Mechanistic Elucidation of Protease–Substrate and Protein–Protein Interactions for Targeting Viral Infections

Inaugural-Dissertation

to obtain the academic degree

Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry, Pharmacy

of Freie Universität Berlin

by

Szymon Piotr Pach

2021

The presented thesis was prepared from 2017 until 2021 under the supervision of Prof. Dr. Gerhard Wolber at the Institute of Pharmacy of the Freie Universität Berlin.

1st Reviewer: Prof. Dr. Gerhard Wolber

2nd Reviewer: Prof. Dr. Christine Goffinet

Date of defense   15.02.2022

MÖBIUS   Was einmal gedacht wurde, kann nicht mehr
            zurückgenommen werden.

            (What was once thought can never be unthought.)

                              Friedrich Dürrenmatt, *Die Physiker*

Acknowledgments

Table of Contents

1. Motivation and Aim of the Thesis

A Zika virus epidemic broke out in 2015 in Brazil[1] causing microcephaly, a rare fetal malformation, in over 600 pregnant women.[2] Due to this severe and rare complication of a Zika infection, the epidemic raised global interest. A concern emerged with the start of the 2016 Summer Olympics in Rio de Janeiro: the virus could spread all over the world together with tourists coming back to their homelands.[3] The broad media awareness triggered a surge in research on the Zika virus (firstly described in 1952,[4] sic) and also other flaviviruses as shown in Figure 1B.
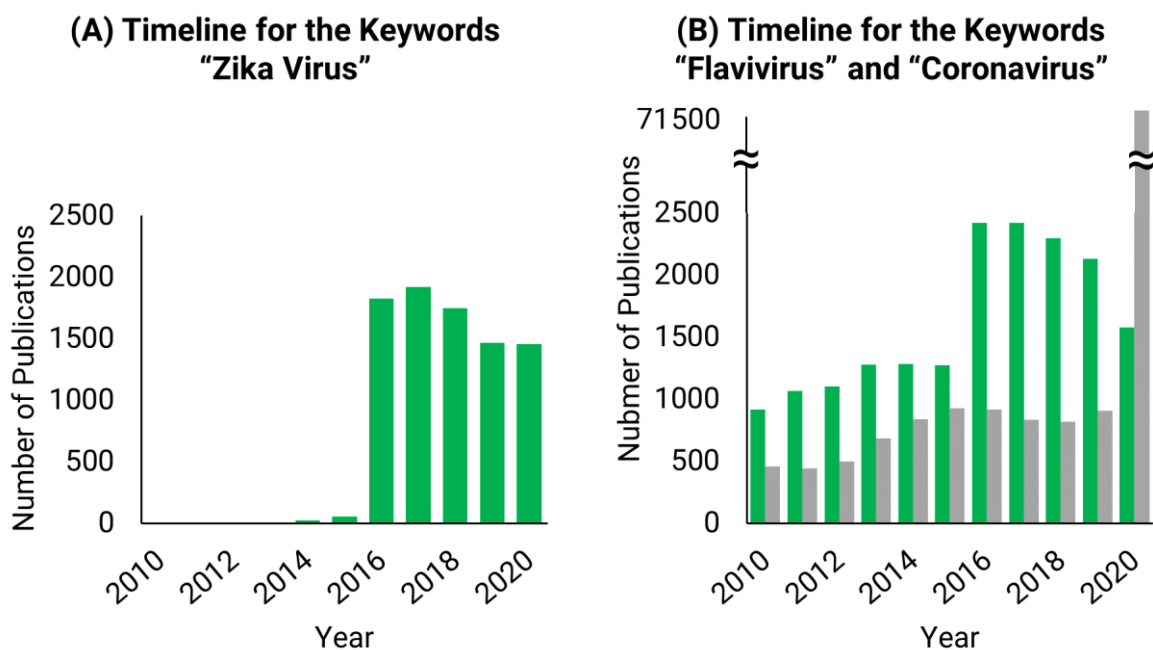
Figure 1. A timeline presenting the surge in research on (A) the Zika virus (green) and (B) flaviviruses (green) after 2015 and coronaviruses (grey) after 2019. The data were obtained from PubMed by searching for keywords "Zika virus", "flavivirus", and "coronavirus" in the period from 2010 to 2020.

A similar timeline development can be observed for members of the genus *Coronavirus* after 2019 (Figure 1B). In short, the research on viruses and antiviral substances undergoes a leapfrog development from one epidemic to another. Keeping this in mind, we decided to establish computational models and tools, which will allow efficient research on viral proteins independently from epidemiological trends.

Here, the target choice for our *in-silico* models will be dependent on the availability of structural data. To validate our *in-silico* models in an experimental setup, we have

established collaborations with four laboratories specialized in biochemical and biological assays for the experimental characterization of antiviral compounds and therapeutic approaches. According to the expertise of the collaborating labs, we focus on (i) emerging viral species, such as the severe acute respiratory syndrome-related coronavirus 2, Dengue virus, and West Nile virus, and (ii) neglected viral species lacking broad scientific interest, such as the Zika virus and equine herpesvirus 1 and 4.

The primary objectives of this thesis are establishing computational workflows for the identification of:

1) novel small-molecule antiviral substances inhibiting the flaviviral NS2B-NS3 proteases,

2) a binding epitope of glycoprotein D to its receptor, the major histocompatibility complex I, and

3) protein-protein interaction descriptors enabling prediction of species susceptibility to coronavirus disease 2019 (COVID-19).

The biochemical and biological evaluation of our predictions is performed in close collaborations with the laboratories of:

1) Dr. Christoph Nitsche, Prof. Jörg Rademann, and Dr. Mila Leuthold for the development of NS2B-NS3 protease inhibitors and

2) Prof. Nikolaus Osterrieder for the prediction of binding epitopes for herpesviral glycoprotein D and development of molecular descriptors of animal susceptibility to COVID-19.

2.   Theoretical Background

2.1   Epidemiology and Replication of Viruses

2.1.1   Viruses Represent a Global Health Threat

Viral infections represent a global health threat, causing thousands of deaths and disabilities each year.[5] Viral infections of plants and animals cause significant economic loss and damage to the environment.[6] Viruses (from lat. *virus*- poison, noxious liquid) represent a long-known threat. Several pandemics of viral infections have been reported historically: smallpox, influenza, yellow fever, measles, Dengue fever, and more recently, acquired immunodeficiency syndrome (AIDS), Ebola fever,[7] and coronavirus disease 2019 (COVID-19).[8] Despite the long-known history of viral infections, viruses were first discovered in the late 19th century.[9, 10]

New viral diseases are emerging due to close human-animal contacts, the spread of viral vectors enhanced by global climate change, and international tourism,[11] rendering the need for antiviral therapies an urgent challenge for mankind.
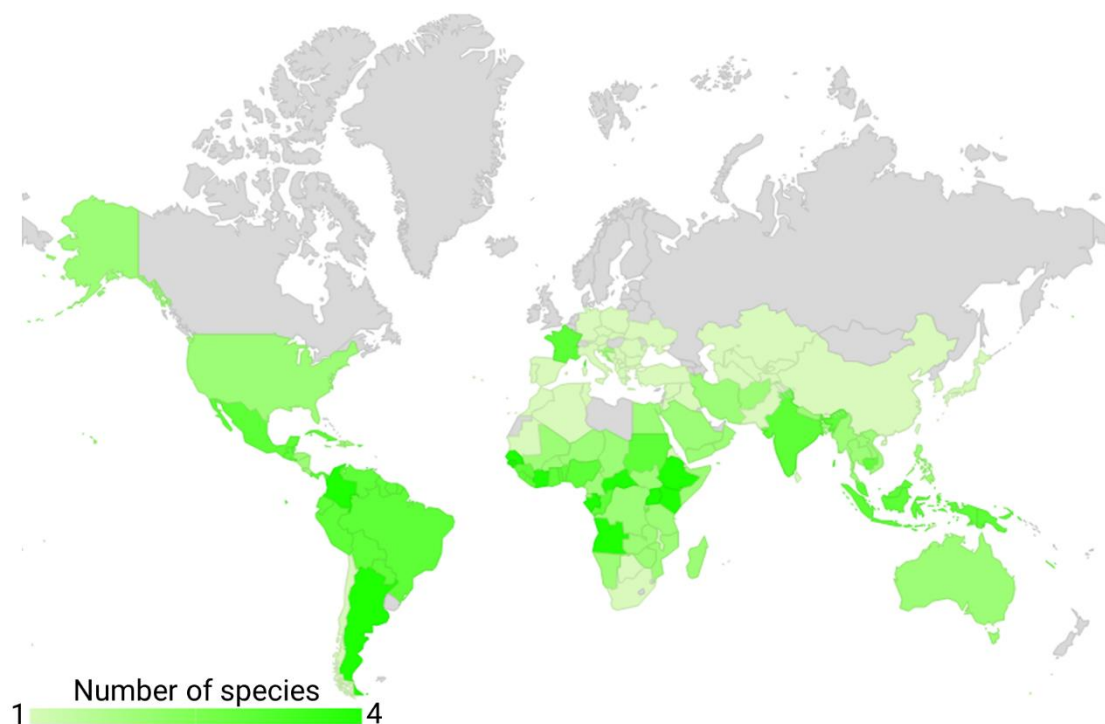
Figure 2. Spread of Dengue, yellow fever, Zika, and West Nile viruses around the world indicated as the number of the flaviviral species present in each country (adapted after Collins et al.[12] and Center for Disease Control and Prevention https://www.cdc.gov/).

Dengue, West Nile, Yellow Fever, and Zika Virus

Due to the spread of tropical mosquito species into the northern hemisphere, *ar*thropod-*bo*rne viruses (arboviruses) represent an increasing threat to all humans. The most prominent arboviruses are flaviviruses, which have caused several epidemics: West Nile virus (WNV, endemic distribution in Africa, North and South America, and South Europe[12]), Zika virus (ZIKV, an epidemic in Oceania and South America[1]), Dengue virus (DENV, endemic spread in Africa, South America, and South Asia[12]), and yellow fever virus (YFV, several epidemics in America and Africa[13], Figure 2).

The mortality of flaviviral infections is moderate to high at around 5% for secondary Dengue fever infection,[14] 8% for neonatal Zika fever,[15] 10% for West Nile encephalitis,[14] and up to 50% for severe yellow fever.[14] The fatal complication of Dengue fever and yellow fever is a hemorrhagic syndrome (Figure 3).[14] West Nile virus leads to encephalitis resulting in a high mortality rate[14] and Zika fever leads to microcephaly in newborns, which can be fatal.[15]

Currently, no specific treatment against arboviral infections is available. In the cases of yellow fever,[16] Japanese encephalitis,[17] tick-borne encephalitis,[18] and Dengue fever,[19] protective vaccines are used. There is still no vaccine against the Zika virus and West Nile virus. Paradoxically, the Dengue virus vaccine increases the risk of severe Dengue infection in some populations due to antibody-dependent enhancement.[20] Therefore, there is an urgent need for the development of small molecular antiarboviral substances.

Severe Acute Respiratory Syndrome-Related Coronavirus 2

The most recent outbreak of a viral infection is the COVID-19 pandemic, which originated in December 2019 in the Chinese city of Wuhan.[8] The causative agent is a new coronavirus species, severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2). Until mid-2021, the COVID-19 pandemics took a toll of over 1 million fatalities registered in all countries worldwide.[5] The fatality rate of the new disease is variable (0-15%[21]) and is significantly higher in older age, in the presence of comorbidities,[22] and in male patients.[23] COVID-19 causes damage to multiple organs, mainly the lungs and the digestive and cardiovascular systems (Figure 3).[22]

The ongoing COVID-19 vaccination campaign raises hope in stopping the viral spread. However, due to potential vaccine escape mutants,[24] more research is needed. Here, novel animal infection models might support the understanding of COVID-19 and the development of novel vaccines.[25]

Equine Herpesvirus 1 and 4

Viral epidemics are also harmful to animals. Most recently, an outbreak of a neurological form of equine herpesvirus 1 (EHV-1) infection during jumping tours in Valencia (Spain) caused several fatalities.[26] EHV-1 causes damage to the nervous system and urogenital and respiratory tracts (Figure 3). It has been also discussed that equid horse-borne herpesviruses represent a threat to the environment by causing fatalities in black bears.[27] Zebra-borne EHV-1 can cause infections of alpaca or Indian rhinceros.[27]

Currently, only limited treatment options transferred from human medicine are available.[6] To prevent an EHV infection, attenuated virus vaccines are used.[6] Their efficacy, however, is strongly limited.[6] Hence, an understanding of immune system evasion by EHV is needed for the development of better vaccines.
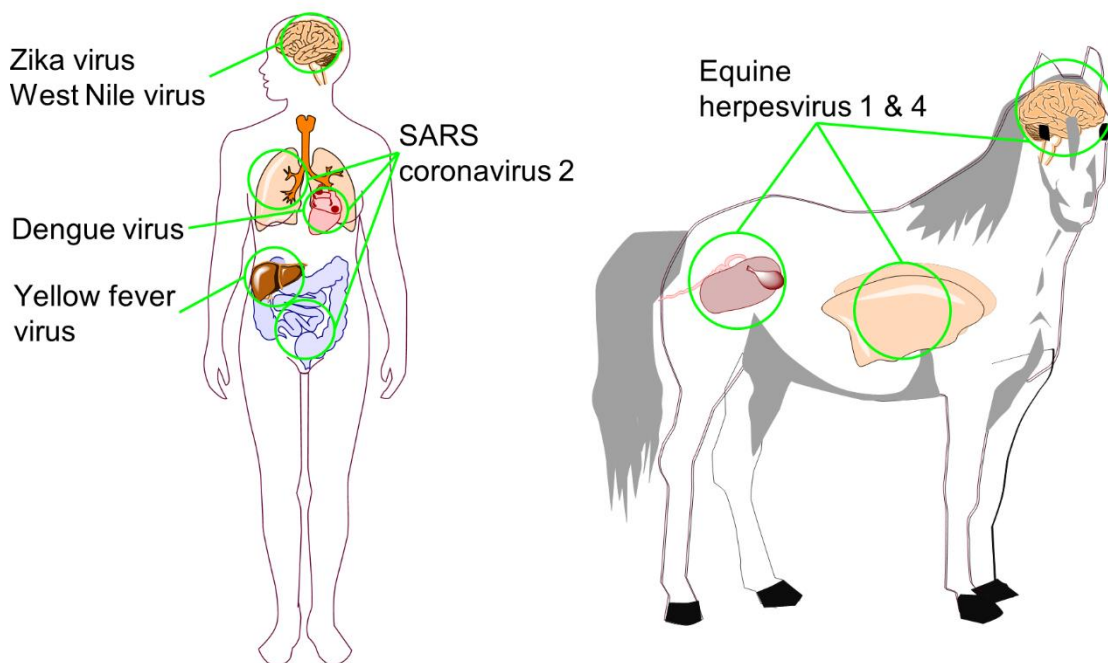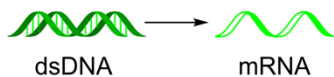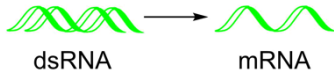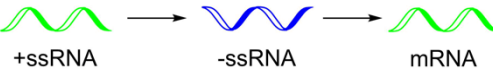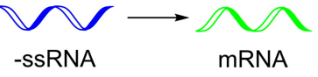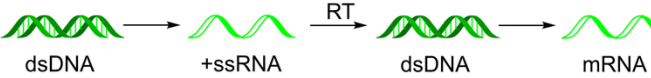
Figure 3. Symbolic representation of the organs and systems infected by the flaviviruses, coronaviruses, and equine herpesviruses.

2.1.2 The Viral Replication Cycle

Viruses are submicroscopic particles containing nucleic acids embedded in a protein-based capsid.[28] They are obligate intracellular parasites and can only replicate when they infect a living cell.[29] Hence, viruses are not classified as living organisms in the broad scientific consensus.[29]

Their categorization as non-animated matter combined with heterogeneous and unclear evolutionary origin make the taxonomy of viruses a complex task.[29, 30] Therefore, a systematic grouping of viruses after common ancestors, as commonly practiced in the taxonomy of animals, plants, or bacteria, is pointless.[29] The International Committee on Taxonomy of Viruses (ICTV) classifies viruses according to the type of contained nucleic acid and the similarity of their sequence, capsid structure, and replication cycle.[31] The underlying concept of viral taxonomy[30] is summarized within the Baltimore classification. David Baltimore, the inventor of this scheme,[32] classified viruses according to their genetic systems, i.e. the type of nucleic acid and its transcription[30]. Currently, seven groups are known (I-VII, Table 1).

Table 1. Baltimore classification of viruses depending on the type, polarity, and transcription of the nucleic acid.

Class	Genetic Material	Simplified Transcription Scheme	Family Examples
I	dsDNA	dsDNA → mRNA	*Herpesviridae**
II	+ssDNA	+ssDNA → dsDNA → mRNA	*Parvoviridae*
III	dsRNA	dsRNA → mRNA	*Reoviridae*
IV	+ssRNA	+ssRNA → -ssRNA → mRNA	*Flaviviridae** *Coronaviridae**
V	-ssRNA	-ssRNA → mRNA	*Filoviridae*
VI	+ssRNA	+ssRNA →(RT) dsDNA → mRNA	*Retroviridae*
VII	dsDNA	dsDNA → +ssRNA →(RT) dsDNA → mRNA	*Hepadnaviridae*

Viral families marked with an asterisk (*) are the subject of the following thesis. Abbreviations: ds- double-stranded nucleic acid, RT- reverse transcriptase, ss- single-stranded nucleic acid.

The design of directly acting antiviral drugs (DAAD) requires detailed knowledge about the replication cycle of the virus. Therefore, in the following section, the replication cycles of viruses subject to this thesis are sketched. In this thesis, only Baltimore groups I and IV will be described, as representatives of these groups were targeted.

Replication of Baltimore Group I Viruses

The replication cycle of the genus *Herpesvirus* starts with an entry step (Figure 4, I.).
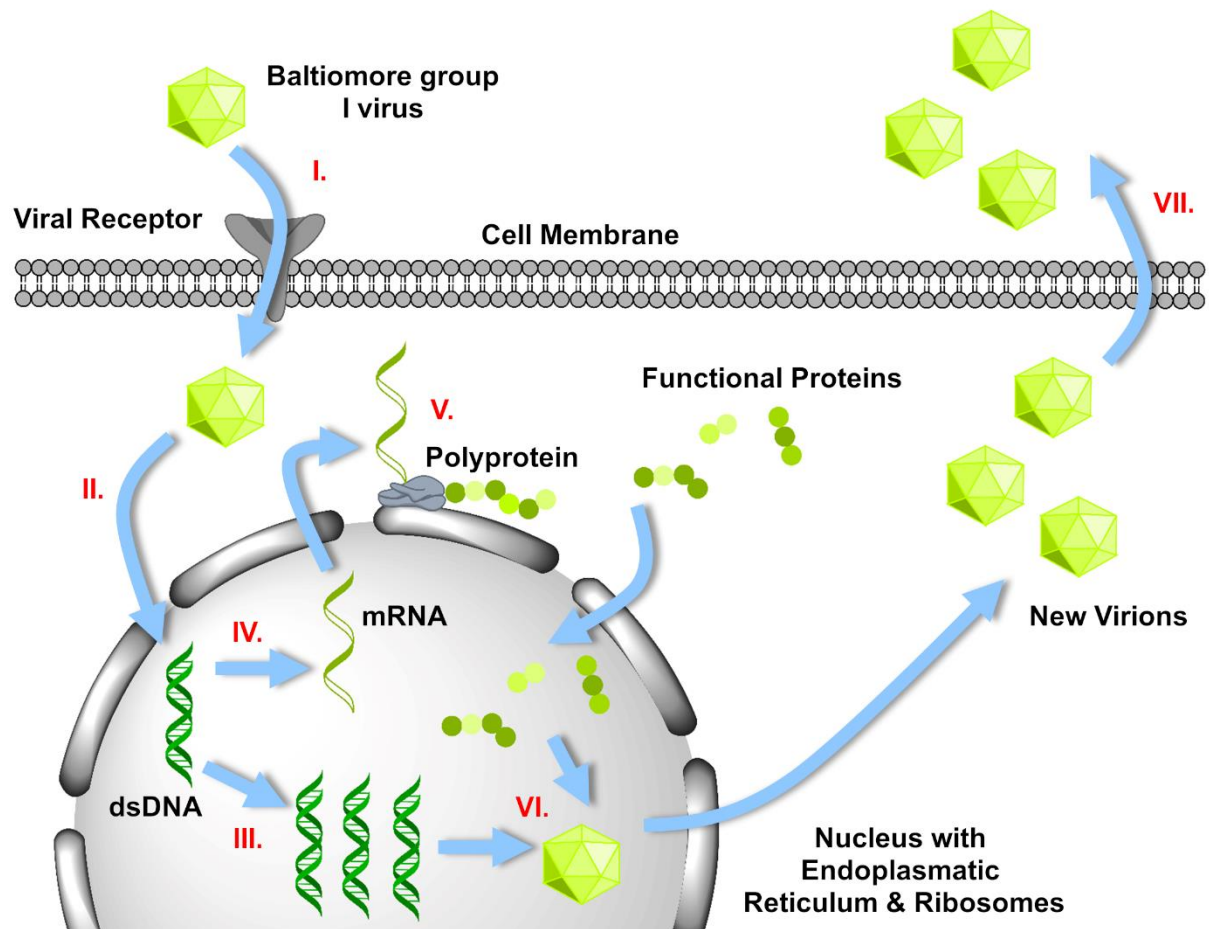
Figure 4. A schematic representation of a replication cycle of Baltimore group I viruses (except Poxviruses). Red roman numerals indicate replication steps: I. viral entry, II. uncoating, III. DNA replication, IV. transcription, V. translation, VI. assembly, and VII. release.

