the oral bioavailability of drug candidates*. J Med Chem, 2002. **45**(12): p. 2615-23.
16. Gleeson, M.P., *Generation of a set of simple, interpretable ADMET rules of thumb*. J Med Chem, 2008. **51**(4): p. 817-34.
17. Leeson, P.D. and B. Springthorpe, *The influence of drug-like concepts on decision-making in medicinal chemistry*. Nat Rev Drug Discov, 2007. **6**(11): p. 881-90.
18. Teague, S.J., et al., *The Design of Leadlike Combinatorial Libraries*. Angew Chem Int Ed Engl, 1999. **38**(24): p. 3743-3748.
19. Vieth, M., et al., *Characteristic physical properties and structural fragments of marketed oral drugs*. J Med Chem, 2004. **47**(1): p. 224-32.
20. Wenlock, M.C., et al., *A comparison of physicochemical property profiles of development and marketed oral drugs*. J Med Chem, 2003. **46**(7): p. 1250-6.

21. Rees, D.C., et al., *Fragment-based lead discovery*. Nat Rev Drug Discov, 2004. **3**(8): p. 660-72.
22. Hajduk, P.J., *SAR by NMR: putting the pieces together*. Mol Interv, 2006. **6**(5): p. 266-72.
23. Pellecchia, M., et al., *Perspectives on NMR in drug discovery: a technique comes of age*. Nat Rev Drug Discov, 2008. **7**(9): p. 738-45.
24. Mayer, M. and B. Meyer, *Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy*. Angewandte Chemie International Edition, 1999. **38**(12): p. 1784-1788.
25. Liedberg, B., C. Nylander, and I. Lunström, *Surface plasmon resonance for gas detection and biosensing*. Sensors and Actuators, 1983. **4**(0): p. 299-304.
26. Perspicace, S., et al., *Fragment-based screening using surface plasmon resonance technology*. J Biomol Screen, 2009. **14**(4): p. 337-49.
27. Proll, F., P. Fechner, and G. Proll, *Direct optical detection in fragment-based screening*. Anal Bioanal Chem, 2009. **393**(6-7): p. 1557-62.
28. Navratilova, I. and A.L. Hopkins, *Fragment Screening by Surface Plasmon Resonance*. ACS Medicinal Chemistry Letters, 2010. **1**(1): p. 44-48.
29. Jhoti, H., et al., *Fragment-based screening using X-ray crystallography and NMR spectroscopy*. Curr Opin Chem Biol, 2007. **11**(5): p. 485-93.
30. Blaney, J., V. Nienaber, and S.K. Burley, *Fragment-based Lead Discovery and Optimization Using X-Ray Crystallography, Computational Chemistry, and High-throughput Organic Synthesis*, in *Fragment-based Approaches in Drug Discovery*. 2006, Wiley-VCH Verlag GmbH & Co. KGaA. p. 215-248.
31. Murray, C.W. and T.L. Blundell, *Structural biology in fragment-based drug design*. Curr Opin Struct Biol, 2010. **20**(4): p. 497-507.
32. Caliandro, R., et al., *Protein crystallography and fragment-based drug design*. Future Med Chem, 2013. **5**(10): p. 1121-40.
33. Kranz, J.K. and C. Schalk-Hihi, *Chapter eleven - Protein Thermal Shifts to Identify Low Molecular Weight Fragments*, in *Methods in Enzymology*, C.K. Lawrence, Editor. 2011, Academic Press. p. 277-298.
34. Holdgate, G.A., et al., *Affinity-based, biophysical methods to detect and analyze ligand binding to recombinant proteins: matching high information content with high throughput*. J Struct Biol, 2010. **172**(1): p. 142-57.
35. Vanwetswinkel, S., et al., *TINS, target immobilized NMR screening: an efficient and sensitive method for ligand discovery*. Chem Biol, 2005. **12**(2): p. 207-16.
36. Torres, F.E., et al., *Higher throughput calorimetry: opportunities, approaches and challenges*. Curr Opin Struct Biol, 2010. **20**(5): p. 598-605.
37. Freire, E., *Do enthalpy and entropy distinguish first in class from best in class?* Drug Discov Today, 2008. **13**(19-20): p. 869-74.