Viral fusion proteins specifically interact with a viral receptor on the host cell surface. Equine herpesviruses 1 and 4 enter the host cells using glycoprotein D as a fusion protein.[33, 34] The major histocompatibility complex I serves as the viral receptor.[33, 34] Next, the genetic material is transported to the nucleus where it is replicated by viral enzymes (*pol* catalytic unit in equine herpesviruses 1 and 4[35]) and transcribed into mRNA (Figure 4, steps II., III., and IV.). The mRNA is then translated (Figure 4, V.) into

early and late proteins, which are subsequently processed by proteases, such as the herpesviral protease.[36] The proteins are then transported back to the nucleus and are assembled with the replicated dsDNA to obtain new virions (Figure 4, VI.). Finally, the new virions leave the nucleus and exit the cell by exocytosis (Figure 4, VII.).[37]

Replication of Baltimore Group IV Viruses.
The replication cycles of the viral genera *Flavivirus* and *Coronavirus* are similar to each other (Figure 5).[31, 38] In the beginning, the virus enters the cell (Figure 5, I.) using specific interactions between the viral fusion protein and the receptor (e.g. protein E - Gas6-AXL tyrosine kinase receptor complex for Zika virus[39] and spike protein - angiotensin-converting enzyme 2 for SARS-CoV-2[8]).
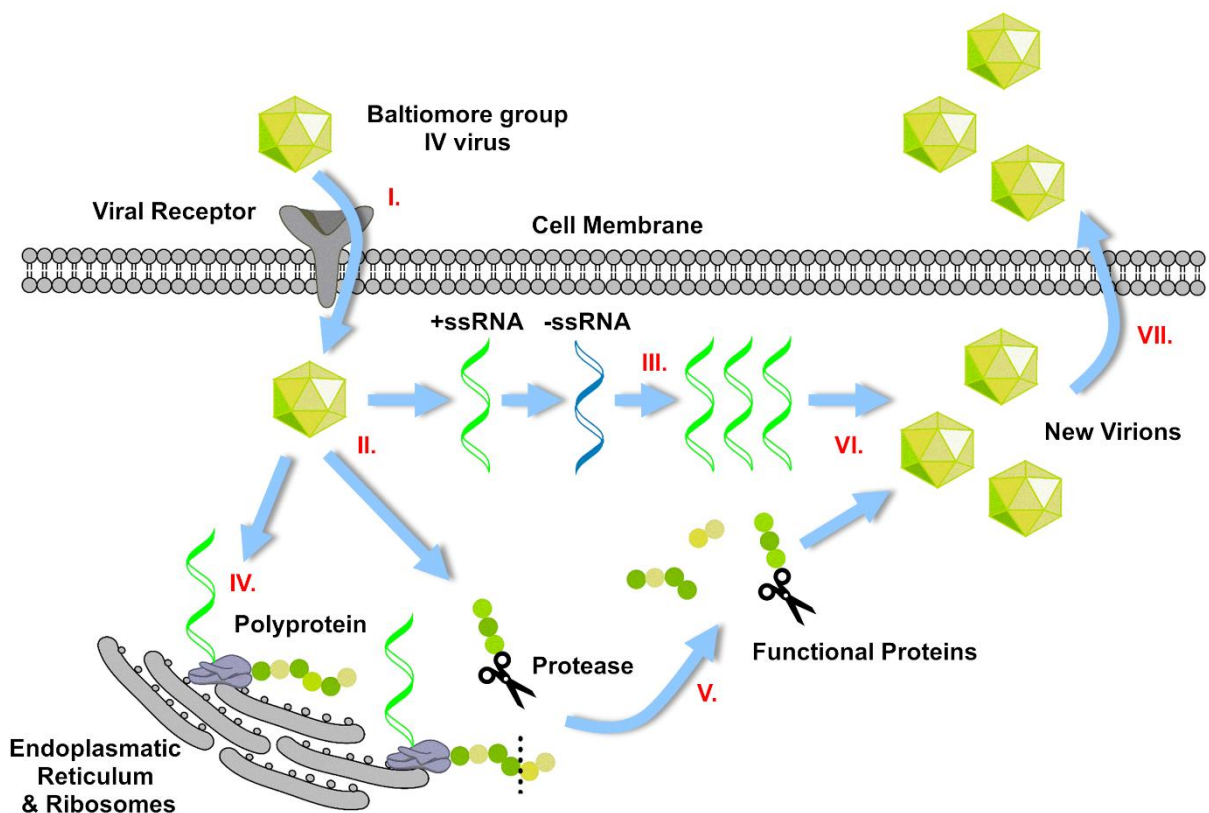
Figure 5. A schematic representation of a replication cycle of Baltimore group IV viruses. Red roman numerals indicate replication steps: I. Viral entry, II. uncoating, III. RNA replication, IV. translation, V. processing, VI. assembly, and VII. release.

Once it intrudes the cell, the virus releases its genetic material (+ssRNA) and non-structural proteins (Figure 5, II. Uncoating). The +ssRNA serves as a template for replication of the viral genome via an -ssRNA intermediate (Figure 5, III.) and
18

translation (Figure 5, IV.).[38] Due to the lack of RNA-dependent RNA polymerase in the human cells, this enzyme is provided by the virus (e.g. NS5 in Zika virus[38]). The translation is performed by cellular components to yield a polyprotein: a string of consecutively synthesized structural and non-structural protein units. The polyprotein is then autocatalytically processed by the viral protease (NS2B-NS3 in Zika virus[38] and M^{pro} in SARS-CoV-2[40]) and host proteases like furin[41, 42] to generate functional proteins (Figure 5, V.). Finally, the replicated +ssRNA and the structural and non-structural proteins are assembled into new virions that are subsequently released from the cell (Figure 5, VI. and VII.).

2.2 Antiviral Substances

2.2.1 History of Antiviral Research

As shown above, viruses can infect humans and other animals to cause life-threatening diseases. Accordingly, there is an urgent unmet need for antiviral substances. The two most promising strategies in the development of antivirals are (i) targeting proteins (whether of viral or host origin) involved in the viral replication cycle or (ii) stimulation of the host's immune response.[43] The first approach will be followed in this thesis as it allows specific targeting of the pathogen, resulting in fewer toxic effects compared to the second approach that bears the danger of insufficient selectivity.[43]

Research on antiviral substances started in the early 1950s with the accidental discovery of the antiviral activity of p-aminobenzaldehydethiosemicarbazone on vaccinia virus-infected eggs and mice (Figure 6).[44] For the first time, it was shown that an antiviral substance is not necessarily cytotoxic. The next surge of antiviral research was caused in the early 1960s by the accidental observation of the antiviral activity of idoxuridine, a derivative of a cytostatic drug.[45] The same team presented a novel cell-based plaque-reduction testing system for potential antiviral substances, similar to that used in antibiotic development.[46] Both discoveries triggered antiviral research on nucleotide analogs leading to a potent antiherpesviral drug, aciclovir.[47]

In parallel, it has been shown that ammonium ions[48] and primary amines[49] inhibit influenza virus replication leading to the discovery of amantadine,[50] an uncoating inhibitor. In the early 1970s, it was discovered that muramic acid derivatives can specifically inhibit the influenza virus neuraminidase, leading to a potent release

inhibitor.[51] As the crystal structures of the enzyme were made available in the mid-1980s, the first rationally designed neuraminidase inhibitor with antiviral *in-vivo* activity, zanamivir, was discovered.[52] For the first time, a structure-based approach was used to successfully identify an antiviral drug.
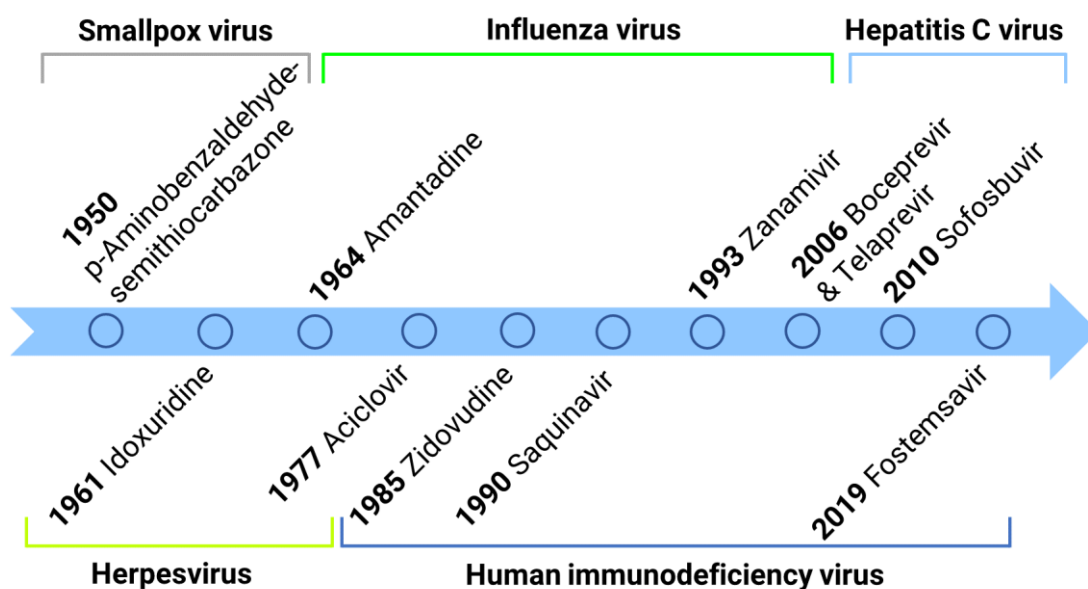
Figure 6. A representative timeline of antiviral research. Each time point represents a breakthrough innovation or the discovery of a first-in-class drug for the therapy of viral infections.

In 1980, a mysterious infectious disease was causing the accumulation of patients with a rare cancer type, Kaposi sarcoma, and opportunistic infections. In 1983, the pathogen was identified[53] and later named human immunodeficiency virus, HIV. Only two years later, in 1985, the first drug against the new virus, zidovudine, was discovered in a high-throughput screening campaign.[54] In 1989, the first structures of HIV enzymes were solved, triggering structure-based development of antiretroviral substances. The first rationally discovered HIV protease inhibitor, saquinavir, was introduced in 1990,[55] only seven years after the identification of the virus. The development of drugs against HIV is still ongoing. The newest antiretroviral drug introduced to the market in 2019 is fostemsavir,[56] an entry inhibitor.

The most recent antiviral drug class was developed against a virus firstly described in 1989, hepatitis C virus (HCV).[57] The treatment of hepatitis C was revolutionized with the introduction of the first protease inhibitors, boceprevir and telaprevir, in 2006.[58, 59]

Notably, these drugs were also discovered using structure-based approaches. The last breakthrough in the treatment of hepatitis C was the discovery of sofosbuvir, an NS5B RNA-polymerase inhibitor, in 2010, rendering combination therapy possible.[60]

2.2.2 Experimental Structure Determination

Structure determination methods play a crucial role in the current drug development process. Despite recent innovations in the field of *ab initio* protein structure prediction,[61] experimentally obtained coordinates of macromolecules are still essential for structure-based drug design (SBDD).
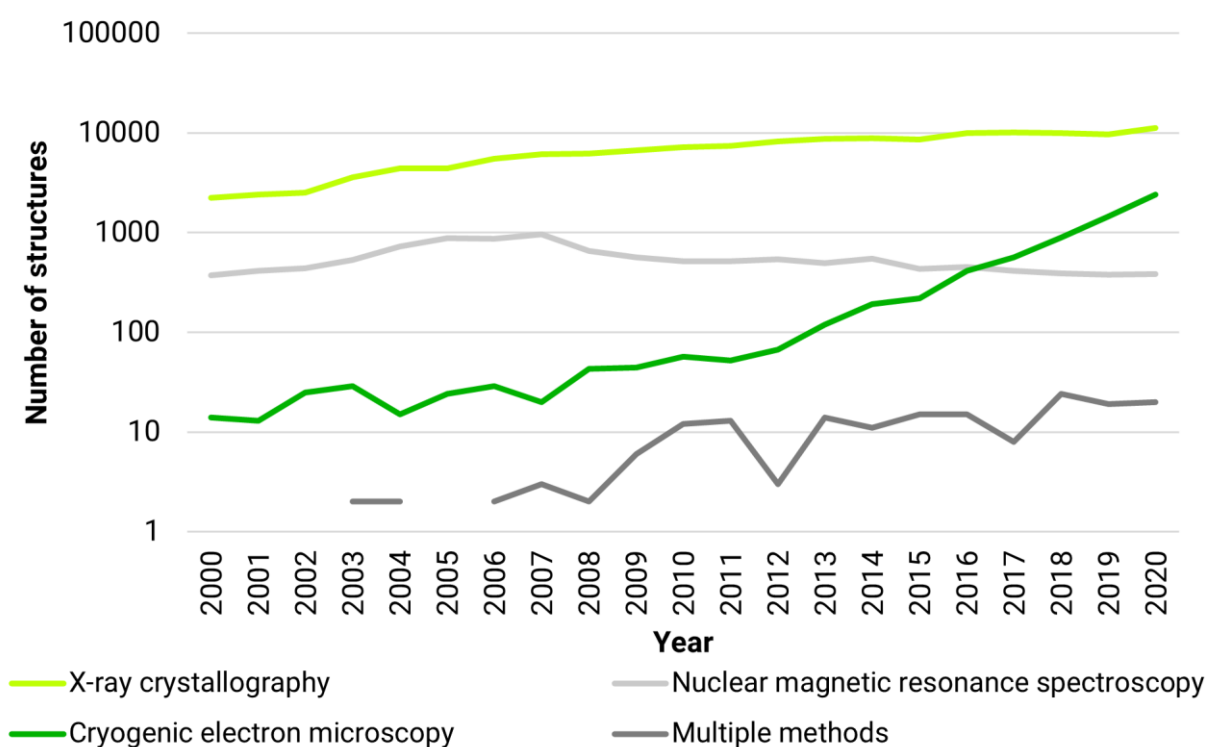
Figure 7. Structural data deposited in the Protein Data Bank (PDB)[62] in the period 2000-2020 after its launch (adapted from https://www.rcsb.org/stats/all-released-structures). Color code: light green- X-ray crystallography, light grey- nuclear magnetic resonance spectroscopy, dark green- cryogenic electron microscopy, dark grey- multiple methods.

The structural data available for the research community is mainly stored in two non-commercial databases: Protein Data Bank (PDB)[62] and Electron Microscopy Data Bank (EMDB).[63] The structural data deposited in these databases is obtained by researchers mostly using X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, or cryogenic electron microscopy (Cryo-EM, Figure 7). The most popular method

applied for structure determination is X-ray crystallography (Figure 7). Briefly, to obtain the three-dimensional data, the pure sample of a biomolecule is treated with chemicals such as salts and buffers to gain a protein crystal. This crystal is then irradiated with X-rays, delivering two-dimensional diffraction patterns depending on the wavelength, the distance between single crystal layers, angle of irradiation, and electronic and structural properties of the crystallized particles. For highly complex molecules like proteins, software (PHASER,[64] PHENIX,[65] or COOT[66]) is used to decipher the diffraction patterns and translate them into three-dimensional electron density maps, which are subsequently virtually filled with atoms and refined to obtain a final structure.[67]

X-ray crystallography provides an attractive approach for structure determination with a relatively low sample size needed for the experiment, thousands of solved structures, and a long tradition of software and hardware development.[68] The main drawbacks of X-ray crystallography are experimental settings (based on a trial and error approach),[69] non-physiological conditions of crystallization (high salt concentrations, usage of heavy metal ions or non-physiological pH values, and solid-state of the sample),[69] and incorrect data handling (e.g. fitting atoms into poorly solved electron density clouds or negative difference density map).[70, 71]

Hence, the preparation of the macromolecular structure for an *in-silico* experiment is an important step to become aware of crystallization artifacts, which may lead to wrong conclusions.[72]

2.2.3 Computational Methods in Structure-Based Drug Design

Computational methods in structure-based design support the understanding of mechanisms behind pharmacological actions of drugs or physiological functions of biomolecules.[73] Their crucial role in drug discovery campaigns was underlined and reviewed several times.[74, 75] In this thesis, the following computational methods were used and will be described in detail: homology modeling, ligand and protein-protein docking, 3D pharmacophore modeling, and molecular dynamics (MD) simulations.

Prediction of Unknown Protein Structures: Homology Modeling

When the protein 3D structure is unknown, homology modeling (or comparative modeling) can be used to obtain this information. There are two distinct approaches:

template-free and template-based modeling.[76] Template-based homology modeling was used in this thesis and will be described herein in detail.

In the first step, a template protein with an experimentally obtained structure and sequence similar to the target sequence must be found. The sequence alignment of the target protein and template can be performed with algorithms such as Clustal Omega[77] or T-Coffee[78] for multiple sequences and Needleman-Wunsch,[79] Smith-Waterman,[80] FASTA,[81] or BLAST[82] algorithms for pairwise alignment. The algorithms differ in gap handling, whereby handling of mutations and similarity between the amino acids are common for all algorithms and mostly based on BLOSUM[83] (Blocks Substitution Matrix).

In the next step, the 3D structure of the target protein is built based on the coordinates of the template. Several software packages and web servers have been developed (Phyre2,[84] iTasser,[85] SWISS-MODEL,[86] MOE) for this purpose. In this thesis, the MOE (Chemical Computing Group ULC, Montreal, Canada) homology model builder[87] was used, hence the methodology will be briefly described here: when the alignment of the template and target sequences is complete, the algorithm constructs an initial model by copying identical residues. In the next step, the backbone is sampled from a database containing high-resolution protein structures. Subsequently, side chain rotamers are constructed and scored by a GB/IV function,[88] which estimates the solvation energy of the side chains and allows prediction of whether a side chain is solvent-exposed or not. In the last step, the final model is selected based on the average structure of generated models and energy minimized.

The final step of homology model building is validation. For this purpose, general quality features should be checked (Ramachandran outliers,[89] global protein fold, atomic clashes, and, if applicable, positioning of disulfide bonds). Moreover, the homology model can be validated by docking known ligandss[90] or incorporating knowledge from mutational studies.[90]

Generation of Binding Hypotheses: Ligand and Protein-Protein Docking

Docking is a method used to obtain 3D models of binding partners, such as a small molecule in a complex with a protein, or one protein bound to another protein.

Molecular docking aims to place a small molecular ligand into a binding site of a macromolecule. Many strategies are implemented to reach this goal:[91] the incremental

reconstruction approach (FlexX[92]), molecular- or surface-shape (DOCK,[93] Surflex[94]), Monte Carlo search (Autodock[95]), systematic search (Glide[96]), and genetic algorithm (GOLD[97]). GOLD was used in this thesis due to comparably good performance.[98, 99] In the following section, the genetic algorithm and implemented scoring functions will be briefly described.

The genetic algorithm[100] used by GOLD translates the information about conformations of ligand and protein (e.g. torsional angles) into a computer-readable number-string termed "chromosome". In the next step, all possible interactions between ligand and protein are detected. The algorithm calculates the energy of the ligand-protein complex (hydrogen bonds, lipophilic contacts) and the internal energy of the ligand (indicating relaxed or restrained ligand conformations). The best-scored parental "chromosomes" are then combined or slightly changed by the software using evolutionary processes (mutations, cross-over, and migration of components of the number-strings) to result in offspring "chromosomes". These are iteratively evaluated once more, undergoing the evolutionary processes to obtain final docking poses.

Docking programs use scoring functions for the assessment of generated docking poses. In this thesis, the ASP[101] and the Chemscore[102] functions were used for scoring and re-scoring, respectively, as they are recommended by GOLD for protease docking. Both are empirical scoring functions, derived from calculations of the binding energy of reported ligand-protein complexes.

In this thesis, the docking score was not included in the process of binding mode selection, as this value neither correlates with the binding affinities nor allows discrimination between native and virtual docking poses.[103] Molecular docking was exclusively used to obtain plausible ligand conformations in the binding pocket. This task is reported to be the only well-performing functionality of docking algorithms.[99]

To obtain a binding hypothesis of more complex partners, like proteins, an approach called protein-protein docking is used. In contrast to small molecule docking, *de-novo* determination of protein-protein binding interfaces (PPI) is a more complex task. Due to the high number of possible binding combinations between two macromolecules,[104] shallow and flat binding pockets, and large interaction areas,[105] knowledge about the nature of the interaction is needed for plausible results. Several software packages were developed for the generation of atomistic protein-protein binding models

(Haddock,[106] Rosetta[107]). In this thesis, Rosetta was used due to its efficient performance for the prediction of protein-protein binding conformations.[108, 109]
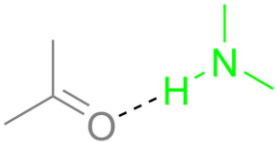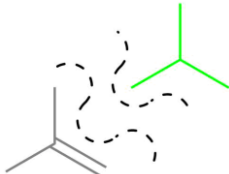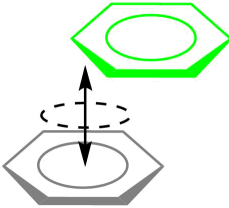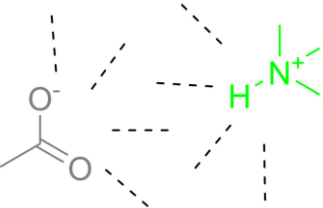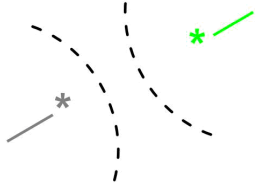
In the first step, Rosetta[107] utilizes a Monte Carlo algorithm to handle proteins as rigid bodies, attempting to adjust the coarse-grain surfaces of the binding partners to each other. In the next step, side chain rotamers are generated to refine the model. After each perturbation step, the algorithm verifies whether the energy of the generated side chain conformations is acceptable according to the Metropolis Criterion.[107, 108] The implemented scoring function calculates lipophilic contacts, hydrogen bond energy, solvation and surface area, electrostatic force, and residue rotamer probability terms for all atoms to obtain energy-minimum protein-protein binding conformations.[107] In total, several thousand docking runs are performed (at least 10,000 runs are recommended for global docking[104]) to sample conformational space. The output of Rosetta delivers several binding modes with their respective scoring values, where the lowest value (which represents the calculated binding energy) should represent the best docking pose. Due to the low accuracy of the scoring function, docking poses should be carefully analyzed using experimentally obtained knowledge to obtain the best result.[107]

Essential Protein-Ligand Interaction Patterns: 3D Pharmacophore Modeling

A pharmacophore is an aggregate of chemically defined interaction points and steric features, which determine the binding affinity and thus the biological activity of a small organic molecule.[110] The following chemical features are considered to be essential for ligand binding:[111] hydrogen bonds, lipophilic contacts, aromatic and ionic interactions, and exclusion volume spheres. A pharmacophore feature contains information about the geometry of the interactions as well as steric hindrance (Table 2).[111]

To generate a pharmacophore, different software can be used (e.g. LigandScout,[112, 113] Catalyst,[114] Pharmer,[115] Phase,[116] or MOE). All available software packages can create a 3D pharmacophore based on ligand structures, and most of them can also generate 3D pharmacophores based on ligand-protein complexes or apo structures of proteins. These pharmacophores can be used for virtual screening[117] to discriminate between potentially active and inactive molecules from compound libraries.

Table 2. Feature types (interactions or steric information) commonly used in pharmacophore modeling.

Feature Type	Interaction Geometry	Interaction Partners[a]
Hydrogen Bond	*Distance*: 2.6 - 3.2 Å [113, 118] *Angle*: 127 - 135° [119]	
Lipophilic Contact	*Distance*: 1.0 - 5.0 Å [113] *Angle*: 360° [113]	
Aromatic Interaction	*Distance*: 3.4 - 3.6 Å [118] *Angle*: 0 - 20° [113]	
Ionic Interaction	*Distance*: explicit calculation up to 5.6 [113] - 12 Å [120] *Angle*: 360° [121]	
Shape / Steric Hindrance	*Distance*: depending on the structure *Angle*: 360° [117]	

[a] Interaction partners are indicated in green for a macromolecular partner and grey for a ligand. Asterisk (*) indicates any atom.

The applicability of the 3D pharmacophore modeling reaches far beyond classical virtual screening campaigns[122-124] (as presented further in this thesis); 3D pharmacophores can be used to mechanistically explain why ligands can trigger different pharmacological actions[125] or to validate homology models.[90, 126] The novel applications of pharmacophore modeling were accompanied by the development of innovative concepts, as dynamic pharmacophores[73, 127] or advanced interaction

sampling with water molecules as molecular probes.[128] Both methods were used in this thesis and will be briefly discussed below.

The dynamic pharmacophores (*Dynophores*)[73, 127] application was developed in-house as a method for analyzing interactions between ligand and macromolecule over the course of an MD simulation. *Dynophore*-App generates interaction density clouds, a graphical representation of spatial interaction probability distribution. Moreover, the *Dynophore* application provides distance distribution and occurrence-frequency graphs. Dynamic pharmacophores overcome issues known from static pharmacophore modeling: underestimation of target and ligand flexibility and rigid geometrical boundaries of interactions. *Dynophores* were successfully used to demonstrate how the flexibility and multiple binding modes of a certain flaviviral protease inhibitor contribute to its superior activity over other ligands reported in this publication.[124] A separate publication shows that *Dynophores* sufficiently explain activity cliffs between ligands sharing almost identical static interaction patterns;[73] analysis of interaction density clouds revealed differences in angle and distance distribution of a crucial hydrogen bond in correlation with ligand activity.

PyRod[128] is a software developed in-house that facilitates the detection of interaction points in a binding pocket by tracing water-protein interactions throughout MD simulations. This method takes into account target flexibility and allows prioritization of pharmacophore features by the duration and geometry of water-protein interactions. It has been shown in this thesis and the associated publication[124] that *PyRod* can generate pharmacophore models that can be successfully used for the prospective identification of novel pan-flaviviral protease inhibitors.

Regardless of how it is derived, a 3D pharmacophore model should be validated.[117] The validation of pharmacophore models is carried out using known active ligands targeting the binding site of interest as well as inactive molecules. The inactive molecules can be either reported as inactive in the literature or assumed to be inactive (in which case they are termed decoys). In this thesis, the decoy strategy was used due to scarce ligand data availability. Decoys can be obtained from web servers (e.g. DUD-E[129]). Briefly, DUD-E uses submitted active ligands to find molecules sharing similar physicochemical properties yet differing in structure (calculated as the difference in Tanimoto coefficient). Finally, the sets of active and decoy/ inactive molecules are screened against the pharmacophore model. The quality of the model

can be characterized based on (i) the absolute number of returned active molecules, (ii) receiver operating characteristic (ROC) curve, or (iii) early enrichment factors (e.g. $EF_{1\%}$).[130] A well-performing pharmacophore model returns a high number of reportedly active molecules during validation. Moreover, the active molecules should show a better fit in the pharmacophore model than inactive molecules during the validation process (indicated as a high value of $EF_{1\%}$). A good model shows a large area under the ROC curve, indicating a high specificity and selectivity of the model.

Criteria for Selecting Virtual Hits: Visual inspection

Visual inspection is a process by which ligand binding hypotheses are manually selected based on rational empirical rules and the researcher's experience on interactions with the target.[131] Although not a computational method, the visual inspection of binding hypotheses and ligand structures inseparably accompanies virtual screening campaigns to allow the filtering of obtained hits.[131] In this thesis, visual inspection was extensively employed for the selection of the final binding hypotheses of small molecular ligands and hence will be briefly described.

A recently-published review by Fischer and colleagues[131] presents statistics on the popularity of visual inspection criteria based on a survey and feedback from around 100 experts in molecular modeling. The survey included 20 different criteria for visual inspection of docking poses illustrating a large number of views on the selection problem. In this thesis, the following four experience-based criteria were taken into account when inspecting binding modes and ligand structures:

- fulfillment of specific ligand-protein interactions (pharmacophore fit),[131]
- exclusion of binding modes exposing lipophilic moieties to the solvent (entropic penalty),[131]
- exclusion of binding modes establishing hydrogen bonds via charged moieties but lacking an ionic interaction,[131] and
- exclusion of binding modes with strained ligand conformations (non-coplanar pi-bond systems, twisted alkyl chains, non-planar aromatic moieties).[131]

The visual inspection criteria of the docking poses used in this thesis are completed by the following specific rules, which exclude non-drug-like structures and structures

that could potentially interfere with biochemical assays, including pan-assay interfering compounds[132] (PAINS) and frequent hitters:[133]

- exclusion of non-drug-like structures (e.g. polyfluorinated or polyaromatic compounds),
- exclusion of reactive moieties (e.g. quinones, alkyl halides, peroxides, aziridines, Michael acceptors, aldehydes, known cytostatic drugs), and
- exclusion of frequent hitters (e.g. flavonoids, curcumin[134]).

Molecular Dynamics Simulations

Molecular dynamics (MD) simulations are calculations used to explore the motion of molecular models. Briefly, to perform an all-atom MD simulation, the system (protein, solvent, salts, ligands) is parameterized using a force field. During the parameterization step, partial charges and atom types (including connectivity, van der Waals radii, and atomic masses) are distributed on the particles.[135] Subsequently, the system undergoes several equilibration steps. The coordinates obtained by X-ray diffraction or Cryo-EM only present an average probability of atom positions leading to non-equilibrium structures of biomolecules.[136] The equilibration is necessary to obtain a system free of structural artifacts.[136] In the main simulation run, random starting velocities based on Boltzmann distribution at a given temperature are assigned to all atoms of the parameterized system.[137] Examples of engines for carrying out MD simulations are Desmond[138] (extensively used in this thesis), Amber,[139] and GROMACS.[140]

During an MD simulation, new positions and velocities are recorded by integration of Newton's law of motion for each time step, similar to a movie consisting of multiple consecutive frames.[141] The potential energy of the system used for the force calculation is described within the force field (e.g. OPLS,[142] GAFF,[143] CHARMM[144]) as a sum of bonded and nonbonded interactions (Figure 8).[145]

The calculation of forces for all bonds and non-bonded interactions is computationally expensive. Therefore, MD simulations are conducted in a nano- to microsecond range to observe the dynamics of a system. Due to the constant improvement of hardware and software, the average simulation time and complexity of the systems are rapidly increasing.[146]
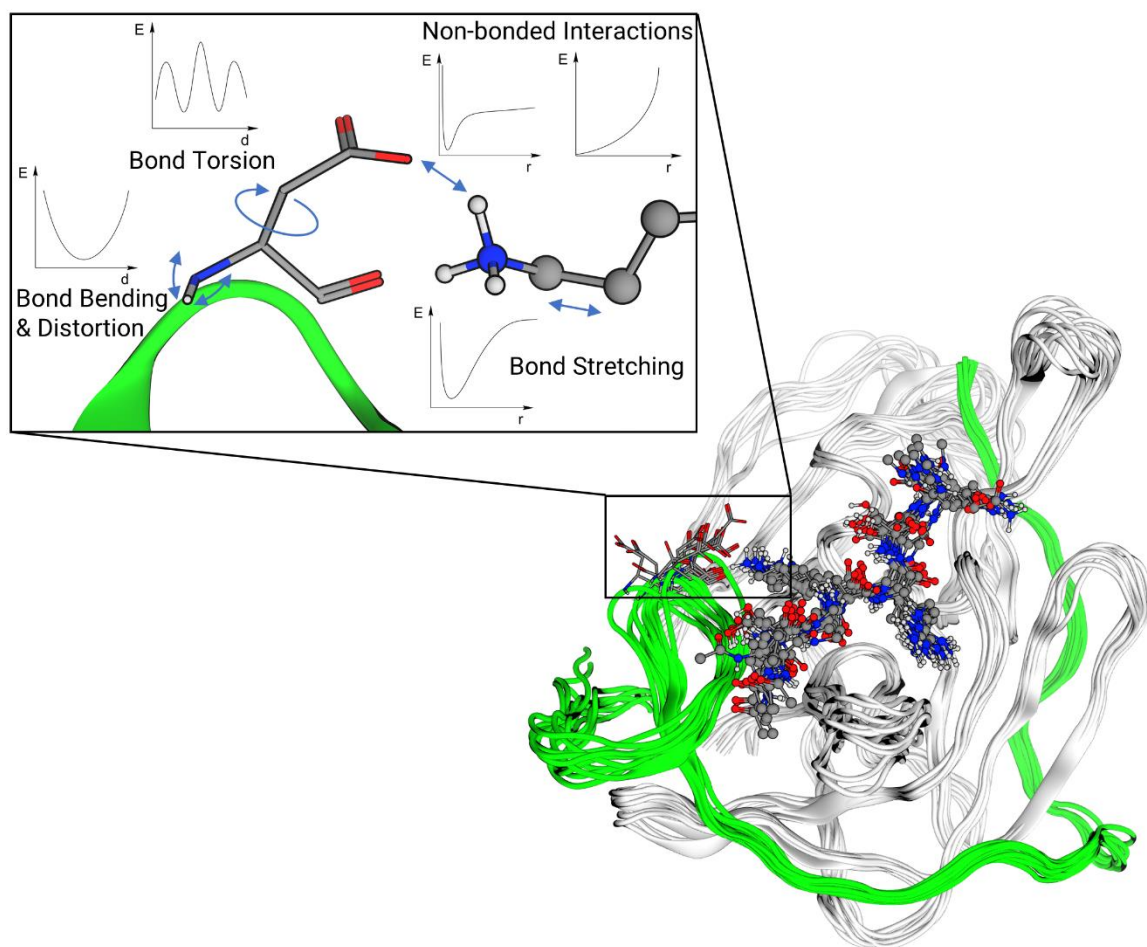
Figure 8. Representative time steps of a molecular dynamics simulation of the West Nile virus protease (green and grey backbone representation) in complex with a substrate (sphere and stick representation). The upper panel presents the forces (blue arrows) calculated during an MD simulation and respective energy-function graphs.

2.3 Proteases as Antiviral Drug Targets

Viral proteases represent validated and valuable targets for directly acting antiviral drugs (DAAD). As of 2021, two protease inhibitor classes are approved for the treatment of human immunodeficiency virus[147] (HIV) and hepatitis C virus[148] (HCV) infections. Beyond these two examples, proteases represent promising drug targets for the treatment of infections with flaviviruses, coronaviruses, or herpesviruses.[11, 149, 150]

2.3.1 Structures of Proteases

Proteases are enzymes that catalyze the hydrolysis of an amide bond in protein substrates. They can either be classified by (i) structure of the active center and

hydrolysis mechanism or (ii) evolutionary origin[151] as presented in the MEROPS database.[152] In the following, the catalytic and general structural protease features according to the MEROPS database will be summarized.

Currently, nine protease families are known.[152] Each family is represented by a letter indicating the catalytic residue: A- aspartic, C- cysteine, G- glutamic, M- metallo, N- asparagine, P- mixed, S- serine, T- threonine, and U- unknown proteases. Viral proteases are represented in every family, except P and T.[152] Families C, G, N, and S contain a cluster of catalytic residues, the *catalytic dyad* or *triad*, in the active center (Figure 9, e.g. the H-C dyad in coronaviral M^{pro} [40] or the D-H-S triad in flaviviral NS2B-NS3[153]). The catalytic mechanism of all families can be classified according to the reaction intermediate (Figure 9), which can be either covalently bound (an ester in families S and T, a thioester in family C, or a hemiaminal in family N) or non-covalently bound (families A, G, and M).[154] The catalytic activity of the proteases allows an increase in the protein hydrolysis rate by the order of 10^{10}.[154] This unique capability is supported by three main structural features: (i) the catalytic center, (ii) the oxyanion hole, and (iii) the polypeptide binding site (Figure 10).[154] Since protease family S is the subject of this thesis, its structural features will be briefly elaborated in the following.
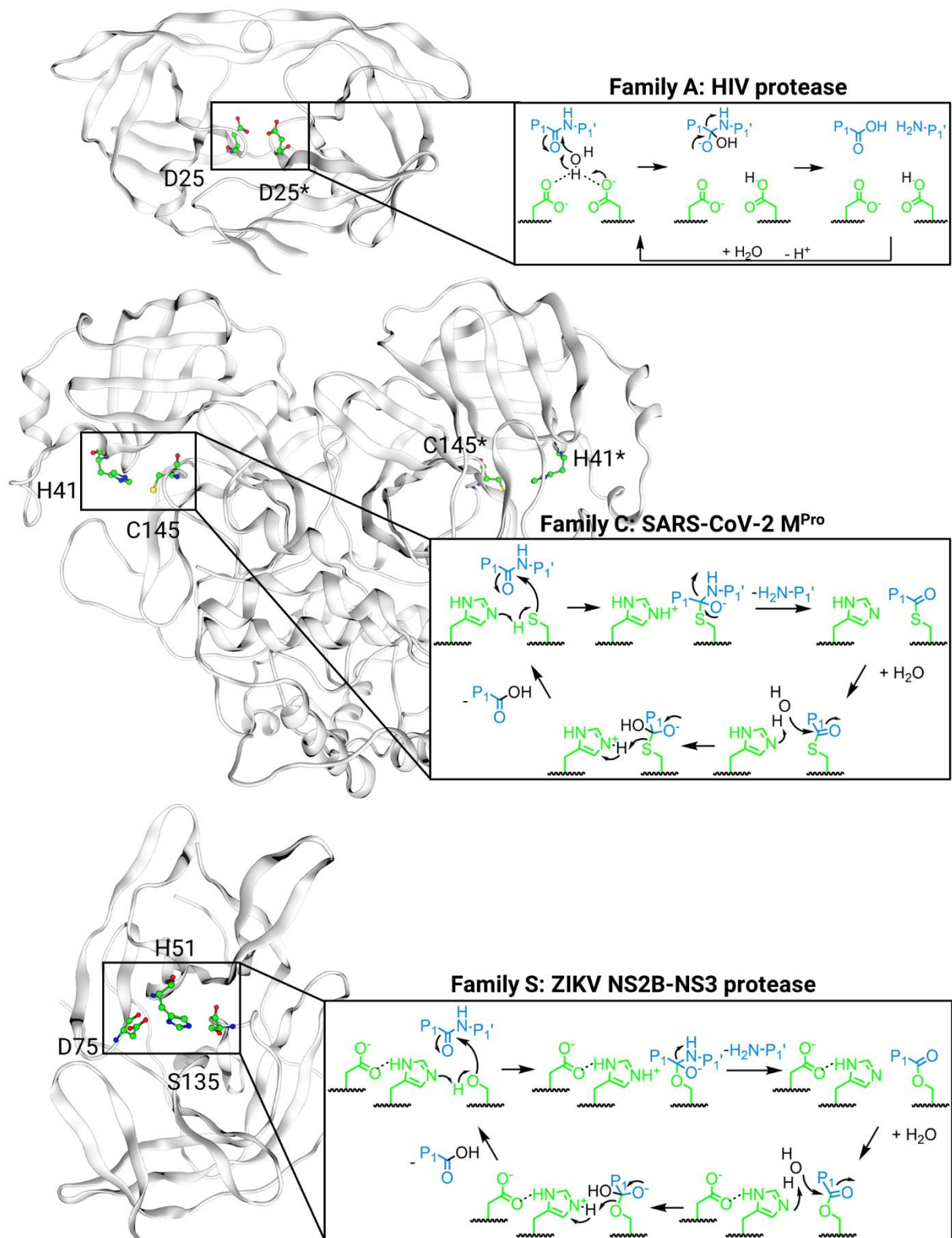
Figure 9. Examples of viral proteases[40, 155, 156] (grey ribbon) from families A, C, and S. Residues of the catalytic center are indicated in stick and sphere representation. The right-hand panels show the proteolysis reaction mechanisms (green atoms- catalytic center, black- water molecules, blue- substrate). Residues marked with an asterisk are a part of protomer B in a homodimer.
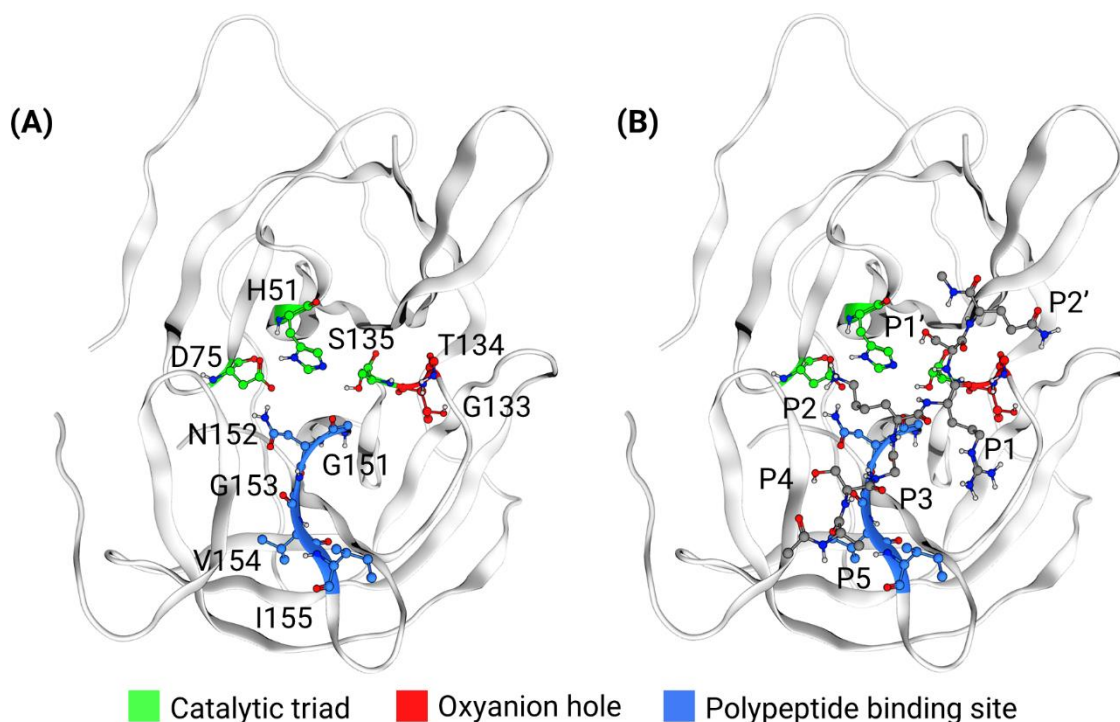
Figure 10. (A) Structural features determining the catalytical activity of a protease (grey backbone, PDB-ID: 5IDK[157]): (i) catalytic triad (green ribbon, sphere and stick model), (ii) oxyanion hole (red ribbon, sphere and stick model), and (iii) polypeptide binding site (blue ribbon, sphere and stick model). (B) Alternating substrate subpockets (indicated with P) in the canonical conformation of the substrate (docking pose, dark grey sphere and stick model).