38. Wyatt, P.G., et al., *Identification of N-(4-piperidinyl)-4-(2,6-dichlorobenzoylamino)-1H-pyrazole-3-carboxamide (AT7519), a novel cyclin dependent kinase inhibitor using fragment-based X-ray crystallography and structure based drug design*. J Med Chem, 2008. **51**(16): p. 4986-99.

39. Howard, N., et al., *Application of fragment screening and fragment linking to the discovery of novel thrombin inhibitors*. J Med Chem, 2006. **49**(4): p. 1346-55.
40. Erlanson, D.A., *Fragment-based lead discovery: a chemical update*. Curr Opin Biotechnol, 2006. **17**(6): p. 643-52.
41. Norder, H., et al., *Picornavirus non-structural proteins as targets for new anti-virals with broad activity*. Antiviral Res, 2011. **89**(3): p. 204-18.
42. Tapparel, C., et al., *Picornavirus and enterovirus diversity with associated human diseases*. Infect Genet Evol, 2013. **14**: p. 282-93.
43. Gilbert, G.L., et al., *Outbreak of enterovirus 71 infection in Victoria, Australia, with a high incidence of neurologic involvement*. Pediatr Infect Dis J, 1988. **7**(7): p. 484-8.
44. Alexander, J.P., Jr., et al., *Enterovirus 71 infections and neurologic disease--United States, 1977-1991*. J Infect Dis, 1994. **169**(4): p. 905-8.
45. Samuda, G.M., et al., *Monoplegia caused by Enterovirus 71: an outbreak in Hong Kong*. Pediatr Infect Dis J, 1987. **6**(2): p. 206-8.
46. Qiu, J., *Enterovirus 71 infection: a new threat to global public health?* Lancet Neurol, 2008. **7**(10): p. 868-9.
47. Junttila, N., et al., *New enteroviruses, EV-93 and EV-94, associated with acute flaccid paralysis in the Democratic Republic of the Congo*. J Med Virol, 2007. **79**(4): p. 393-400.
48. Dougherty, W.G. and B.L. Semler, *Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes*. Microbiol Rev, 1993. **57**(4): p. 781-822.
49. Palmenberg, A.C., *Proteolytic processing of picornaviral polyprotein*. Annu Rev Microbiol, 1990. **44**: p. 603-23.
50. Kitamura, N., et al., *Primary structure, gene organization and polypeptide expression of poliovirus RNA*. Nature, 1981. **291**(5816): p. 547-53.
51. Flanagan, J.B., et al., *Covalent linkage of a protein to a defined nucleotide sequence at the 5'-terminus of virion and replicative intermediate RNAs of poliovirus*. Proc Natl Acad Sci U S A, 1977. **74**(3): p. 961-5.
52. Lee, Y.F., et al., *A protein covalently linked to poliovirus genome RNA*. Proc Natl Acad Sci U S A, 1977. **74**(1): p. 59-63.
53. Minor, P.D., *Polio eradication, cessation of vaccination and re-emergence of disease*. Nat Rev Microbiol, 2004. **2**(6): p. 473-82.
54. Clark, M.E., et al., *Poliovirus proteinase 3C converts an active form of transcription factor III<sub>C</sub> to an inactive form: a mechanism for inhibition of host cell polymerase III transcription by poliovirus*. EMBO J, 1991. **10**(10): p. 2941-7.
55. Clark, M.E., et al., *Direct cleavage of human TATA-binding protein by poliovirus protease 3C in vivo and in vitro*. Mol Cell Biol, 1993. **13**(2): p. 1232-7.
56. Porter, A.G., *Picornavirus nonstructural proteins: emerging roles in virus replication and inhibition of host cell functions*. J Virol, 1993. **67**(12): p. 6917-21.
57. Yalamanchili, P., K. Weidman, and A. Dasgupta, *Cleavage of transcriptional activator Oct-1 by poliovirus encoded protease 3C<sub>pro</sub>*. Virology, 1997. **239**(1): p. 176-85.
58. Weidman, M.K., et al., *Poliovirus 3C protease-mediated degradation of transcriptional activator p53 requires a cellular activity*. Virology, 2001. **291**(2): p. 260-71.