On the one hand, the catalytic center (Figure 9, Figure 10A) enables a nucleophilic attack on the carbonyl C of the substrate, allowing the amine moiety, a bad leaving group, to detach from the reaction center.[154] On the other hand, the catalytic center polarizes the O-H bond of a water molecule to make it more nucleophilic, which allows the hydrolysis of the covalent intermediate.[154] As shown in Figure 9, all hydrolysis mechanisms include a negatively charged tetrahedral species derived from the carbonyl group of the substrate. This charge is stabilized by a protease feature called the *oxyanion hole*, which consists of partially positively charged N-H hydrogen atoms of the protease backbone adjacent to the catalytic residue (Figure 10A).[154] This feature is essential for the catalytic activity of proteases.[154] The last structural feature contributing to catalytic activity is the polypeptide binding site (Figure 10A), which facilitates positioning of the substrate on the protease and allows the scissile bond to be attacked by the catalytic center.[154] The polypeptide binding site is mostly an antiparallel β-sheet establishing hydrogen bonds to the backbone of the substrate.[154]

This conformation is called the "canonical conformation"[154] and forces the side chains of the substrate to occupy alternating protease subpockets (Figure 10B). According to the Schechter-Berger nomenclature,[158] protease subpockets are indicated with the letter P and a number. The numbering starts on the residues adjacent to the scissile bond, the N-terminal part of the substrate is numbered with Arabic numerals (designating the non-prime site), and the C-terminal part is marked with an additional " ' " symbol (designated the prime site, Figure 10B). The first subpocket of the non-prime site (P1) is called the specificity pocket.[154]

In this thesis, three closely related proteases were targeted: the NS2B-NS3 proteases from Zika, West Nile, and Dengue virus 2. The NS2B-NS3 protease accepts substrates with basic side chains (R or K) in the P1 and P2 subpockets (Figure 11).[159]

| | | | Scissile bond | | | | | | | Scissile bond | |
Species	P4	P3	P2	P1	P1'	Species	P4	P3	P2	P1	P1'
Zika virus NS2B-NS3	S	G	K	R	S	Dengue virus 2 NS2B-NS3	S	K	K	R	S
	T	G	K	R	S		K	K	Q	R	A
	A	G	K	R	G		A	G	R	K	S
	V	K	R	R	G		N	T	R	R	G
West Nile virus NS2B-NS3	N	R	K	R	G						
	Y	T	K	R	G						
	S	G	K	R	S						
	G	L	K	R	G						

Figure 11. Substrate spectra of NS2B-NS3 proteases from Zika, West Nile, and Dengue virus 2.[159]

The flaviviral NS2B-NS3 is a heterodimeric protein consisting of three structural features:

- the helicase, a C-terminal NS3 domain[160] responsible for the unpacking of double-stranded RNA intermediates produced during replication,
- the N-terminal NS3 serine protease that processes the viral polyprotein into functional units.[161] The protease is connected to the helicase via a short flexible loop;

- the NS2B is a cofactor spanning the protease domain of NS3. The NS2B subunit is essential for the protease activity enabling proper folding of the binding site and substrate binding.[161, 162]

The substrate binding site of the NS2B-NS3 protease is flat and hydrophilic.[163] All substrate binding pockets of flaviviral proteases are highly conserved (Figure 12A-C).[163] The non-prime site is divided into subpockets S1-S4. For the localization of the S3 and S4 subpockets only vague hypotheses exist: based on the X-ray crystal structures of peptidomimetic inhibitors (e.g. Bz-nKRR-H in PDB-ID: 2FP7[162]), Braun and colleagues[164] place the S3 subpocket below the S2 subpocket, and contend that the S4 subpocket is solvent-exposed (Appendix Figure 1). The analysis conducted by Chappell and colleagues[159] based on the docking studies of the substrate to the WNV protease renders the first hypothesis implausible: due to a crystallization artifact of the tetrapeptide inhibitor, the aromatic ring establishes a π-cation interaction to the guanidine group in the S1 subpocket, leading to a spiral-like inhibitor conformation (Appendix Figure 1). We support the second hypothesis that localizes the S3 and S4 subpockets to alternating positions according to the canonical conformation (Figure 12D).[159]

The suggested binding mode of a substrate is briefly characterized based on the example of ASGKR↓SQ (scissile bond indicated by arrow) in complex with the WNV protease. The S2' and S1' subpockets are nonspecifically recognizing a variety of residues. Mutational studies on the closely related DENV2 protease suggest that NS3-Y34 plays an important role in the binding of peptide inhibitors.[165] The transition from the prime to the non-prime binding site is marked by the catalytic triad NS3-D75, H51, S135.[166] The oxyanion hole is composed of NS3-S135, T134, and G133. The polypeptide binding site consists of NS3-G151, G153, and I155. The S1 subpocket establishes cationic interactions and hydrogen bonds between NS3-D129 and the arginine of the substrate. The residue NS3-D129 is crucial for enzymatic activity.[159] In the S2 subpocket, NS3-D75 and NS2B-N84 interact with the amino group of the substrate.[159] The S3 pocket is nonspecific and solvent-exposed. Docking studies suggest that lipophilic residues of the substrate establish lipophilic contacts to NS3-Y161.[159] Finally, the S4 subpocket is nonspecific and substrate binding is supported by multiple NS2B and NS3 residues.
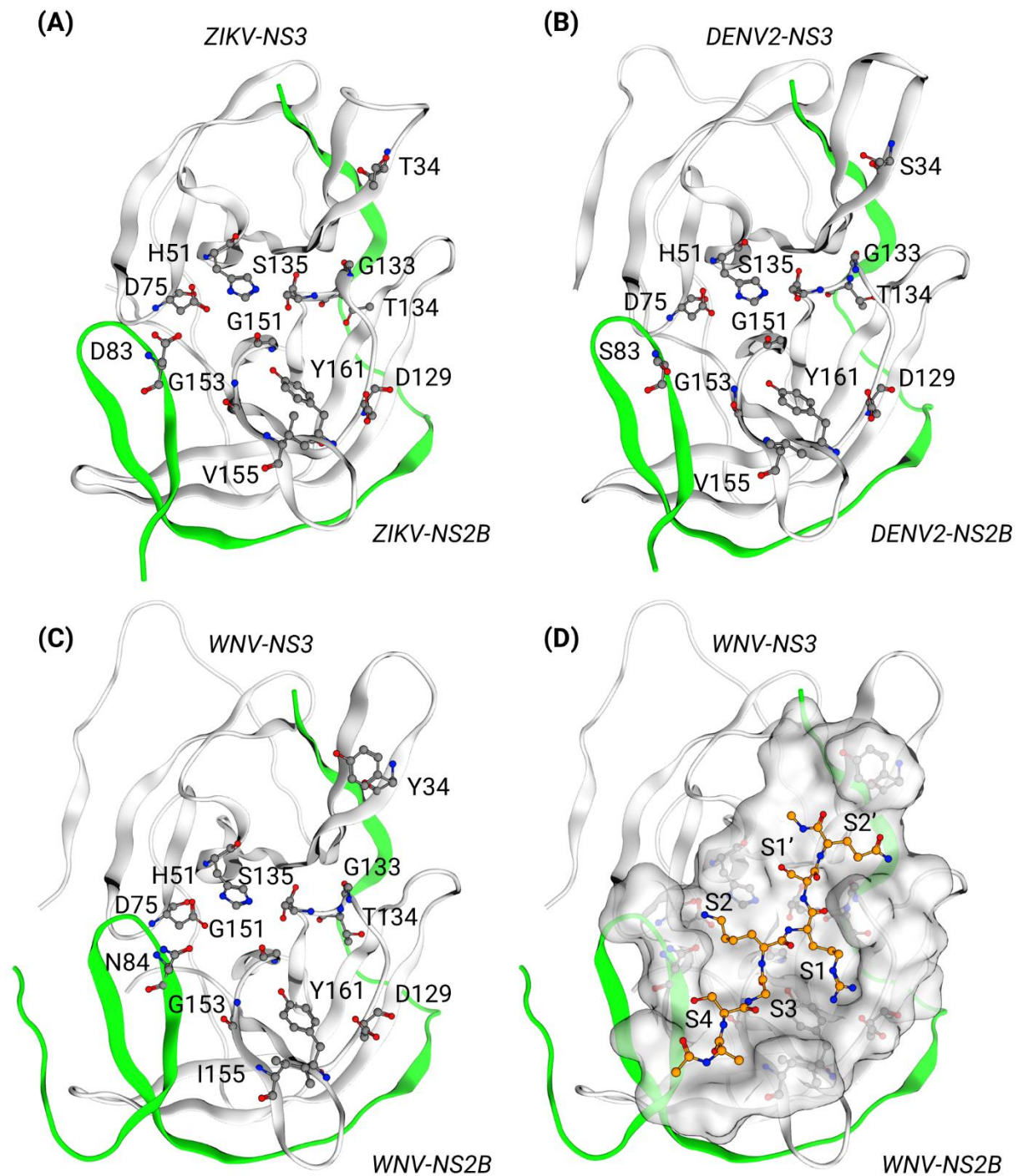
Figure 12. Global view of the NS2B-NS3 protease of Zika virus (A) and focus on competitive binding sites of Dengue 2 (B) and West Nile virus (C). (D) A suggested binding mode of a substrate (ASGKR↓SQ) of the West Nile virus protease. Color code: green backbone- NS2B, white backbone- NS3, grey spheres- protein carbon atoms, orange sticks and spheres-substrate carbon atoms, red spheres and sticks- oxygen atoms, blue spheres and sticks-nitrogen atoms.

Known inhibitors of NS2B-NS3 protease mostly mimic the substrate structure, leading to peptidomimetic inhibitors such as **cn-716**[153] or **Bz-nKRR-H**[167], which display potencies in the high nanomolar range (Figure 13). Due to undesired pharmacokinetic properties of this chemotype, several non-peptidic inhibitors were developed, of which the most potent examples are compound **6**[168] and PanThera **166347**[169] (Figure 13). The small molecular inhibitors show moderate potency in the one- to two-digit micromolar range. The high molecular masses of the reported molecules (436 Da and 611 Da for compound **6** and PanThera **166347**, respectively) render them poor starting points for optimization campaigns. Hence, novel and more potent inhibitors are needed.

cn-716 ($K_{i\,ZIKVPro}$ = 0.04 µM)   **Bz-nKRR-H** ($K_{i\,DENV3Pro}$ = 5.6 µM)

PanThera **166347** ($K_{i\,WNVPro}$ = 1.2 µM)   Compound **6** ($K_{i\,WNVPro}$ = 6.2 µM)
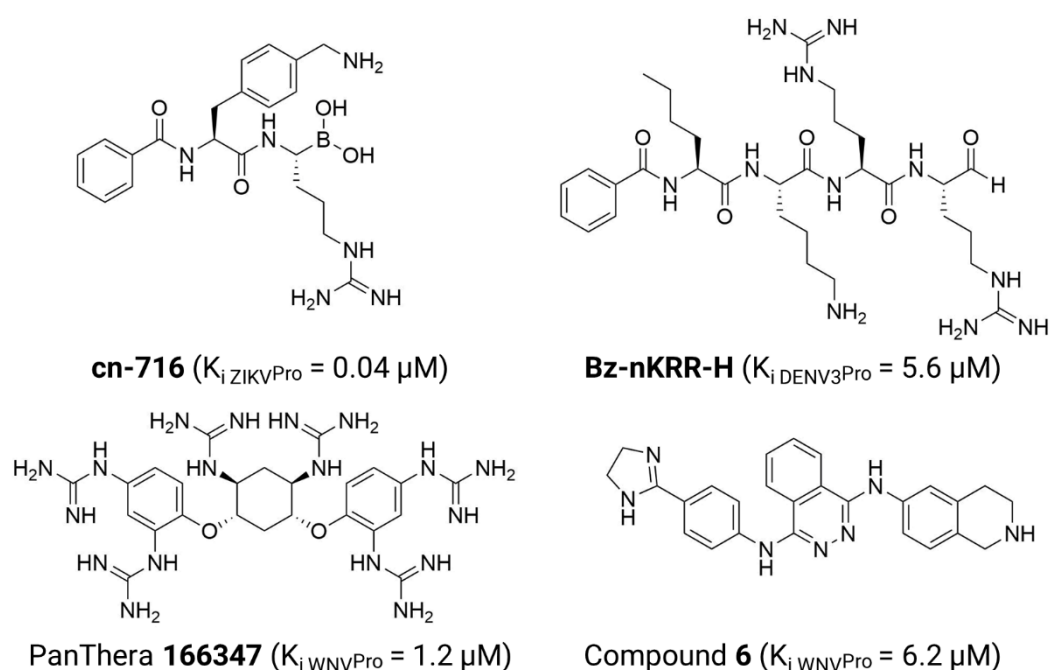
Figure 13. Examples of competitive inhibitors of NS2B-NS3 protease from Zika (ZIKV),[153] West Nile (WNV),[168, 169] and Dengue 3 virus (DENV3).[167]

2.3.2 Quantifying Inhibition: Available Protease Assays

In order to biochemically characterize proteases, several assays were developed, e.g. absorption-, fluorescence polarization-, and Förster resonance energy transfer (FRET)-based assays.[170] The principle of these assays is the measurement of substrate hydrolysis rate using short peptide sequences marked with a chromophore. To characterize the inhibitory potency of the compounds discovered in the course of this thesis, a fluorescence polarization assay was used, hence it will be briefly described.

During a fluorescence polarization assay, the release of a fluorogenic dye (e.g. 7-Amino-4-methylcoumarin, AMC) is measured (Figure 14). The protease cleaves an amide bond downstream of the specificity sequence (in the case of NS2B-NS3 protease, this site is located after a dibasic peptide). The released dye can rotate faster in the solution due to its lower molecular weight. When irradiated with polarized light, the free fluorophore has a higher chance of light emission than the bound dye. Subsequently, the increasing fluorescence intensity over time is measured as a surrogate parameter for protease activity.[170]
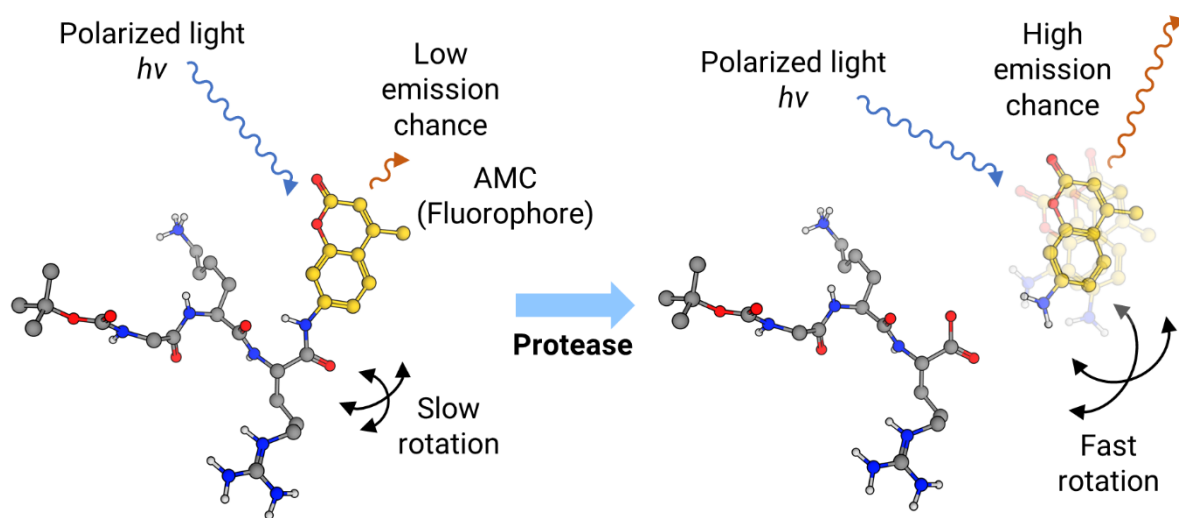
Figure 14. Principle of a fluorescence polarization assay shown using the example of the fluorogenic flaviviral NS2B-NS3 protease substrate Boc-GKR-AMC.

Based on the fluorescence intensity graphs obtained from the described assay, the following parameters can be derived to characterize protease activity: K_M (Michaelis-Menten constant) and k_{cat} (reaction rate, Figure 15A). According to Michaelis-Menten kinetics,[171, 172] K_M indicates the substrate affinity of the enzyme and k_{cat} the reaction rate constant in a two-step reaction model (Figure 15A).

Protease inhibitors modify either K_M or v_{MAX} (maximal reaction rate at infinite substrate concentration) values, depending on the inhibition mode (Figure 15B). Competitive inhibitors, which compete with the substrate for the binding site, increase the K_M value. This means that a higher substrate concentration is needed to reverse inhibitor binding. Non-competitive inhibitors (inhibitors binding to an allosteric pocket) decrease the maximal reaction rate by binding to the enzyme and enzyme-substrate complex. Inhibitory activity can be described using IC_{50} (half-inhibitory concentration)

38

or K_i (inhibitory equilibrium constant) values for non-covalent or fast-reversible covalent inhibitors and with k_{inact} (inactivation reaction rate) or k_{inact}/K_i values for slow-reversible or irreversible covalent inhibitors.
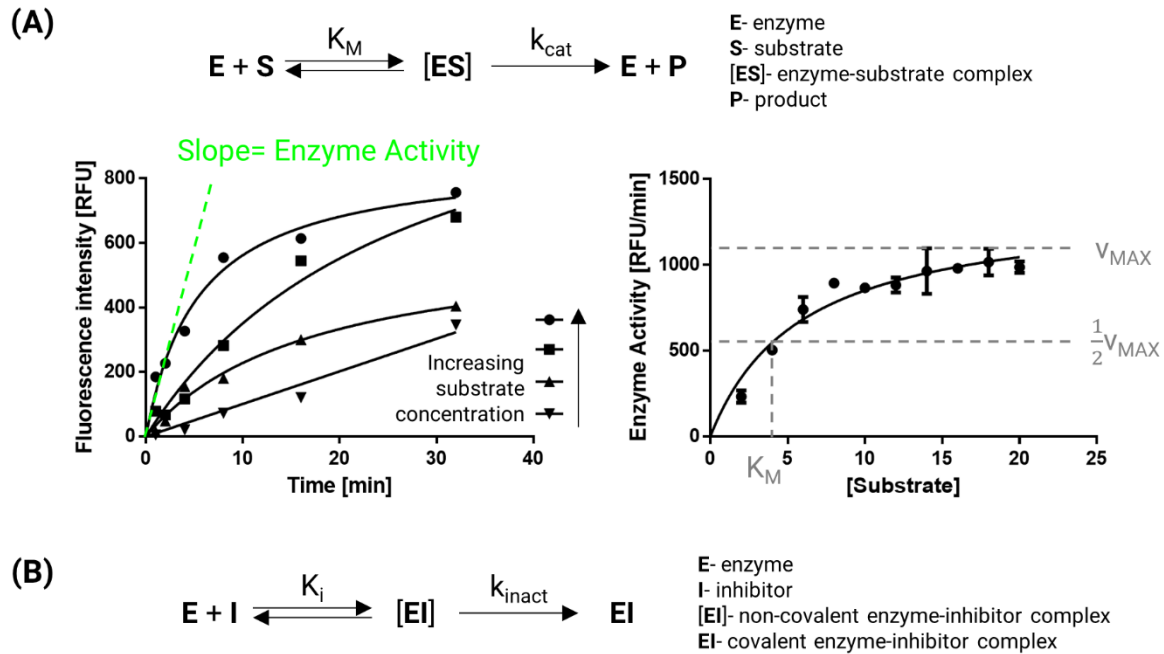
Figure 15. (A) The Michaelis-Menten kinetics equation is based on a two-step reaction model and can be derived from the fluorescence curves obtained from protease assays. (B) Reaction model when an inhibitor is added to an enzyme-substrate mixture.

The parameters describing inhibitory activity can be derived from the fluorescence curves obtained from the protease assays. To obtain a dose-response curve, the protease activity is measured under constant substrate concentration (usually around the K_M value) and varying inhibitor concentrations, as indicated in Figure 16.

The IC_{50} value describes the inhibitor concentration that causes 50% inhibition of the enzyme. This parameter is broadly used to report inhibitory activity. The IC_{50} value is strongly dependent on substrate concentration and therefore not comparable between two different assay setups. More comparable is the K_i value, which can be derived from IC_{50} values measured under varying substrate concentrations using the Cheng-Prusoff equation.[173]
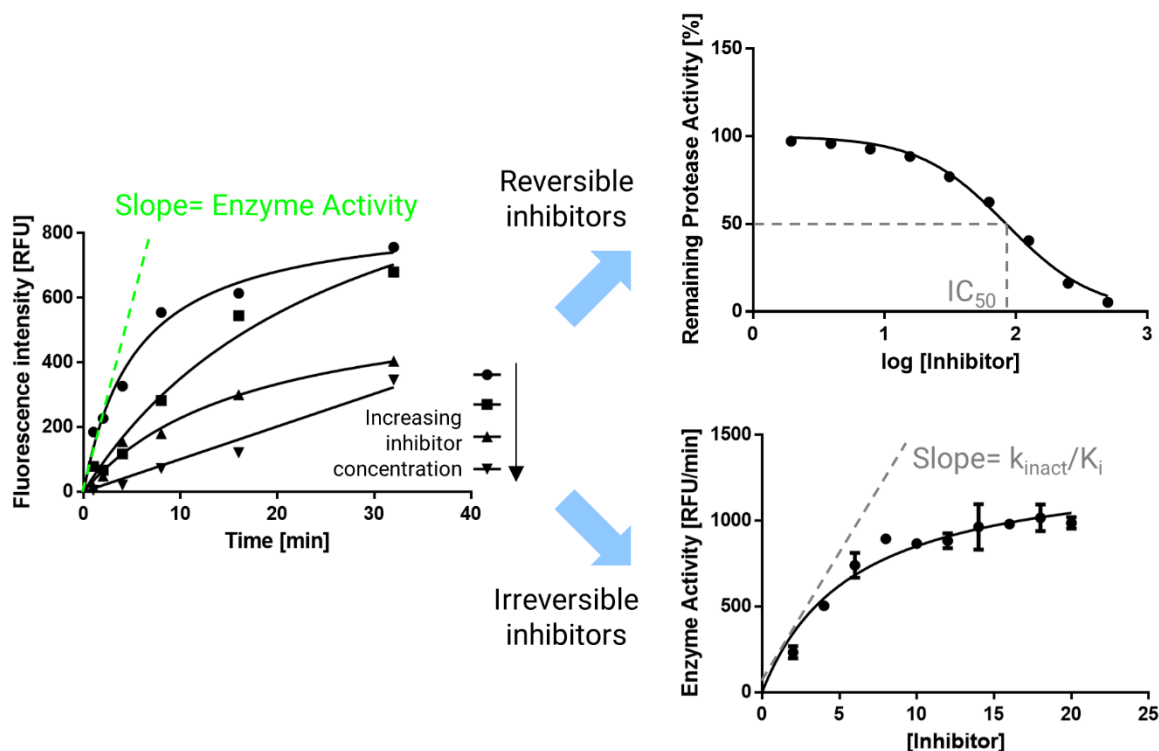
Figure 16. Dose-response curves obtained from protease activity measurements under constant substrate concentrations for non-covalent or fast-reversible covalent inhibitors (upper panel) and slow-reversible or irreversible covalent inhibitors (lower panel).

In order to characterize slow-reversible or irreversible covalent inhibitors, additional kinetic parameters are needed due to non-equilibrium binding kinetics. Commonly, the k_{inact}/K_i value, which describes the inhibition reaction rate (k_{inact}) and affinity of the enzyme to the inhibitor (K_i), is employed. This parameter can be calculated from dose-response curves based on the pseudo-first-order kinetics equation and plotting the obtained k_{obs} (pseudo-first-order reaction rate constant) against the inhibitor concentrations (Figure 16).

2.4 Viral Fusion Proteins as Drug Targets

Fusion proteins (FP) enable coalescence of the membrane of enveloped viruses with the host-cell membrane.[174] It has been shown that the membrane fusion step (called entry) is pivotal for the infectious cycle.[175] Hence, targeting viral FP represents a promising starting point for the development of novel antiviral substances. Currently, two drug classes targeting viral FP are present on the market. The first class includes the HIV entry inhibitors enfuvirtide[175] and fostemsavir[56] interfering with the HIV envelope glycoproteins 41 and 120, respectively, disrupting the fusion of viral

40

membrane with the human cell membranes effectively. The second group is represented by the hepatitis D virus (HDV) entry inhibitor bulevirtide.[176] Bulevirtide allosterically binds to the sodium-bile acid cotransporter hindering HDV replication. Another application of mechanistic studies on FP is the possibility of antibody and vaccine development[177] or prediction of the spectrum of hosts susceptible to viral infection.[178]

In this thesis, computational studies were used to support potential equine herpesvirus vaccine development and prediction of potential hosts for coronaviral infections. The understanding of FP − receptor interactions requires detailed knowledge about the structures of the proteins. Therefore, in the following section, the general structural features of FP will be described.

2.4.1 Structures of Viral Fusion Proteins

The C-terminal part of viral FP is anchored in the viral envelope and the N-terminal part is exposed to the environment. Hence, viral FP are categorized as type I transmembrane proteins.[177] Three classes of FP can be distinguished depending on the orientation towards the envelope, main fold features, and posttranslational processing (I-III, Table 3).

Table 3. Classification of viral fusion proteins and their general features (adapted from White et al.[174]).

Class	**Class I**	**Class II**	**Class III**
Major structural feature	α Helix	β Sheet	α Helix and β sheet
Orientation in native form	Perpendicular to the envelope	Parallel to the envelope	Unknown or perpendicular to the envelope
Priming	Yes	Yes	No
Examples	SARS-CoV-2 spike	ZIKV protein E	EHV-1 glycoprotein D

Abbreviations: EHV-1- equine herpesvirus 1, SARS-CoV-2- severe acute respiratory syndrome coronavirus 2, ZIKV- Zika virus.

Independently of the global fold of an FP, two functional regions can be distinguished: the receptor-binding domain (RBD) and the fusion subunit.[174] The RBD recognizes and binds to the viral receptor on the host cell surface. RBD-receptor binding or low pH values induce[174] subsequent large-scale conformational changes of the FP leading to the insertion of the fusion subunit into the host-cell membrane.[179] In the next step, a transition state, termed *hemifusion*, is formed, where the single membrane layers of the virus and the host cell merge. Finally, both membrane layers fuse to build a pore, allowing the viral capsid to enter the host cell (Figure 17).

An example of the class I FP and subject of this thesis is the spike protein (S)[174] of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).
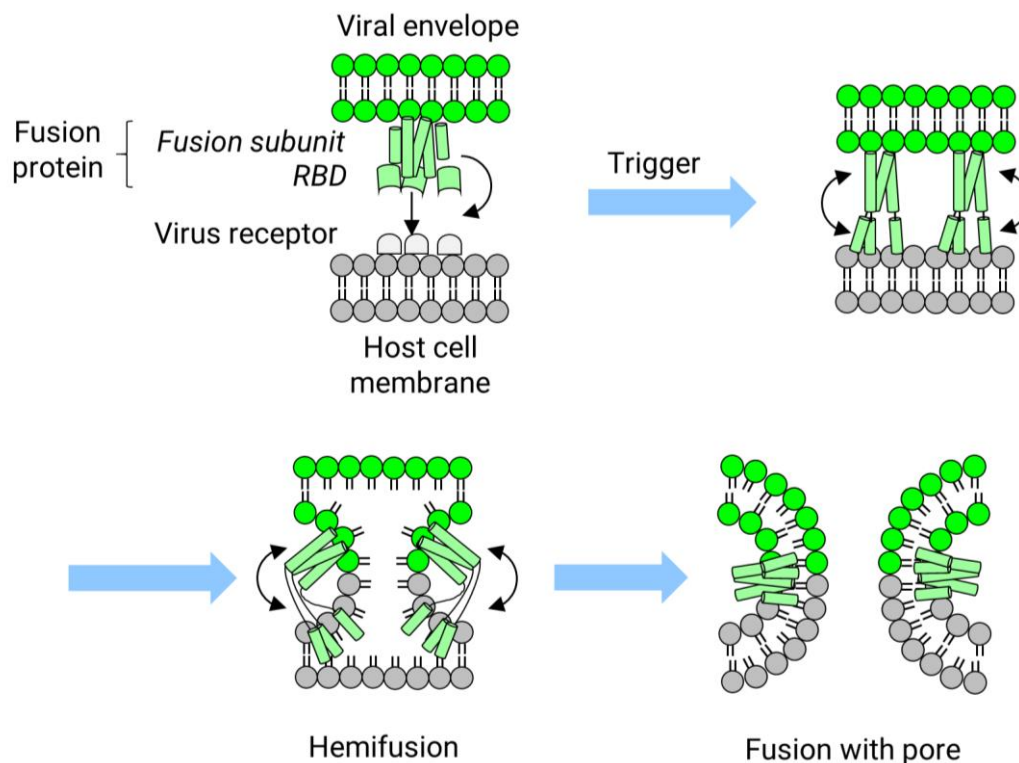
Figure 17. Schematic membrane fusion process of class I fusion proteins (adapted from White et al.[174]). The principle of membrane fusion for classes II and III fusion proteins is similar and hence not explicitly depicted. The receptor-binding domain (RBD) and virus receptor are not shown after the triggering step for the sake of clarity.

The name "spike protein" originates from the visible thorn-like form of the protein observed in images of whole coronaviruses. S contains a large α-helical core comprising the fusion subunit and the RBD consisting of β-sheets and flexible loops (Figure 18). S is processed by the host proteases furin and transmembrane protease

42

serine subtype 2 (TMPRSS2) to gain its fusogenic function.[41] The RBD of S specifically binds to the angiotensin-converting enzyme 2 (ACE2), a protease present on the surface of host cells.[8] It has been shown in an animal model that ACE2 binding is a crucial step for viral entry and is pivotal for the susceptibility of an animal species to SARS-CoV infection.[180] Hence, investigation of the S-ACE2 binding interface represents a viable strategy for the prediction of species susceptibility to coronavirus disease 2019 (COVID-19).

The structure of the ACE2-binding epitope of SARS-CoV-2 S is well characterized (Figure 19).[181] Three binding pockets can be distinguished in the first published X-ray crystal structure of S in complex with human ACE2 (PDB-ID: 6M0J) called here A, B, and C. S binds to a non-competitive binding site localized in the subdomain I of ACE2.
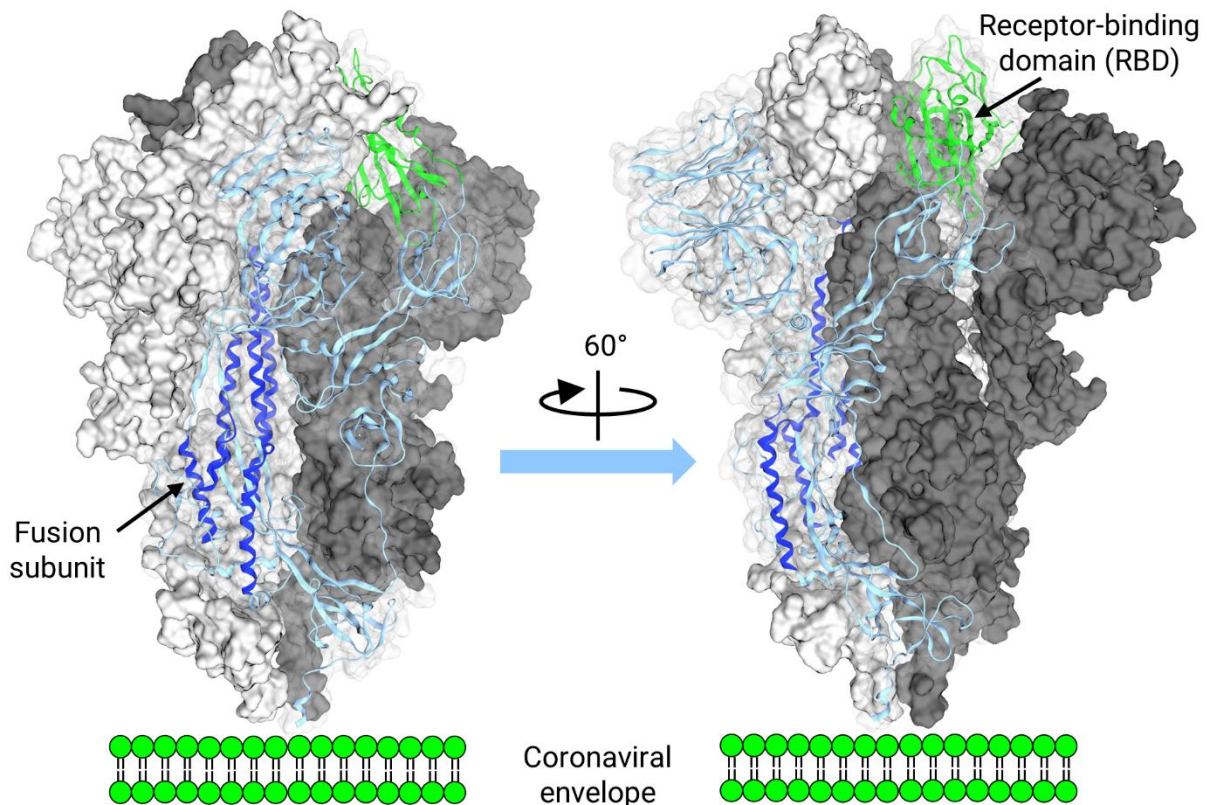
Figure 18. Extramembranous part of the spike protein from severe acute respiratory syndrome-related coronavirus 2 in its closed pre-fusion state (PDB-ID: 6VXX[182]). The protomer A is indicated with a white ghost molecular surface, other protomers (B and C) of the homotrimer are indicated with solid grey surfaces. Color code: dark blue ribbon- the α-helical core of the fusion subunit, pale blue ribbon- non-helical parts of the fusion subunit, green ribbon- receptor-binding domain.
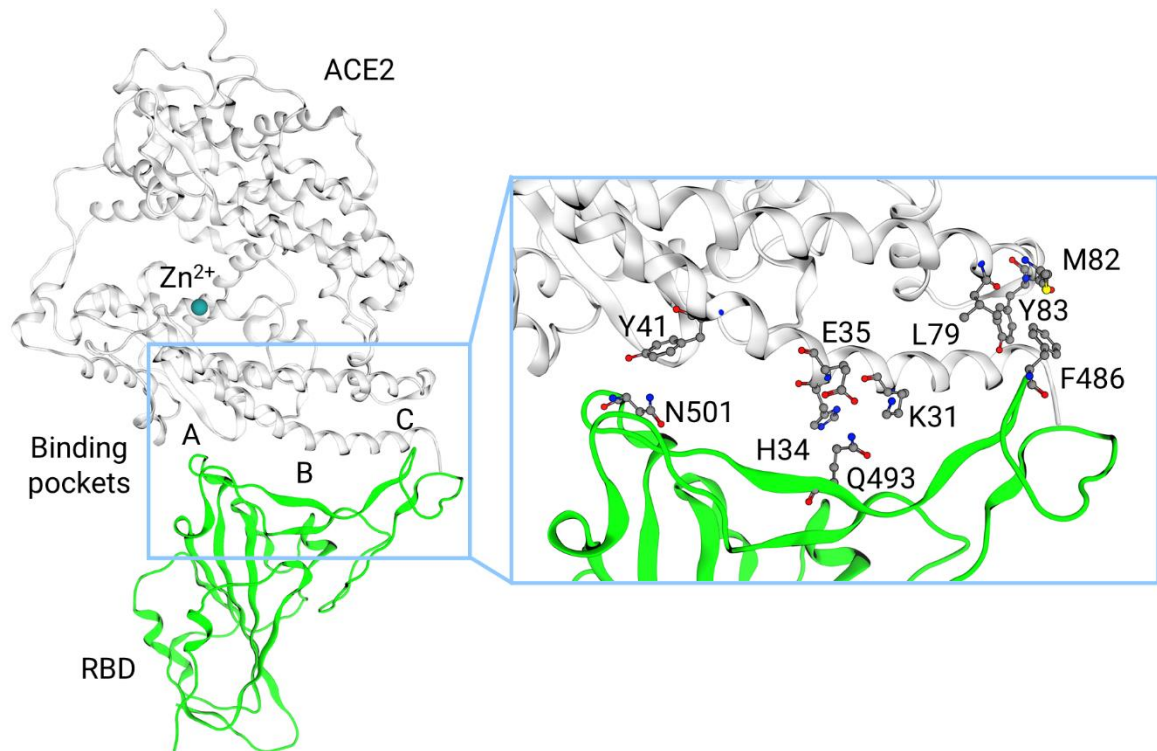
Figure 19. The receptor-binding domain of the spike protein in complex with its ligand, human angiotensin-converting enzyme 2 (ACE2). Color code: green ribbon- receptor-binding domain, white ribbon- human ACE2, cyan sphere- catalytic zinc ion, gray spheres- carbon atoms, red spheres- oxygen atoms, blue spheres- nitrogen atoms, yellow sphere- sulfur atom.

In binding pocket A, S is reported[181] to establish hydrogen bonding via N501 to the ACE2 residue Y41. In binding pocket B, S residue Q493 interacts with ACE2 residues K31, H34, and E35 via a hydrogen bond network. Finally, in binding pocket C, F486 from S establishes lipophilic contacts to ACE2-L79, M82, and Y83. The mechanistic understanding of ACE2-S interactions and the role of single residues in the protein-protein interface remain elusive. First insights into this matter are reported in this thesis and an accompanying paper.[178]

The second example of an FP investigated in this thesis is glycoprotein D (gD) from equine herpesviruses 1 and 4 (EHV-1, EHV-4). This FP belongs presumably to class III, like its human herpesvirus 1 homolog,[174] and is not yet extensively characterized. Currently, only the structure[183] (Figure 20) and the function of the RBD are known. The EHV-1 and EHV-4 gD bind with low millimolar affinity[183] to the equine major histocompatibility complex I (MHC-I)[33, 34] and are essential for viral entry into the equine cells.[33] Hence, investigating the gD-MHC-I binding interface might support the development of antiviral therapies, such as small molecule drugs or vaccines.
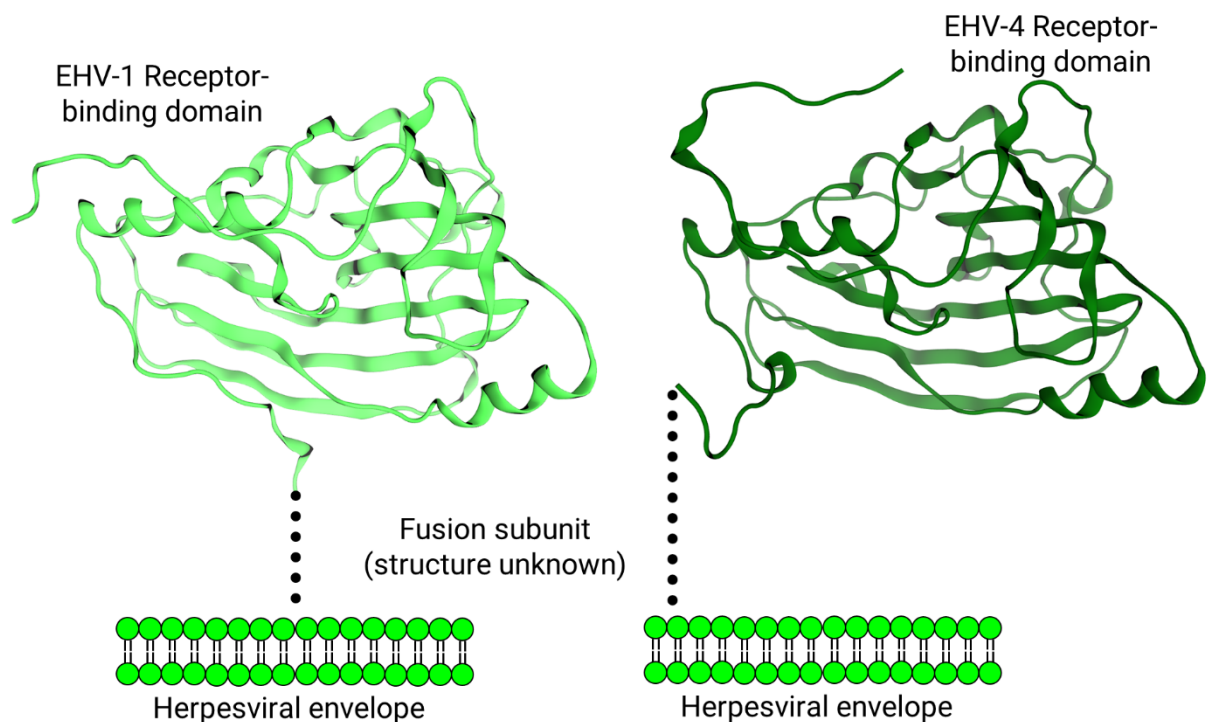
Figure 20. Structures of the receptor-binding domain of glycoprotein D from equine herpesvirus 1 (left panel, PDB-ID: 6SQJ) and 4 (right panel, PDB-ID: 6TM8).[183]

2.4.2 Quantifying Ligand Association: Available Assays

In order to biochemically characterize FP, parameters such as the dissociation constant of the receptor or ligand binding (K_D), molecular mass, and glycosylation patterns can be used. In this thesis, K_D data of EHV gD binding to its receptor MHC-I was analyzed to explain differences in predicted binding interfaces. The data was obtained by surface plasmon resonance (SPR) by Viviane Kremling under the supervision of Walid Azab in the Osterrieder Lab in Berlin (Germany), hence this method will be briefly described.

During an SPR experiment, a protein or ligand of interest is rinsed over a metal surface with an immobilized protein binding partner. The surface is irradiated with polarized light, causing propagation of an electromagnetic wave on the metallic surface (called surface plasmon polarization) and reflection of the light.[184] Due to the binding of protein or ligand to the immobilized partner protein, the thickness of the surface increases. This leads to a shift in the refractive index of the surface and therefore to a shift of the reflection angle of the light.[184] Based on the obtained association and dissociation rate constants, the equilibrium constant can be calculated.

2.5 Classification and Architecture of Protein-Protein Interactions

Protein-protein binding interfaces render highly interesting, yet challenging, targets for drug design campaigns due to their large size, topology, and restricted structural information.[105] Several approved antiviral substances interfere with viral-host protein-protein interactions. This strategy is valuable for rational drug design and is used to develop novel therapies. Currently, two drug classes are present on the market: entry inhibitors of HIV and HDV. Most recently, a peptidic entry inhibitor, bulevirtide,[176] was initiated to the market, representing a breakthrough in the therapy of hepatitis D. The analysis of protein-protein interface (PPI) architecture and mechanistic studies was used in this thesis to support the identification of a potentially immunogenic epitope for a vaccine, and development of an animal model for viral infection introduced in the previous chapters.