59. Weng, K.F., et al., *Enterovirus 71 3C protease cleaves a novel target CstF-64 and inhibits cellular polyadenylation*. PLoS Pathog, 2009. **5**(9): p. e1000593.
60. Walker, P.A., L.E. Leong, and A.G. Porter, *Sequence and structural determinants of the interaction between the 5'-noncoding region of picornavirus RNA and rhinovirus protease 3C*. J Biol Chem, 1995. **270**(24): p. 14510-6.
61. Ohlenschlager, O., et al., *The structure of the stemloop D subdomain of coxsackievirus B3 cloverleaf RNA and its interaction with the proteinase 3C*. Structure, 2004. **12**(2): p. 237-48.
62. Zell, R., et al., *Determinants of the recognition of enteroviral cloverleaf RNA by coxsackievirus B3 proteinase 3C*. RNA, 2002. **8**(2): p. 188-201.
63. Blaum, B.S., et al., *Functional binding of hexanucleotides to 3C protease of hepatitis A virus*. Nucleic Acids Res, 2012. **40**(7): p. 3042-55.
64. Mosimann, S.C., et al., *Refined X-ray crystallographic structure of the poliovirus 3C gene product*. J Mol Biol, 1997. **273**(5): p. 1032-47.
65. Wang, J., et al., *Crystal structures of enterovirus 71 3C protease complexed with rupintrivir reveal the roles of catalytically important residues*. J Virol, 2011. **85**(19): p. 10021-30.
66. Matthews, D.A., et al., *Structure-assisted design of mechanism-based irreversible inhibitors of human rhinovirus 3C protease with potent antiviral activity against multiple rhinovirus serotypes*. Proc Natl Acad Sci U S A, 1999. **96**(20): p. 11000-7.
67. Matthews, D.A., et al., *Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein*. Cell, 1994. **77**(5): p. 761-71.
68. Cordingley, M.G., et al., *Substrate requirements of human rhinovirus 3C protease for peptide cleavage in vitro*. J Biol Chem, 1990. **265**(16): p. 9062-5.
69. Binford, S.L., et al., *Conservation of amino acids in human rhinovirus 3C protease correlates with broad-spectrum antiviral activity of rupintrivir, a novel human rhinovirus 3C protease inhibitor*. Antimicrob Agents Chemother, 2005. **49**(2): p. 619-26.
70. Tong, L., *Viral proteases*. Chem Rev, 2002. **102**(12): p. 4609-26.
71. Rawlings, N.D., D.P. Tolle, and A.J. Barrett, *Evolutionary families of peptidase inhibitors*. Biochem J, 2004. **378**(Pt 3): p. 705-16.
72. Shepherd, T.A., et al., *Small peptidic aldehyde inhibitors of human rhinovirus 3C protease*. Bioorganic & Medicinal Chemistry Letters, 1996. **6**(23): p. 2893-2896.
73. Kaldor, S.W., et al., *Glutamine-derived aldehydes for the inhibition of human rhinovirus 3C protease*. Bioorganic & Medicinal Chemistry Letters, 1995. **5**(17): p. 2021-2026.
74. Webber, S.E., et al., *Tripeptide aldehyde inhibitors of human rhinovirus 3C protease: design, synthesis, biological evaluation, and cocrystal structure solution of P1 glutamine isosteric replacements*. J Med Chem, 1998. **41**(15): p. 2786-805.
75. De Palma, A.M., et al., *Selective inhibitors of picornavirus replication*. Med Res Rev, 2008. **28**(6): p. 823-84.
76. Sham, H.L., et al., *Potent inhibitors of the HIV-1 protease with good oral bioavailabilities*. Biochem Biophys Res Commun, 1995. **211**(1): p. 159-65.
77. Dragovich, P.S., et al., *Structure-based design of ketone-containing, tripeptidyl human rhinovirus 3C protease inhibitors*. Bioorg Med Chem Lett, 2000. **10**(1): p. 45-8.



78. Hanzlik, R.P. and S.A. Thompson, *Vinyllogous amino acid esters: a new class of inactivators for thiol proteases*. J Med Chem, 1984. **27**(6): p. 711-2.
79. Liu, S. and R.P. Hanzlik, *Structure-activity relationships for inhibition of papain by peptide Michael acceptors*. J Med Chem, 1992. **35**(6): p. 1067-75.