Protein-protein interactions can be classified based on their stability (obligate/ non-obligate), affinity (permanent/ transient, strong/ weak), and composition (homooligomeric/ heterooligomeric, Figure 21).[185] Despite large PPI areas consisting of a large number of residues, it has been shown that only a few crucial residues (called hot spots) contribute to the stability and affinity of a PPI.[186] Several concepts describing the contribution of single residues in PPI to the total binding energy have been developed (e.g. O-ring theory,[186] double water exclusion hypothesis[187]). In this thesis, the O-ring theory was extensively used and hence will be described in detail.
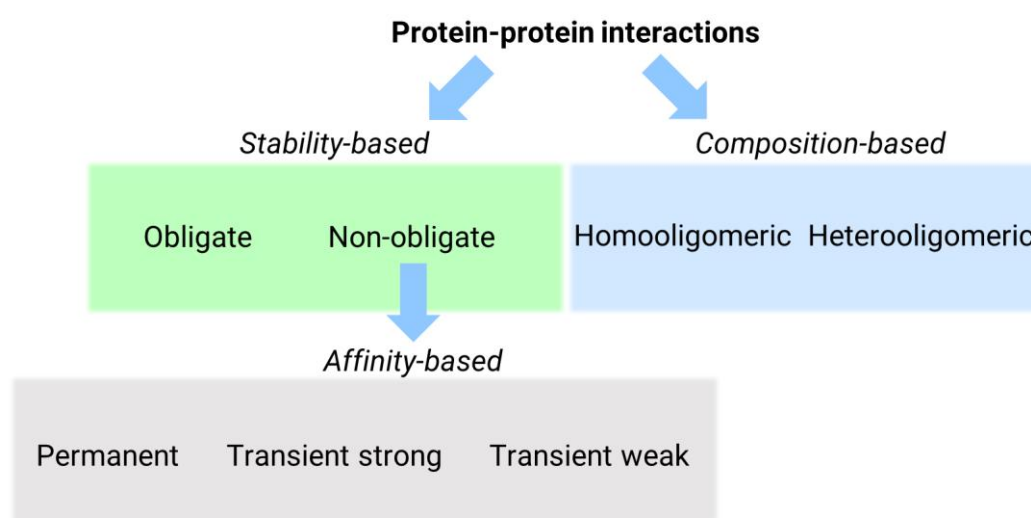
Figure 21. Classification of protein-protein interactions.

The O-ring theory is based on a statistical evaluation of over 2000 published alanine mutations in protein-protein binding interfaces with known structural data. Bogan et al. have calculated the fold of enrichment of each proteogenic amino acid in the dataset of hot spots over the dataset of all alanine mutations. A hot spot residue was defined as a residue that contributes more than 2 kcal/mol to the total binding energy. Bogan et al. found out that only a small group of acidic (D), basic (K, R), aromatic (H, W, Y), and lipophilic residues (I, P) are statistically enriched in the hot spots. The authors assume that these residues suffer a lower entropic penalty due to short side chains (D or P) or fewer rotatable bonds (H, W, or Y) when placed in a PPI, compared to other lipophilic or ionic amino acids. Moreover, they observed that the hot spot residues are occluded by less relevant residues sealing the binding interface, the O-ring (Figure 22). The O-ring is presumed to prevent bulk solvent from protruding into the PPI and interacting with the hot spot residues.

While the O-ring theory provides a relatively narrow set of residues enriched in PPI hot spots, more recent analysis suggests that lipophilic amino acids, in general, are enriched in obligate protein-protein contacts.[187] This observation was included for the analysis of PPI architecture in addition to the O-ring theory.
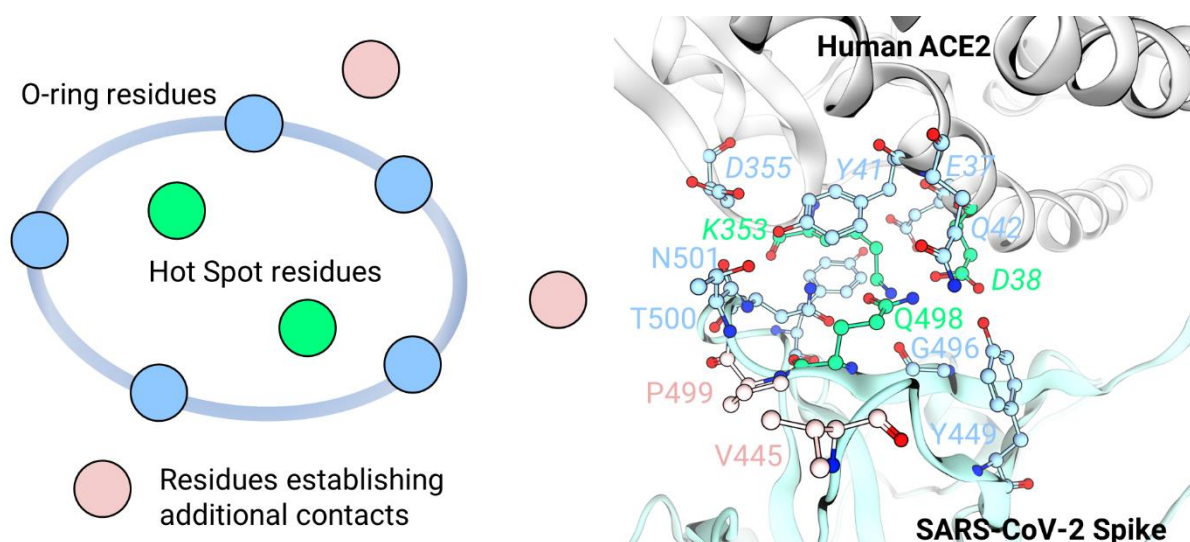
Figure 22. Scheme representing the O-ring theory (left) and a presumed O-ring[178] in the severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) Spike – human angiotensin-converting enzyme 2 (ACE2) binding interface (PDB-ID: 6M0J,[181] residues in non-italic: SARS-CoV-2 spike, residues in italic: human ACE2).

3. Results

In the following section, the computational work undertaken to understand the mechanisms behind virus-host cell interactions is described in detail. The biological and biochemical data supporting our models are shown and briefly described. In the first part (Chapter **3.1**), the development of small molecular NS2B-NS3 protease inhibitors against the emerging flavivirus family is presented. In the next chapter (**3.2**), a binding epitope of glycoprotein D from equine herpesvirus 1 and 4 to its receptor, equine MHC-I, is predicted. The last chapter (**3.3**) presents the structure-based models of ACE2 orthologs in their respective complexes with the spike protein used for the prediction of animal susceptibility to coronavirus disease 2019.

3.1   Development of Small Molecules Targeting NS2B-NS3 Proteases

3.1.1 Molecular Dynamics-Based 3D Pharmacophores Enable Identification of Potential WNV^{Pro} and ZIKV^{Pro} Inhibitors

In order to develop flaviviral inhibitors, we focused on the NS2B-NS3 protease as a promising target for directly acting antiviral drugs (DAAD).[11] We decided to target the substrate binding site of NS2B-NS3 since competitive inhibitors bear a higher resistance barrier than allosteric inhibitors.[188]

Due to a low number of reported inhibitors sharing a peptidomimetic character (Chapter 2.3.1 "Structures of Proteases", page 33), we aimed to identify novel inhibitors of the Zika virus protease (ZIKV^{Pro}) with a small molecular structure. To validate our models, high-quality inhibitor data was needed. We therefore searched for comparable NS2B-NS3 proteases in closely related flaviviruses. Several small-molecular competitive inhibitors are reported for the closely related West Nile virus protease (WNV^{Pro}), which shares high sequence similarity and identity with ZIKV^{Pro} (NS2B: 82% and 62%, respectively, NS3: 82% and 69%, respectively, Appendix Figure 2). This renders WNV^{Pro} a perfect candidate for comparative modeling of ZIKV^{Pro} inhibitors.

The substrate binding site of NS2B-NS3 is a challenging target for drug discovery, in literature even referred to as a *featureless binding site*.[189] The high hydrophilicity, shallowness,[163] and flexibility[162] render the structure-based inhibitor design highly complicated. To address these issues, we used a novel application called *PyRod*.[128]
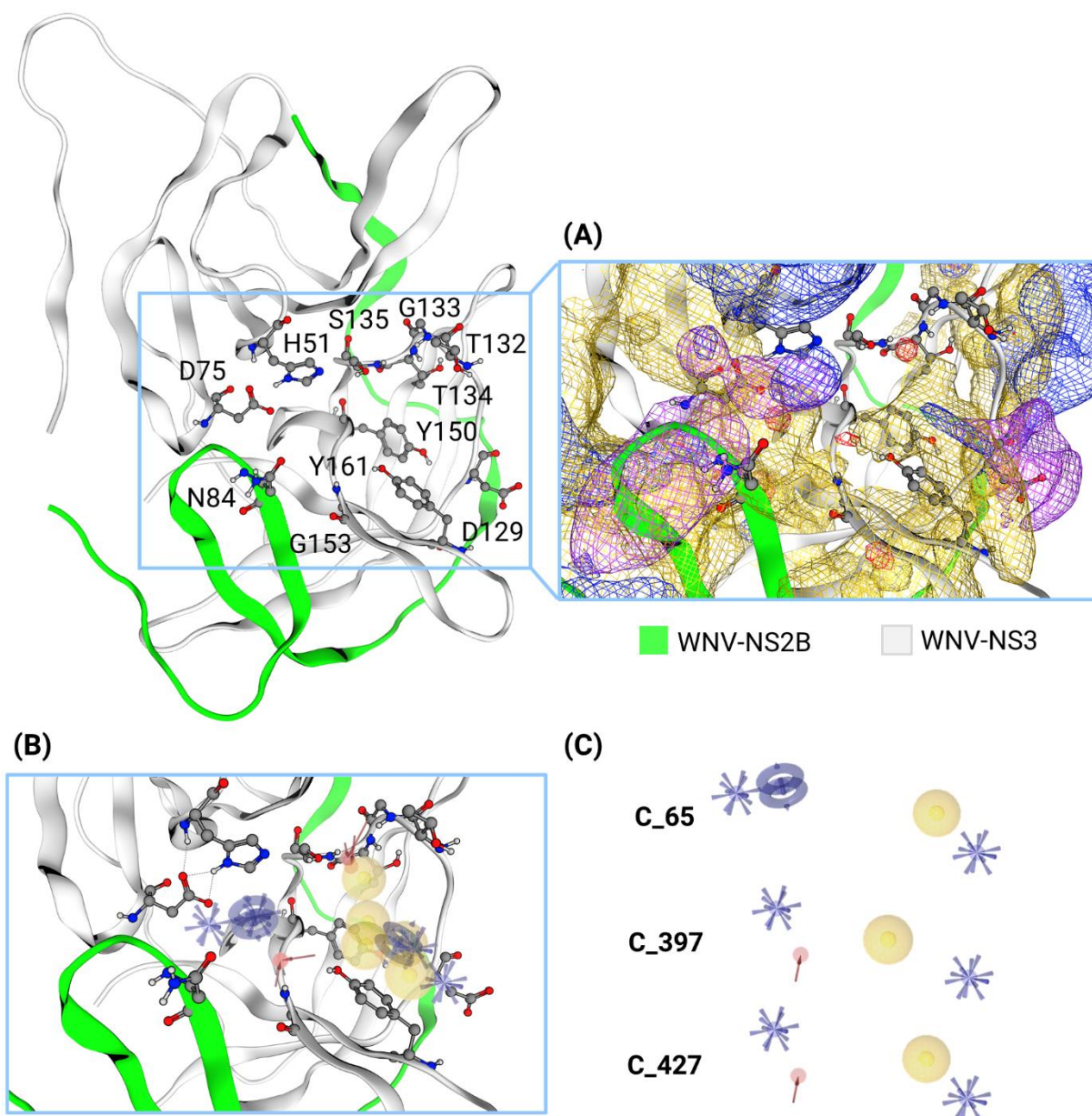
Figure 23. (A) Dynamic molecular interaction fields (dMIFs) generated with *PyRod* from molecular dynamics simulations of the apo-structure of the West Nile virus protease. (B) Generated focused pharmacophore **B** and (C) derived best-performing combinatorial pharmacophores (**C_65**, **C_397**, **C_427**). Color code: yellow clouds and spheres- lipophilic contacts, red clouds and arrows- hydrogen bond acceptors, blue clouds and rings- aromatic interactions, and purple clouds and stars- cationic contacts.

PyRod counters high flexibility of the target by using conformations obtained through molecular dynamics (MD) simulations. The software samples dynamic molecular interaction fields (dMIFs) by tracing water molecules in the protein environment. The

generated dMIFs (Figure 23A) are automatically translated into interaction points that can be applied to a virtual screening campaign.

Due to the large area of the binding pocket included in the workflow (1118 Å^2 as calculated with POVME3[190] as described in the Experimental section), the derived general 3D pharmacophore model contains too many interaction points covering a too large area to perform virtual screening. In order to rationally generate pharmacophore models suitable for a virtual screening campaign, we focused on the highly conserved S1 and S2 binding subpockets. This led to the focused pharmacophore **B1** (Figure 23B), which still contains too many interaction points for virtual screening. To reduce the number of interactions per pharmacophore, we prepared a combinatorial pharmacophore library using an integrated *PyRod* tool generating permutations of a given general pharmacophore. To limit the number of possible combinations, we decided to keep the crucial cationic interaction[191] to the NS3-residue D129 in all combinations. In addition, we restrained the number of interaction points to 3-6 per combination (with a maximum of 3 hydrophobic contacts, 4 hydrogen bonds, 3 aromatic interactions, and 3 cationic interactions in each combination). This resulted in a total of 3022 unique 3D pharmacophore models. In order to identify the best candidates for a virtual screening campaign, we validated all models using a manually curated set of active WNV[Pro] inhibitors and generated decoy molecules. The set of 17 active inhibitors[168, 169, 192-194] (with reported IC_{50} or K_i values lower than 50 µM, Appendix Table 1) was obtained by a manual literature search and the set of 667 decoys was generated using DUD-E server[129] based on the structure of active ligands.

All 3022 pharmacophore models were automatically screened against the active and decoy libraries. We chose three combinatorial pharmacophore models (**C_65**, **C_397**, **C_427**, Figure 23C) according to their superior performance characterized by their highest early enrichment factors ($EF_{1\%}$) and largest absolute number of recovered active inhibitors.

These three pharmacophore models were used for a virtual screening campaign of a database of 7.6 million commercially available compounds. The screening returned 1079 hits (10 for **C_65**, 712 for **C_397**, and 357 for **C_427**). In order to filter the most promising potential NS2B-NS3 inhibitors, we docked the obtained hits into the substrate binding site of the WNV[Pro] and obtained plausible binding poses. These poses were scored against the combinatorial pharmacophore models to retain

validated interaction patterns. In the next step, we visually inspected all selected hits to exclude implausible binding modes (such as lipophilic moieties pointing towards the solvent or ionized moieties not involved in a salt bridge) and non-drug-like moieties[132] (such as quinones or alkyl halides). The remaining 15 compounds in complex with the WNV[Pro] were simulated in a single 20 ns molecular dynamics (MD) simulation each, to ensure their capability of binding to the highly flexible NS2B-NS3 protease. We visually inspected all MD simulations, determining whether the ligand performs large movements. The last filtering step yielded five final hits showing no dissociation events in the MD simulations (Figure 24).
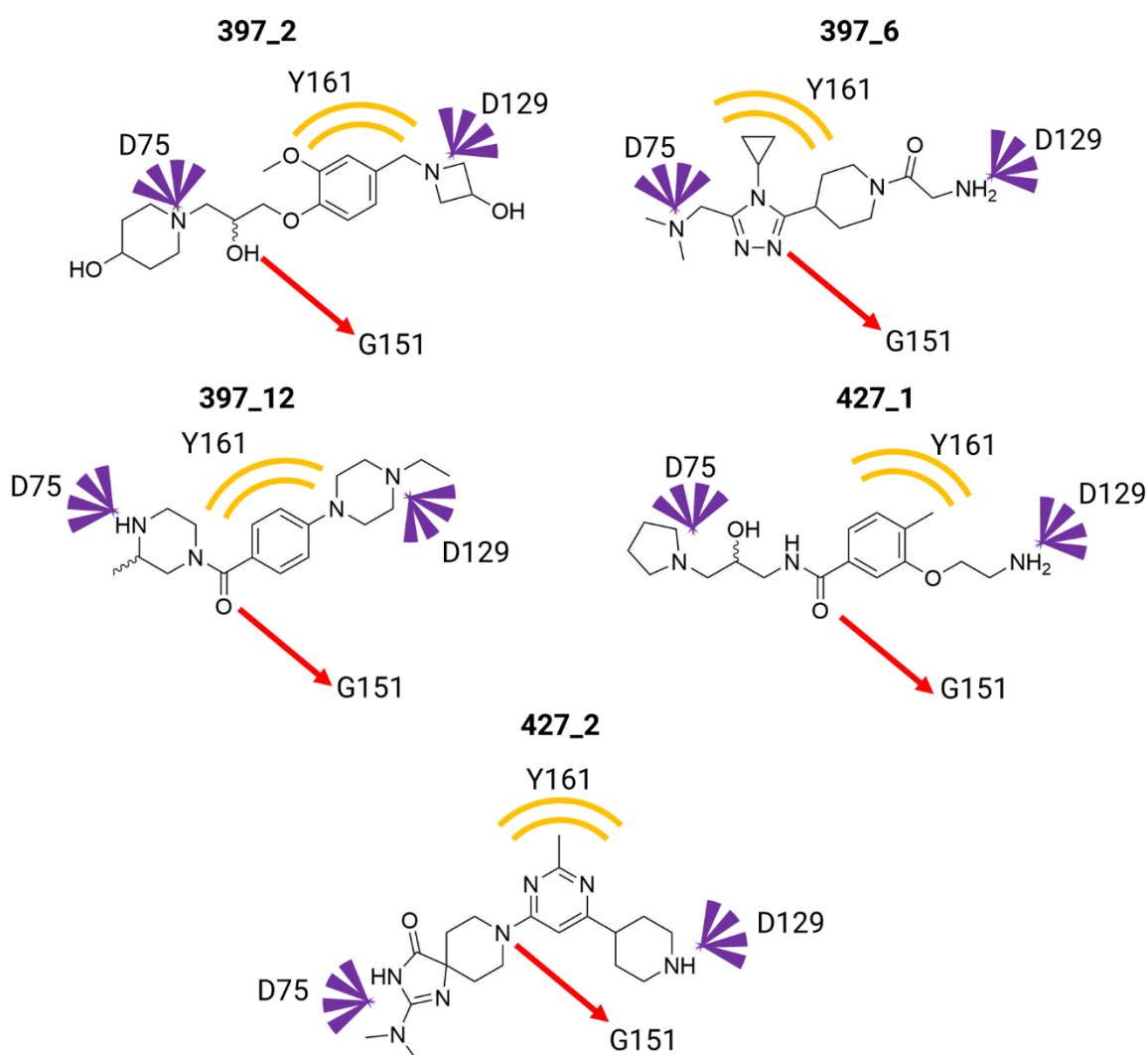
Figure 24. Potential flaviviral protease inhibitors obtained from virtual screening. Purple stars indicate cationic interactions, yellow lines lipophilic contacts, and red arrow hydrogen bonds with the protease according to the presented pharmacophore models.

In the next step, we aimed to confirm the ability of the identified ligands to inhibit ZIKV^{Pro}. The *PyRod* analysis of the substrate binding pocket of ZIKV^{Pro} resulted in the focused pharmacophore model **B2** (Figure 25, right). The pharmacophore shared substantial interaction points with the focused WNV^{Pro} pharmacophore **B1** (Figure 25, left), i.e. cationic interactions in the S1 and S2 subpockets, lipophilic interactions in the S1 subpocket, and hydrogen bond acceptor features in the polypeptide binding region and oxyanion hole. Due to the shared pharmacophoric properties, we predicted that our ligands would be active on both, ZIKV^{Pro} and WNV^{Pro}.

Figure 25. Comparison of the focused pharmacophores for West Nile virus protease (**B1**) and Zika virus protease (**B2**).

3.1.2 Biochemical Characterization of the Selected Compounds Confirms Inhibitory Activity

The selected hits were biochemically characterized on ZIKV^{Pro} by Tim M. Sarter in the Nitsche Lab in Canberra (Australia), and on WNV^{Pro} by Rafe Yousef in the Rademann Lab in Berlin (Germany). In the first step, we measured the remaining NS2B-NS3 activity under constant ligand concentration (500 µM for WNV^{Pro} and 100 µM for ZIKV^{Pro}). In the next step, the K$_i$ values were measured for the two most promising compounds **397_2** and **427_1** (Table 4, Appendix Figure 3).

3.1.3 Dynamic Pharmacophores Explain Determined Activity Cliffs

While the K$_i$ values for compound **397_2** remain in the same order of magnitude as predicted by our models, the activity of **427_1** shows an activity cliff of one order of

magnitude (25.5 µM on WNV^{Pro} and 2.3 µM on ZIKV^{Pro}). To explain this, we performed a series of 50 ns MD simulations of **427_1** in complex with ZIKV^{Pro} and WNV^{Pro}.

Table 4. Measured inhibitory activity of compounds selected from the virtual screening workflow.

Compound	K_i on DENV2^{Pro} [µM]	K_i on WNV^{Pro} [µM]	K_i on ZIKV^{Pro} [µM]
397_2	n.d.	7.4 ± 1.3	11.5 ± 0.5
397_6	n.d.	n.d.	n.d.
397_12	n.d.	n.d.	n.d.
427_1	0.09 ± 0.03	25.5 ± 11.8	2.3 ± 0.4
427_2	n.d.	n.d.	n.d.

n.d.: not determined; K_i values were only determined for compounds within a cutoff below 50 µM. Abbreviations: DENV2- Dengue 2 virus, Pro- protease, WNV- West Nile virus, ZIKV- Zika virus.