80. 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, 1998. **41**(15): p. 2806-18.
81. Dragovich, P.S., et al., *Structure-based design, synthesis, and biological evaluation of irreversible human rhinovirus 3C protease inhibitors. 4. Incorporation of P1 lactam moieties as L-glutamine replacements*. J Med Chem, 1999. **42**(7): p. 1213-24.
82. 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, 1999. **42**(7): p. 1203-12.
83. Witherell, G., *AG-7088 Pfizer*. Curr Opin Investig Drugs, 2000. **1**(3): p. 297-302.
84. Patick, A.K., et al., *In vitro antiviral activity of AG7088, a potent inhibitor of human rhinovirus 3C protease*. Antimicrob Agents Chemother, 1999. **43**(10): p. 2444-50.
85. Kaiser, L., C.E. Crump, and F.G. Hayden, *In vitro activity of pleconaril and AG7088 against selected serotypes and clinical isolates of human rhinoviruses*. Antiviral Res, 2000. **47**(3): p. 215-20.
86. Zhang, K.E., et al., *Liquid chromatography-mass spectrometry and liquid chromatography-NMR characterization of in vitro metabolites of a potent and irreversible peptidomimetic inhibitor of rhinovirus 3C protease*. Drug Metab Dispos, 2001. **29**(5): p. 729-34.
87. Hayden, F.G., et al., *Phase II, randomized, double-blind, placebo-controlled studies of rupintrivir nasal spray 2-percent suspension for prevention and treatment of experimentally induced rhinovirus colds in healthy volunteers*. Antimicrob Agents Chemother, 2003. **47**(12): p. 3907-16.
88. Hsyu, P.H., et al., *Pharmacokinetics and safety of an antirhinoviral agent, rupintrivir, in healthy volunteers*. Antimicrob Agents Chemother, 2002. **46**(2): p. 392-7.
89. 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 Chemother, 2005. **49**(6): p. 2267-75.
90. 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, 2003. **46**(21): p. 4572-85.
91. Lee, E.S., et al., *Development of potent inhibitors of the coxsackievirus 3C protease*. Biochem Biophys Res Commun, 2007. **358**(1): p. 7-11.
92. Lee, C.C., et al., *Structural basis of inhibition specificities of 3C and 3C-like proteases by zinc-coordinating and peptidomimetic compounds*. J Biol Chem, 2009. **284**(12): p. 7646-55.
93. Kim, B.K., et al., *Development of anti-coxsackievirus agents targeting 3C protease*. Bioorg Med Chem Lett, 2012. **22**(22): p. 6952-6.
94. Kuo, C.J., et al., *Design, synthesis, and evaluation of 3C protease inhibitors as anti-enterovirus 71 agents*. Bioorg Med Chem, 2008. **16**(15): p. 7388-98.

95. Tan, J., et al., *3C protease of enterovirus 68: structure-based design of Michael acceptor inhibitors and their broad-spectrum antiviral effects against picornaviruses*. J Virol, 2013. **87**(8): p. 4339-51.
96. Venkatraman, S., et al., *Design, synthesis, and evaluation of azapeptides as substrates and inhibitors for human rhinovirus 3C protease*. Bioorg Med Chem Lett, 1999. **9**(4): p. 577-80.
97. Hill, R.D. and J.C. Vederas, *Azodicarboxamides: A New Class of Cysteine Proteinase Inhibitor for Hepatitis A Virus and Human Rhinovirus 3C Enzymes*. The Journal of Organic Chemistry, 1999. **64**(26): p. 9538-9546.
98. Murray, M.A., et al., *Peptidyl diazomethyl ketones inhibit the human rhinovirus 3C protease: effect on virus yield by partial block of P3 polyprotein processing*. Antivir Chem Chemother, 2001. **12**(5): p. 273-81.
99. Xian, M., et al., *S-nitrosothiols as novel, reversible inhibitors of human rhinovirus 3C protease*. Bioorg Med Chem Lett, 2000. **10**(18): p. 2097-100.
100. Chen, S.H., et al., *Synthesis and evaluation of tripeptidyl alpha-ketoamides as human rhinovirus 3C protease inhibitors*. Bioorg Med Chem Lett, 2003. **13**(20): p. 3531-6.