We analyzed the interaction patterns occurring over the course of the MD simulations using the *Dynophore* application.[73, 127] We observed that the inhibitor can adopt two binding modes in the binding pocket of ZIKV^{Pro} (Figure 26). In the alternative binding mode, **427_1** adopts a conformation occupying S2 and S4 subpockets. The primary amino group of **427_1** migrates from the S1 subpocket and loses the ionic contact to NS3-D129 in favor of a salt bridge to NS2B-D83 in the S4 subpocket. Furthermore, the inhibitor establishes new lipophilic contacts to NS3-V155 in addition to the lipophilic contacts to NS3-Y161 also present in the initial binding mode. We suggest that the additional interactions seen in the second binding mode provide an entropic gain, explaining the improved activity of **427_1** on ZIKV^{Pro}.

The second binding mode of **427_1** can be also observed in the MD simulations of the inhibitor in complex with WNV^{Pro}. Due to the polymorphic mutation in the NS2B-unit (N84 in WNV^{Pro} versus D83 in ZIKV^{Pro}), no stabilizing salt bridge to the primary amino group in the S4 subpocket can be observed. In a single MD simulation, **427_1** leaves the binding pocket of WNV^{Pro}, which may correlate with a lower activity of the inhibitor on this protease.
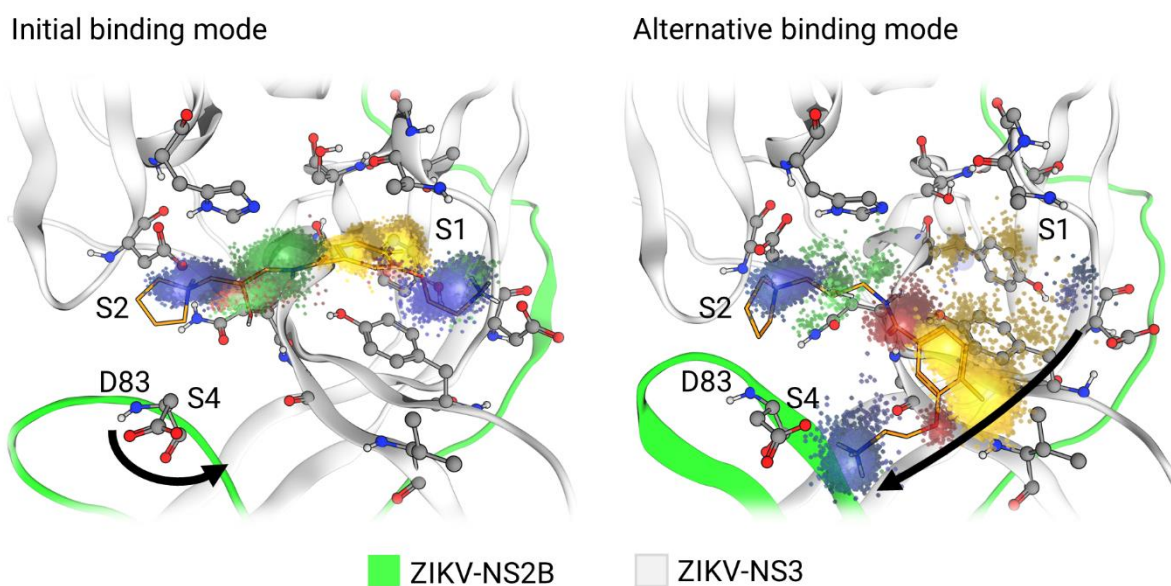
Figure 26. Representative spatiotemporal interaction clouds (*Dynophores*) found in molecular dynamics simulations of **427_1** (stick model) in complex with Zika virus protease (ribbon, sphere and stick model). Color code: orange sticks- carbon atoms of **427_1**, grey spheres- protein carbon atoms, blue sticks and spheres- nitrogen atoms, red sticks and spheres- oxygen atoms, pale grey sticks and spheres- hydrogen atoms, green ribbon- NS2B, white ribbon- NS3, purple point-clouds- cationic interactions, red or green point clouds- hydrogen bond acceptors or donors, respectively, yellow point clouds- lipophilic contacts.

3.1.4 PyRod Predicts the Activity of Our WNV^Pro and ZIKV^Pro Inhibitors on Closely Related DENV2^Pro

After obtaining encouraging results at WNV^Pro and ZIKV^Pro, we investigated potential inhibitory activity on another test system available in the Rademann Lab: Dengue 2 virus protease (DENV2^Pro). Since no experimentally determined crystal structure of the DENV2^Pro in the closed conformation was available, we developed a homology model using MOE (Chemical Computing Group ULC, Montreal, Canada). We chose the DENV3^Pro X-ray structure (PDB-ID: 3U1J) as a template due to its high sequence similarity and identity (77% similarity and 54% identity for NS2B, 86% similarity and 71% identity for NS3, Appendix Figure 2). The obtained DENV2^Pro model shows no Ramachandran outliers[89] (Appendix Figure 4), suggesting a plausible backbone geometry. The substrate binding pocket of the DENV2^Pro model contains all typical features of the flaviviral proteases: the catalytic triad consisting of conserved residues NS3-D75, H51, and S135, and the polypeptide binding site (NS3-G151, G153). This

suggests that the obtained homology model of DENV2^{Pro} is suitable for the modeling of competitive inhibitors.

In the next step, we performed MD simulations of DENV2^{Pro} in an aqueous environment sampling the substrate binding pocket using *PyRod*. The obtained focused pharmacophore **B3** (Figure 27, right) is similar to the focused pharmacophore **B1** of WNV^{Pro} (Figure 27, left). This suggests that the identified ligands could also inhibit DENV2^{Pro}.
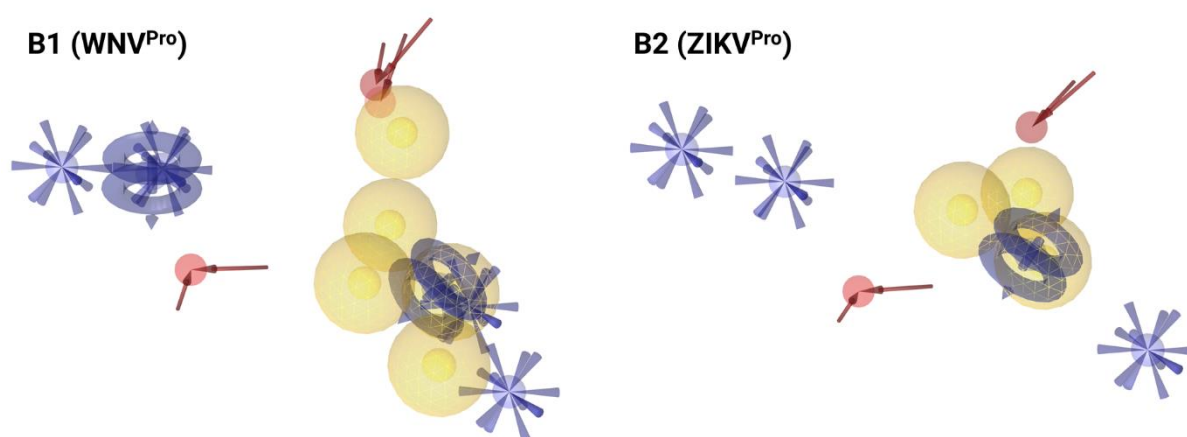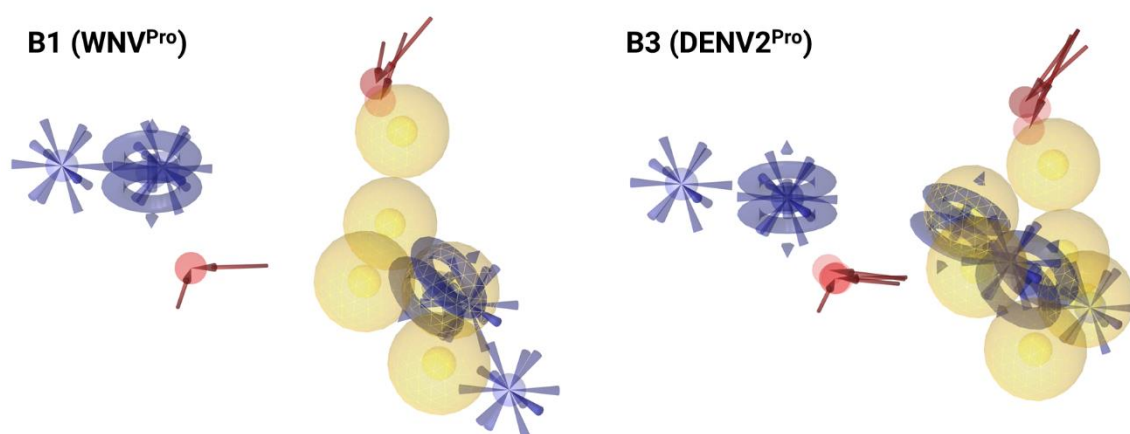

Figure 27. Comparison of the focused pharmacophores for West Nile virus protease (**B1**) and Dengue 2 virus protease (**B3**).

The two most active compounds, **397_2** and **427_1**, were docked into the DENV2^{Pro} model and the complexes were simulated to obtain dynamic pharmacophores. Visual inspection of the obtained trajectories of **427_1** in complex with DENV2^{Pro} revealed that the NS2B-loop of the substrate binding pocket (Figure 28) adopts a conformation featuring a larger distance to the NS3 subunit than observed in WNV^{Pro} and ZIKV^{Pro}. This observation is in line with the findings reported by Christiane Schüler in her master thesis, performed in our group at an earlier date.[195] In this study, it was shown that a single polymorphism, NS2B-F85M (WNV^{Pro} – DENV2^{Pro}), might be responsible for the destabilization of NS2B-NS3 contacts, allowing the DENV2-NS2B-loop to strongly fluctuate in the MD simulations. The extent of the NS2B-loop fluctuation was characterized in this thesis by measuring the distance between the backbone Cα atom of the NS2B-residue 85 in DENV2^{Pro} (M85) and WNV^{Pro} (F85) and the backbone Cα atom of NS3-G153 in DENV2^{Pro} and WNV^{Pro}. We applied this distance descriptor to characterize the NS2B-loop – NS3 core distance fluctuation in our simulations (Figure

28). The measured distribution of the distance between NS2B-residue 85 and NS3-G153 confirmed that the NS2B-loop in the **427_1**-DENV2^{Pro} complex is likely to adopt a larger distance to the NS3-core than in WNV^{Pro} and ZIKV^{Pro}, allowing the S2 subpocket of DENV2^{Pro} to be wider in comparison. We surmised that the positively charged N atom in the pyrrolidine moiety of **427_1** can establish closer ionic contacts to NS3-D75. To demonstrate this, we analyzed the distance between the positively charged N atom in the pyrrolidine moiety of the inhibitor to the carboxylate of DENV2^{Pro} NS3-D75 obtained from the dynamic pharmacophores of **427_1**. Indeed, the distance distribution shows a higher probability of closer ionic contact between the pyrrolidine moiety and NS3-D75 in DENV2^{Pro} than in WNV^{Pro} and ZIKV^{Pro} (Figure 28).

The analysis of the MD trajectories of **397_2** in complex with DENV2^{Pro} did not reveal any differences compared to WNV^{Pro} or ZIKV^{Pro}. Based on the analysis of the trajectories, we predicted **427_1** to be more active at DENV2^{Pro} than on ZIKV^{Pro} and **397_2** to be active in the same order of magnitude as on WNV^{Pro} and ZIKV^{Pro}.

In the next step, all final hits were biochemically characterized on DENV2^{Pro} by Silke Bergmann and Christoph Arkona in the Rademann Lab, following a similar protocol to that for the WNV^{Pro} assay. According to our prediction, the most active compound **427_1** showed higher inhibitory activity on DENV2^{Pro} than on ZIKV^{Pro} (Table 4). Surprisingly, inhibitor **397_2** was detected as inactive in the assay. To explain this surprising result, we analyzed the DENV2^{Pro} dMIFs obtained from *PyRod*. We observed that the S1 subpocket of DENV2^{Pro} is more lipophilic than in ZIKV^{Pro} and WNV^{Pro} (Figure 27). The comparison of **397_2** and **427_1** in complex with DENV2^{Pro} reveals an important difference: inhibitor **397_2** exposes a methoxy moiety in the S1 subpocket compared to a more lipophilic methyl group of **427_1**. The reduced lipophilicity of the methoxy group in **397_2** might contribute to the lower activity of the inhibitor.

The two most promising candidates (**397_2** and **427_1**) were selected for *in-vivo* testing. The DENV2^{Pro}-inhibition cell-based assays were performed by Mila Leuthold in Heidelberg (Germany). Briefly, the novel assay is performed on NS2B-NS3-expressing HeLa-cells with a luciferase reporting system.[196] When active, the protease cleaves luciferase from the flaviviral polyprotein and enables the generation of a fluorescent reagent, luciferin. The measured EC$_{50}$ values for **397_2** (47.8 μM) and **427_1** (26.8 μM) suggest cellular activity of the DENV2^{Pro} inhibitors (Appendix Figure 5).
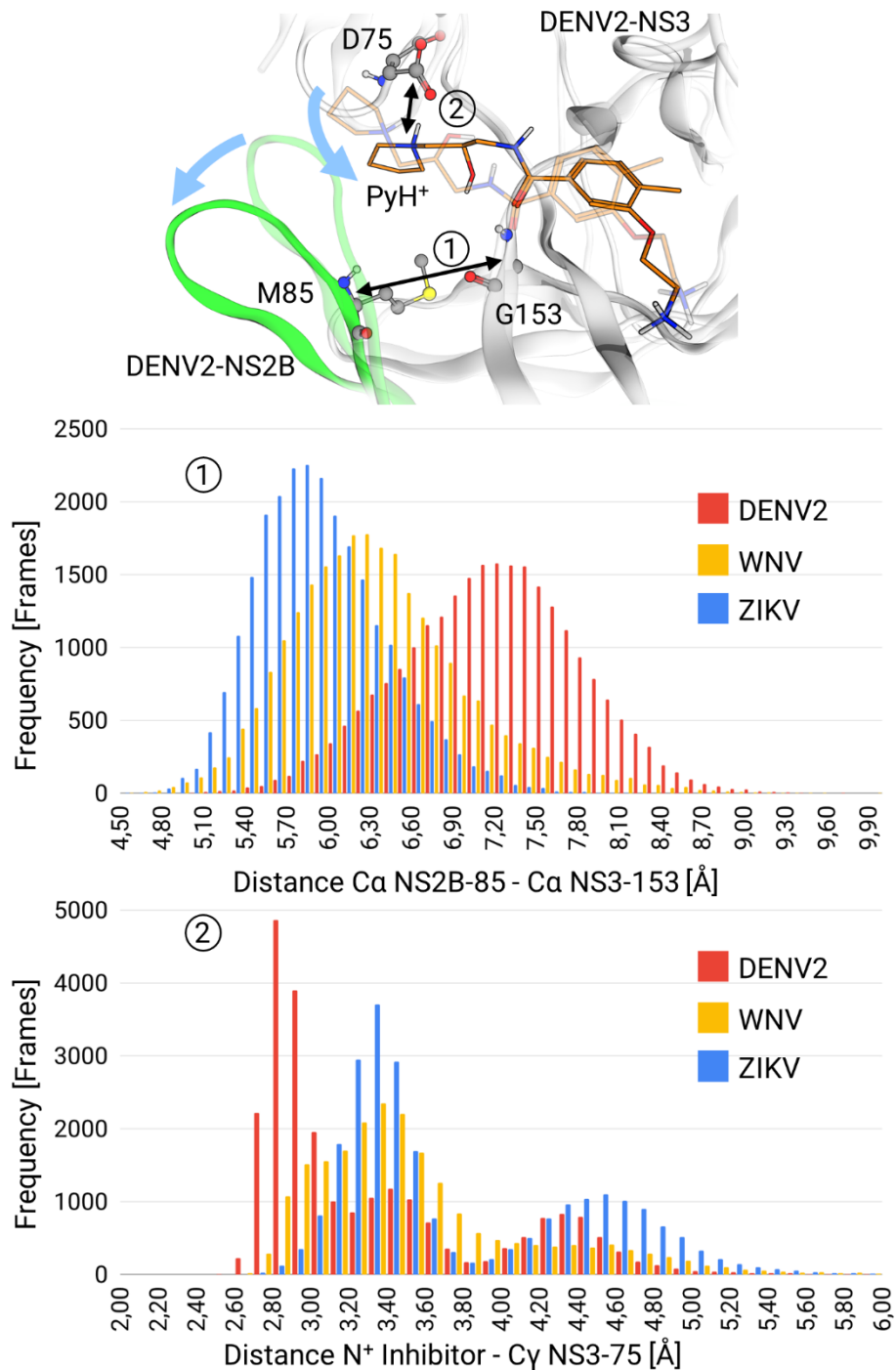
Figure 28. Distribution of the backbone-distances of the -NS2B-loop and NS3-core in Dengue 2, West Nile, and Zika virus proteases (upper histogram, 1). Distance distribution of the positively charged nitrogen atom in the pyrrolidine moiety of **427_1** and the carboxylate of NS3-D75 in Dengue 2, West Nile, and Zika virus proteases (lower histogram, 2).

3.2 Equine Herpesvirus: Targeting the MHC-I − Glycoprotein D Interface for Vaccine Development

As described in the Theoretical Background section (Chapter 2.1.2 "The Viral Replication Cycle", page 13), glycoproteins D (gD) from equine herpesvirus 1 and 4 (EHV-1, -4) recognize the major histocompatibility complex I (MHC-I) on host cells. The gD − MHC-I binding represents the crucial step of EHV entry. Due to the importance of EHV-gD for the replication cycle, gD is a promising starting point for the development of novel vaccines. While the crystal structures of EHV-1 and 4 gD are available, the specific binding region of EHV gD to MHC-I is unknown. The aim of this thesis section is the development of plausible gD MHC-I complex binding geometries in order to identify plausible immunogenic epitopes.

3.2.1 Crystal Structures of Equine MHC-I Lack Pivotal A173 Leading to the Construction of an MHC-I Homology Model

The EHV-1 and -4 enter cells via the major histocompatibility complex I (MHC-I) that can be found in 27 variants (reported in the UniProt database) classified into 18 characterized genotypes (Appendix Table 2).[34] At the beginning of this project, only scarce data on gD-MHC-I binding was available. Sasaki et al. have described the importance of MHC-I-A173 for the EHV-1 gD using a gain-of-function MHC-I-mutant Q173A.[197] Azab and colleagues extended these experiments to several MHC-I genotypes and EHV-4, confirming the pivotal role of MHC-I-A173 (Figure 29).[34] Surprisingly, genotypes 3.2 and 3.7 were inert to EHV-1-infection despite harboring an alanine at MHC-I-position 173.

With this knowledge, we manually inspected all available equine MHC-I crystal structures (PDB-ID: 4ZUS, 4ZUT, 4ZUU, 4ZUV, 4ZUW[198]) to find one featuring EHV entry. Unfortunately, all the structures originate from the same genotype 1.18.7-6, which harbors a glutamic acid residue at position 173. This suggests a negligible binding to gD. Hence, we carefully developed a homology model of MHC-I genotype 3.1 as used in the experimental setup by our collaborators, Viviane Kremiling and Walid Azab from Osterrieder Lab in Berlin (Germany). Genotype 3.1 reportedly supports EHV entry[34] and features an alanine residue at position A173. The homology model was developed based on the template with the best resolution (PDB-ID: 4ZUU[198]).

The MHC-I genotype 3.1 shares 85% identity and 88% similarity in the heavy chain (Appendix Figure 6) with the template genotype 1.18.7-6. This high sequence similarity allows construction of a high-quality homology model of the heavy chain of MHC-I. Subsequently, we built a homology model of the light chain of MHC-I called β_2-microglobulin. The X-ray crystal structure contains a murine β_2-microglobulin sharing 63% identity and 82% similarity with its equine ortholog, allowing the construction of a high-quality model.

The obtained homology models of the heavy and light chains featured no Ramachandran outliers[89] (Appendix Figure 7). Furthermore, the positions of the disulfide bridges characteristic for the MHC-I fold at positions 101-164, 203-259 (heavy chain), and 25-80 (light chain) were present and the geometry of the sulfur-sulfur bonds was correct. These characteristics suggest a high geometric quality of the obtained homology models. Both chains of the homology models were subsequently assembled to a full-length equine MHC-I genotype 3.1 structure. The assembled model was relaxed in a molecular dynamics (MD) simulation as described in the Experimental section. The coordinates for docking of the relaxed model were obtained after the root mean square deviation (RMSD) of the backbone had reached a stable plateau of 3 Å (Appendix Figure 8).

In its physiological form, the MHC-I carries a short peptide 8-12 residues in length in a groove between the α-helices (Figure 29). Our collaborators used a nonapeptide (SDYVKVSNI) in their cell-based infection models, which is different from the nonapeptide co-crystallized in the template structure 4ZUU (CTSEEMNAF). Hence, we decided to model the nonapeptide SDYVKVSNI into our MHC-I genotype 3.1 homology model.

Only scarce data on the interactions between MHC-I and peptide are available. The studies conducted by Yao and colleagues indicate that second and third residues as well as the C-terminus specifically establish interactions with MHC-I.[198] The conformation of the complexed peptide is strongly dependent on its length.[198] Hence, we manually fitted the SDYVKVSNI into the MHC-I homology model using the conformation of the co-crystallized nonapeptide CTSEEMNAF (Figure 29).