101. Reich, S.H., et al., *Substituted benzamide inhibitors of human rhinovirus 3C protease: structure-based design, synthesis, and biological evaluation*. J Med Chem, 2000. **43**(9): p. 1670-83.
102. Maugeri, C., et al., *New anti-viral drugs for the treatment of the common cold*. Bioorg Med Chem, 2008. **16**(6): p. 3091-107.
103. Webber, S.E., et al., *Design, synthesis, and evaluation of nonpeptidic inhibitors of human rhinovirus 3C protease*. J Med Chem, 1996. **39**(26): p. 5072-82.
104. Jungheim, L.N., et al., *Inhibition of human rhinovirus 3C protease by homophthalimides*. Bioorganic & Medicinal Chemistry Letters, 1997. **7**(12): p. 1589-1594.
105. Wang, Q.M., et al., *Dual inhibition of human rhinovirus 2A and 3C proteases by homophthalimides*. Antimicrob Agents Chemother, 1998. **42**(4): p. 916-20.
106. Im, I., et al., *Structure-activity relationships of heteroaromatic esters as human rhinovirus 3C protease inhibitors*. Bioorg Med Chem Lett, 2009. **19**(13): p. 3632-6.
107. Kuo, C.J., et al., *Individual and common inhibitors of coronavirus and picornavirus main proteases*. FEBS Lett, 2009. **583**(3): p. 549-55.
108. Ramtohul, Y.K., et al., *Pseudoxazolones, a new class of inhibitors for cysteine proteinases: inhibition of hepatitis A virus and human rhinovirus 3C proteinases*. Chemical Communications, 2001(24): p. 2740-2741.
109. Ramtohul, Y.K., et al., *Synthesis of pseudoxazolones and their inhibition of the 3C cysteine proteinases from hepatitis A virus and human rhinovirus-14*. Journal of the Chemical Society, Perkin Transactions 1, 2002(11): p. 1351-1359.
110. 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, 2002. **45**(10): p. 2016-23.
111. Singh, S.B., et al., *Structure of stereochemistry of thysanone: a novel human rhinovirus 3C-protease inhibitor from Thysanophora penicilloides*. Tetrahedron Letters, 1991. **32**(39): p. 5279-5282.

112. Kadam, S., et al., *Citrinin hydrate and radicinin: human rhinovirus 3C-protease inhibitors discovered in a target-directed microbial screen*. J Antibiot (Tokyo), 1994. **47**(7): p. 836-9.
113. Brill, G.M., et al., *Novel triterpene sulfates from Fusarium compactum using a rhinovirus 3C protease inhibitor screen*. J Antibiot (Tokyo), 1996. **49**(6): p. 541-6.
114. Baxter, A., et al., *Non-covalent inhibitors of rhinovirus 3C protease*. Bioorg Med Chem Lett, 2011. **21**(2): p. 777-80.
115. Studier, F.W., *Protein production by auto-induction in high density shaking cultures*. Protein Expr Purif, 2005. **41**(1): p. 207-34.
116. Sambrook, J. and D.W. Russell, *Preparation and Transformation of Competent E. coli Using Calcium Chloride*. CSH Protoc, 2006. **2006**(1).
117. Kabsch, W., *Xds*. Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 125-32.
118. Evans, P., *Scaling and assessment of data quality*. Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 1): p. 72-82.
119. Potterton, E., et al., *A graphical user interface to the CCP4 program suite*. Acta Crystallogr D Biol Crystallogr, 2003. **59**(Pt 7): p. 1131-7.
120. McCoy, A.J., et al., *Phaser crystallographic software*. J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.
121. Murshudov, G.N., A.A. Vagin, and E.J. Dodson, *Refinement of macromolecular structures by the maximum-likelihood method*. Acta Crystallogr D Biol Crystallogr, 1997. **53**(Pt 3): p. 240-55.
122. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics*. Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.
123. van der Linden, L., et al., *Application of a cell-based protease assay for testing inhibitors of picornavirus 3C proteases*. Antiviral Res, 2014. **103**: p. 17-24.
124. Piotto, M., V. Saudek, and V. Sklenar, *Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions*. J Biomol NMR, 1992. **2**(6): p. 661-5.
125. Bjorndahl, T.C., et al., *NMR solution structures of the apo and peptide-inhibited human rhinovirus 3C protease (Serotype 14): structural and dynamic comparison*. Biochemistry, 2007. **46**(45): p. 12945-58.
126. Costenaro, L., et al., *Structural basis for antiviral inhibition of the main protease, 3C, from human enterovirus 93*. J Virol, 2011. **85**(20): p. 10764-73.
127. Patick, A.K., *Rhinovirus chemotherapy*. Antiviral Res, 2006. **71**(2-3): p. 391-6.
128. Kathman, S.G., Z. Xu, and A.V. Statsyuk, *A fragment-based method to discover irreversible covalent inhibitors of cysteine proteases*. J Med Chem, 2014. **57**(11): p. 4969-74.
129. Lu, G., 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, 2011. **85**(19): p. 10319-31.
130. Kim, Y., et al., *Broad-spectrum antivirals against 3C or 3C-like proteases of picornaviruses, noroviruses, and coronaviruses*. J Virol, 2012. **86**(21): p. 11754-62.
131. Klein, J., et al., *Detecting Binding Affinity to Immobilized Receptor Proteins in Compound Libraries by HR-MAS STD NMR*. Journal of the American Chemical Society, 1999. **121**(22): p. 5336-5337.



132. Meyer, B., et al., *Saturation transfer difference NMR spectroscopy for identifying ligand epitopes and binding specificities*. Ernst Schering Res Found Workshop, 2004(44): p. 149-67.
133. Lepre, C.A., J.M. Moore, and J.W. Peng, *Theory and applications of NMR-based screening in pharmaceutical research*. Chem Rev, 2004. **104**(8): p. 3641-76.
134. Campos-Olivas, R., *NMR screening and hit validation in fragment based drug discovery*. Curr Top Med Chem, 2011. **11**(1): p. 43-67.
135. Hajduk, P.J., J.R. Huth, and S.W. Fesik, *Druggability indices for protein targets derived from NMR-based screening data*. J Med Chem, 2005. **48**(7): p. 2518-25.
136. van Dongen, M., et al., *Structure-based screening and design in drug discovery*. Drug Discovery Today, 2002. **7**(8): p. 471-478.
137. Rademacher, C., et al., *Targeting norovirus infection-multivalent entry inhibitor design based on NMR experiments*. Chemistry, 2011. **17**(27): p. 7442-53.
138. Pantoliano, M.W., et al., *High-density miniaturized thermal shift assays as a general strategy for drug discovery*. J Biomol Screen, 2001. **6**(6): p. 429-40.
139. Cummings, M.D., M.A. Farnum, and M.I. Nelen, *Universal screening methods and applications of ThermoFluor*. J Biomol Screen, 2006. **11**(7): p. 854-63.
140. Vogt, G. and P. Argos, *Protein thermal stability: hydrogen bonds or internal packing?* Folding and Design, 1997. **2**, **Supplement 1**(0): p. S40-S46.
141. Hubbard, R.E. and J.B. Murray, *Experiences in fragment-based lead discovery*. Methods Enzymol, 2011. **493**: p. 509-31.
142. Basse, N., et al., *Toward the rational design of p53-stabilizing drugs: probing the surface of the oncogenic Y220C mutant*. Chem Biol, 2010. **17**(1): p. 46-56.
143. Jonsson, U., et al., *Real-time biospecific interaction analysis using surface plasmon resonance and a sensor chip technology*. Biotechniques, 1991. **11**(5): p. 620-7.
144. Huber, W. and F. Mueller, *Biomolecular interaction analysis in drug discovery using surface plasmon resonance technology*. Curr Pharm Des, 2006. **12**(31): p. 3999-4021.
145. Cui, S., et al., *Crystal structure of human enterovirus 71 3C protease*. J Mol Biol, 2011. **408**(3): p. 449-61.
146. Wang, Q.M. and R.B. Johnson, *Activation of human rhinovirus-14 3C protease*. Virology, 2001. **280**(1): p. 80-6.
147. Hajduk, P.J., J.R. Huth, and C. Tse, *Predicting protein druggability*. Drug Discov Today, 2005. **10**(23-24): p. 1675-82.