Finally, to identify the gD binding epitope, we performed protein-protein docking using our MHC-I homology model and the EHV-1 gD crystal structure obtained by our collaborators (PDB-ID: 6SQJ[183]). Since the complexed nonapeptide might interfere

with a potential binding epitope due to its proximity to the residue A173, we decided to perform two docking experiments, one with and the other without the nonapeptide. Finally, the resulting binding modes were compared to evaluate the influence of the nonapeptide on gD binding.
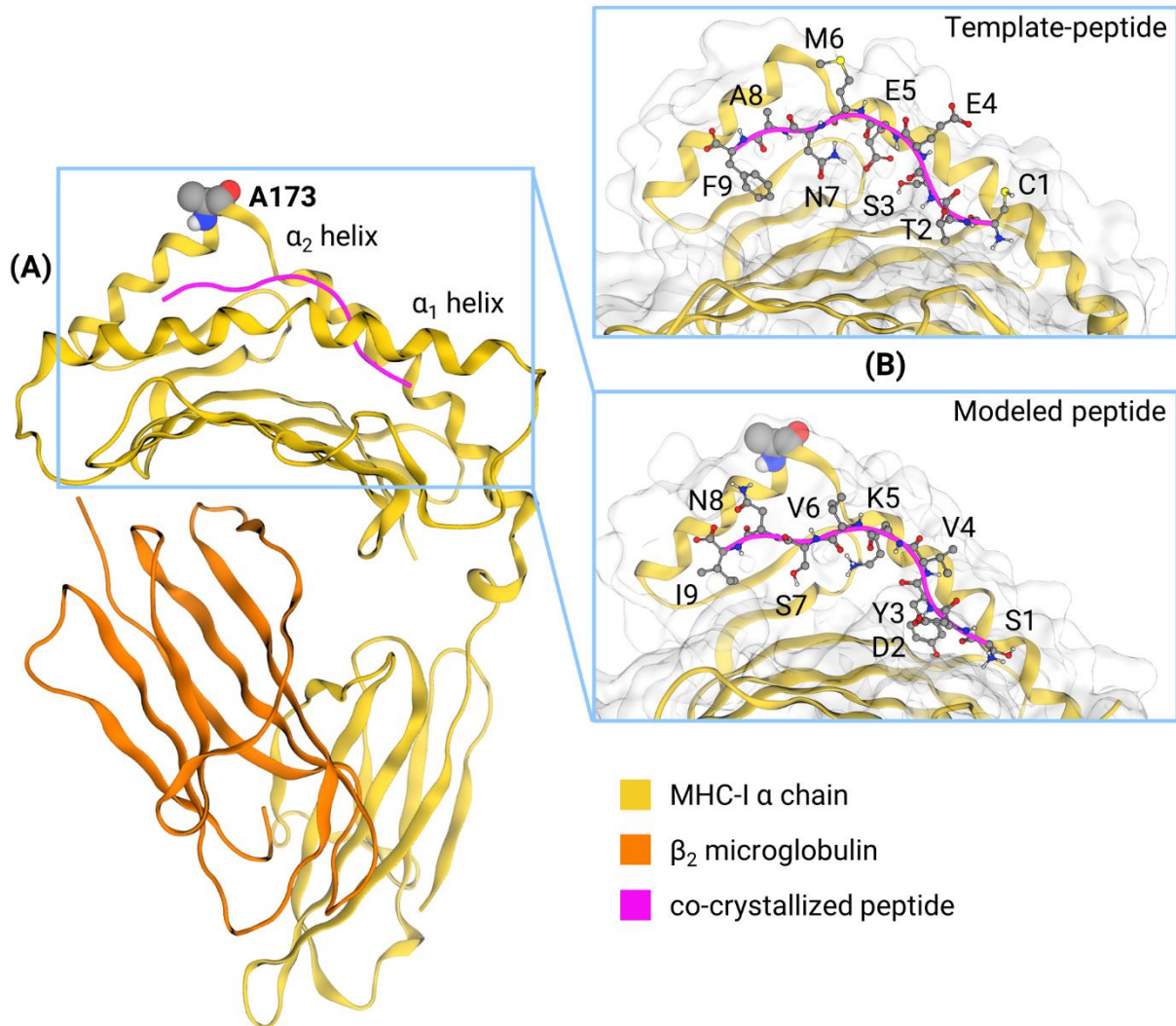
Figure 29. (A) Global structure of the equine major histocompatibility complex I (MHC-I) homology model with indicated A173, a residue crucial for equine herpesvirus entry. (B) The conformation of the co-crystallized template-peptide CTSEEMNAF (top) and modeled peptide SDYVKVSNI used in the cell-based infection model (bottom); for the sake of clarity, the α1 helix is not shown. Color code: golden ribbon – MHC-I α chain, orange ribbon- β2 microglobulin, and magenta ribbon- peptide.

3.2.2 Docking of EHV-1 gD onto the Peptide-Free MHC-I Model Reveals Proximity of the gD-Binding Interface to the Peptide-Binding Region

Both docking experiments, peptide-free and peptide-bound, were performed using Rosetta by generating 10,000 docking runs. The peptide-free docking round of EHV-1 gD on the MHC-I homology model yielded around 7,200 docking solutions. According to the Rosetta protocol,[108] a collection of recommendations for the usage of Rosetta, the most energetically favored binding modes are scored with the lowest value of the *Rosetta energy*, a surrogate parameter for the free energy of binding. We picked the first ten best docking poses and applied heuristic filtering rules (Table 5) to choose the most plausible binding modes.

Table 5. Heuristic filtering rules to find the most plausible docking pose.

Filtering Rule	Rationale
Docking poses harboring the C-terminus of gD in the PPI are discarded.	The C-terminal part of the gD is anchored in the viral envelope.[199]
Docking poses without lipophilic contacts in the PPI are discarded.	Residues with lipophilic side chains are enriched in PPI.[187]
Docking poses containing residues D, H, I, K, P, R, W, Y in the PPI are accepted.	According to the O-ring theory,[186] these residues are statistically enriched in PPI.

Abbreviations: gD- glycoprotein D, PPI- protein-protein interfaces.

The four selected docking poses were prepared and simulated (see Experimental Details, Chapter 6.2 "Equine Herpesvirus: Targeting the MHC-I – Glycoprotein D Interface for Vaccine Development") to examine the stability of the binding interfaces.[105] Only one docking solution shows an RMSD of 6 Å during the MD simulations (Appendix Figure 9), indicating negligible global conformational changes throughout the trajectory.

To characterize the obtained protein-protein interface (PPI), we analyzed steric contacts between the MHC-I and EHV-1 gD using the *PyContact* application (Table 6).[200] The role of all residues involved in the EHV-1 gD – MHC-I contact was assessed based on the O-ring theory[186] described in the Theoretical Background section (Chapter 2.5 "Classification and Architecture of Protein-Protein Interactions").

The protein-protein interface (PPI) displays two main residue patches (Figure 30): The first patch consists of residues building three loops of the gD-core (loop 1: T161,

loop 2: I211, F213, and loop 3: W257, D261) and the α2-chain of MHC-I (Q165, I166, R168, R169). The second patch contains a single N-terminal R43 of gD contacting Y108 placed in the α1-chain of MHC-I.

Figure 30. The binding hypothesis of equine herpesvirus 1 glycoprotein D (EHV-1 gD) to the equine major histocompatibility complex I (MHC-I) genotype 3.1 in its peptide-free form. Color code: green ribbon and pale gray spheres- gD, golden, orange ribbon, and grey spheres- MHC-I.

The binding hypothesis confirms our assumption of the proximity of the hypothetical binding epitope to the peptide-binding groove of MHC-I. Hence, we performed an additional docking round on the relaxed equine MHC-I genotype 3.1 homology model with the modeled peptide of sequence SDYVKVSNI.

Table 6. Frequencies of contacts between the EHV-1 gD and equine MHC-I genotype 3.1 in a peptide-free binding hypothesis.

MHC-I residue	Contact frequency [%]	EHV-1 gD residue	Hypothesized role
Y108	90	R43	O-ring sealing I166-W257 contact
Q165	100	T161	O-ring sealing I166-W257, R169-D261 contacts
I166	**100**	**W257**	**Hot spot contact**
R168	98	F213	O-ring sealing I166-W257, R169-D261 contacts
	100	I211 (backbone)	Additional contact stabilizing the PPI
	100	F213	Additional contact stabilizing the PPI
R169	100	W257	O-ring sealing R169-D261 contact
	100	**D261**	**Hot spot contact**

Abbreviations: EHV-1- equine herpesvirus 1, gD- glycoprotein D, MHC-I- major histocompatibility complex I, PPI- protein-protein interface.

3.2.3 Docking of EHV-1 and 4 gD onto the Peptide-Bound MHC-I Model Yields Plausible Binding Modes

Protein-protein docking was performed using the Rosetta application followed by an MD simulation-based filtering as described for the peptide-free docking round in the previous section 3.2.2 on page 57. The docking yielded around 7,100 possible binding geometries. Among those, we searched for docking solutions containing the crucial contact gD-D261 – MHC-I-R169, indicated as a distance of 4.5 Å between Cγ atom of gD-D261 and Cζ atom of MHC-I-R169, using an MDAnalysis-based[201, 202] script. The script was kindly provided by David Machalz from our group.

The filtering revealed five docking solutions containing the steric proximity we searched for. Visual inspection and O-ring theory-based filtering (Table 5) yielded a single docking pose. The selected complex belongs to the top 3% of all docking poses according to the Rosetta energy score, suggesting a low-energy binding geometry (Figure 31). The peptide-bound docking pose of EHV-1 gD in complex with MHC-I was

simulated in a 100 ns MD simulation and the PPI was characterized with *PyContact* (Table 7), as described for the peptide-free docking. The complex showed a stable RMSD-plateau at 3.5 Å (Appendix Figure 10) suggesting negligible global conformational changes throughout the MD simulation.
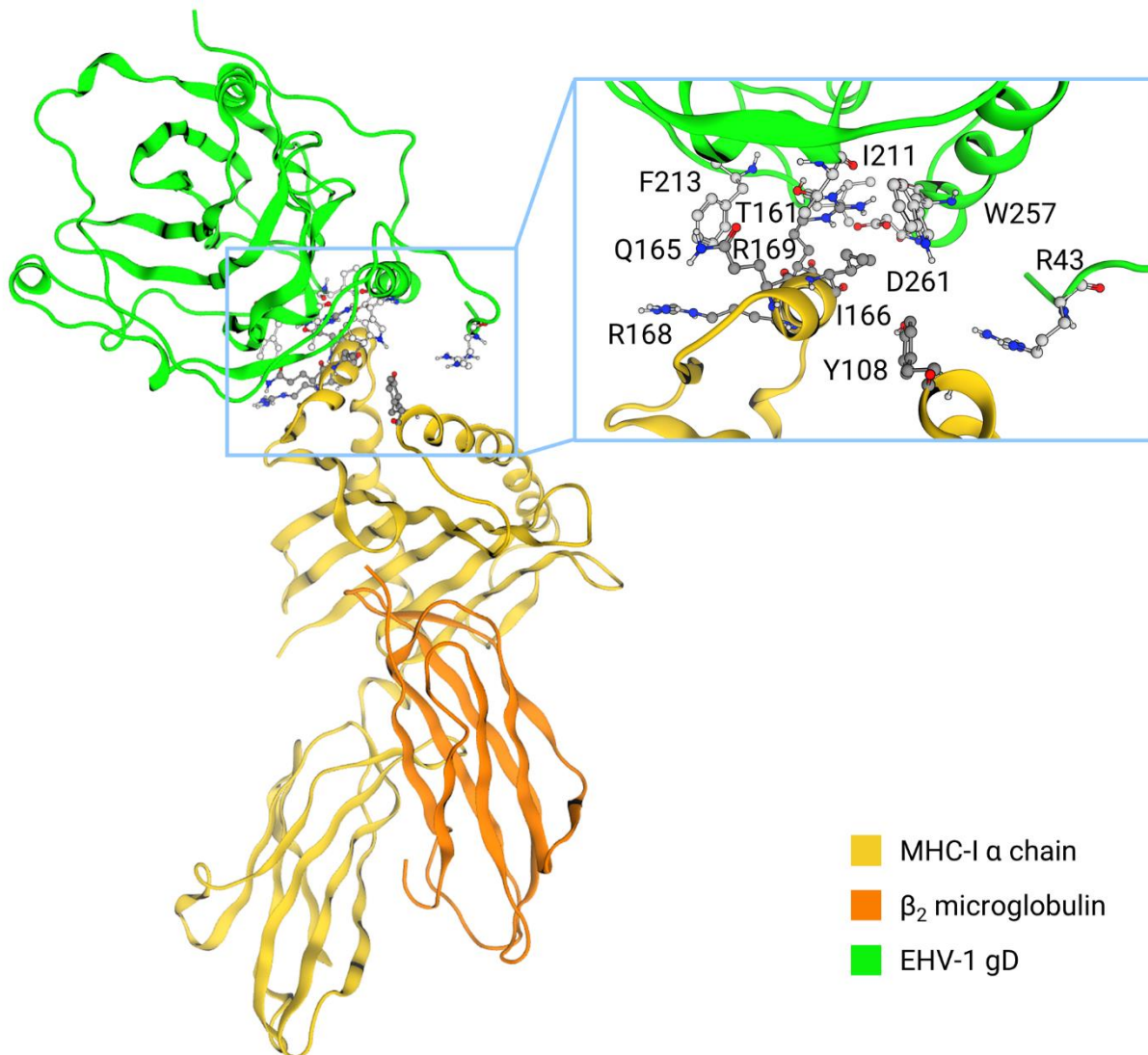

Figure 31. The binding hypothesis of equine herpesvirus 1 glycoprotein D (EHV-1 gD) to the equine major histocompatibility complex I (MHC-I) genotype 3.1 in its peptide-bound form. Color code: green ribbon and pale gray spheres- gD, golden, orange ribbon, grey spheres- MHC-I, and magenta ribbon- peptide.

Table 7. The frequencies of contacts between the EHV-1 gD and equine MHC-I genotype 3.1 in the peptide-bound binding hypothesis.

MHC-I residue	Contact frequency [%][a]	EHV-1 gD residue	Assumed role
R103	100 ± 0	E242	Additional contact stabilizing the PPI
Y108	100 ± 0[b]	A157	O-ring sealing I166-F213 contact
N110	92 ± 5	R222	O-ring sealing I166- F213 contact
E113	98 ± 3	R238	Additional contact stabilizing the PPI
I166	**97 ± 4**	**F213**	**Hot spot contact**
R169	**100 ± 0**	**D261**	**Hot spot contact**
A173	97 ± 1	W257	O-ring sealing R169-D261 contact

[a] Molecular dynamics simulation was performed in triplicate, the value is given as average ± standard deviation. [b] The frequency of MHC-I-Y108 − EHV-1 gD-A157 summarizes lipophilic side chain contacts and hydrogen bonding between the Y108-OH group and the A157 backbone. Abbreviations: EHV-1-equine herpesvirus 1, gD- glycoprotein D, MHC-I- major histocompatibility complex I, PPI- protein-protein interface.

The EHV-1 gD − MHC-I binding hypothesis in the peptide-bound form bears three residue patches (Figure 31):

(i)   the hot spot patch containing MHC-I residues placed in the α2-chain (I166, R169, and A173) and gD residues F213, W257, and D261,

(ii)   a small lipophilic patch with MHC-I residue Y108 of the α1-chain involved in the coordination of the nonapeptide and contacting gD residue A157,

(iii)   hydrogen bonding patch containing MHC-I residues placed in the α1-chain (R103) and loops (N110 and E113) contacting gD residues R222, R238, and E242.

Interestingly, this binding hypothesis may explain why residues with larger side chains than alanine at position 173 of MHC-I hinder viral entry. MHC-I-A173 seals the crucial salt bridge MHC-I-R169 − gD-D261 via a lipophilic contact to gD-W257. Larger side chains at the MHC-I-position 173, such as threonine in genotype 3.4, glutamic acid in

3.5, or valine in 3.6 may represent a steric hindrance and prevent the sealing of the hot spot contact. Additionally, our binding hypothesis explains why the MHC-I genotype 3.7 does not allow viral entry despite having an alanine at position 173. The MHC-I genotype 3.7 bears residue E174 (instead of A174 as in genotype 3.2) that points towards the lipophilic gD-residues W253 and W257 in our model. This potentially leads to an entropic penalty, preventing the sealing of the hot spot salt bridge.

In the next step, we aimed to establish an EHV-4 gD binding hypothesis. At this time, our collaborators had successfully solved the crystal structure of EHV-4 gD (PDB-ID: 6TM8[183]), which was subsequently docked onto peptide-bound MHC-I. Due to the homology of EHV-1 and -4 gD's, we expected to find an EHV-4 gD binding pose exploiting identical patches to those in EHV-1 gD. The protein-protein docking yielded approximately 7,100 poses. The subsequent search for the crucial gD-D261 – MHC-I-R169 contact yielded 21 complexes. Visual inspection and O-ring-based filtering (Table 5) returned a single binding hypothesis (Figure 32) belonging to the top 11% docking poses, according to the Rosetta energy score. The peptide-bound EGV-4 gD – MHC-I complex was subsequently simulated in a 100 ns MD simulation and the PPI was characterized again using *PyContact* (Table 8). The complex showed a stable RMSD-plateau at 6.5 Å (Appendix Figure 11) during the whole simulation, suggesting low conformational changes.

The EHV-4 gD epitope contains the same residue patches as EHV-1 gD, confirming our hypothesis of homologous binding sites. We observed that the MHC-I-R103 – gD-E242 contact frequently present in the EHV-1 complex was not present in the EHV-4 gD binding hypothesis. The lower number of interactions in the EHV-4 gD – MHC-I interface might explain the lower binding affinity of EHV-4 gD (K_D= 4413 μM) compared to EHV-1 gD (K_D= 3996 μM, as measured by Vivane Kremling).[183]

Due to the peptide binding by MHC-I in its physiological state, we suggest the peptide-bound gD-MHC-I docking poses as our final and valid models (Figure 32). Interestingly, no contacts between the peptide and gD were observed, suggesting that the peptide sequence might not influence the gD binding.
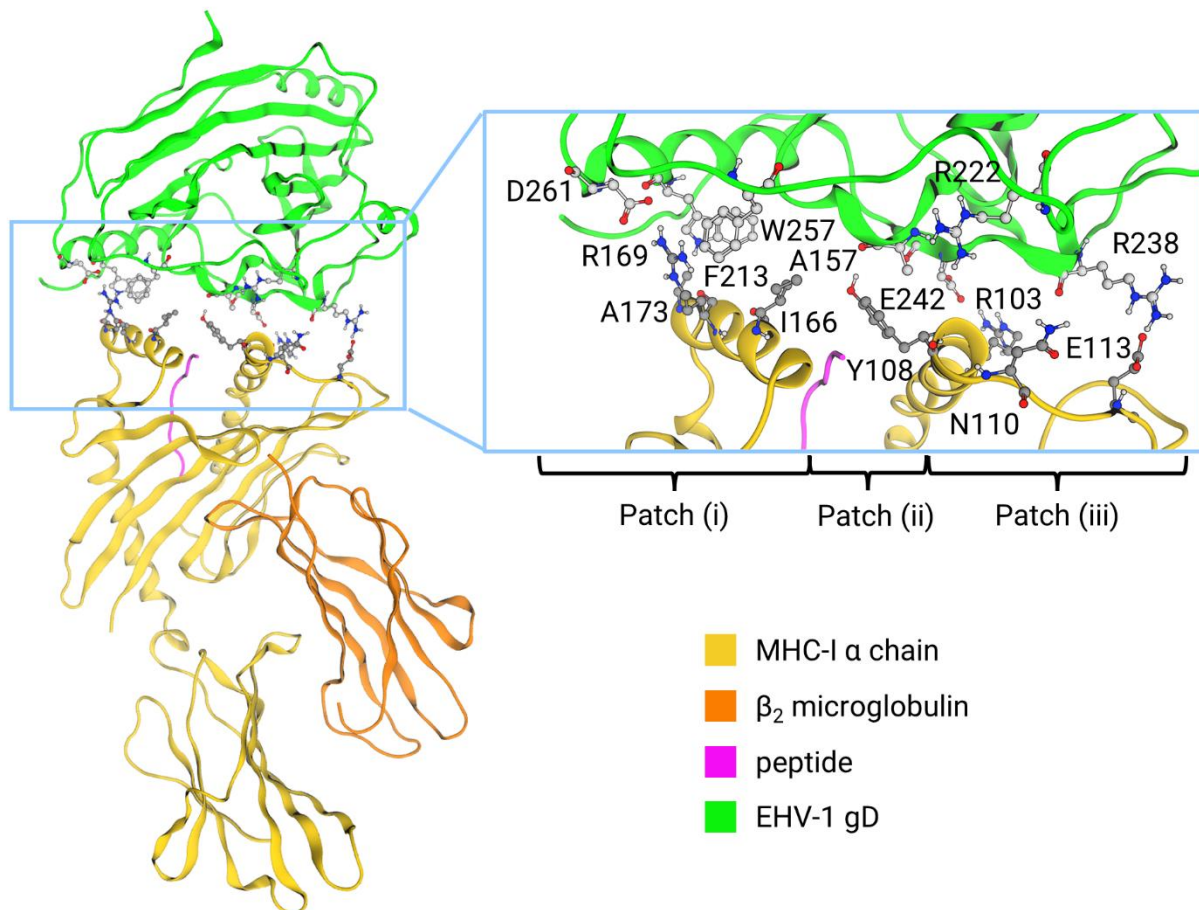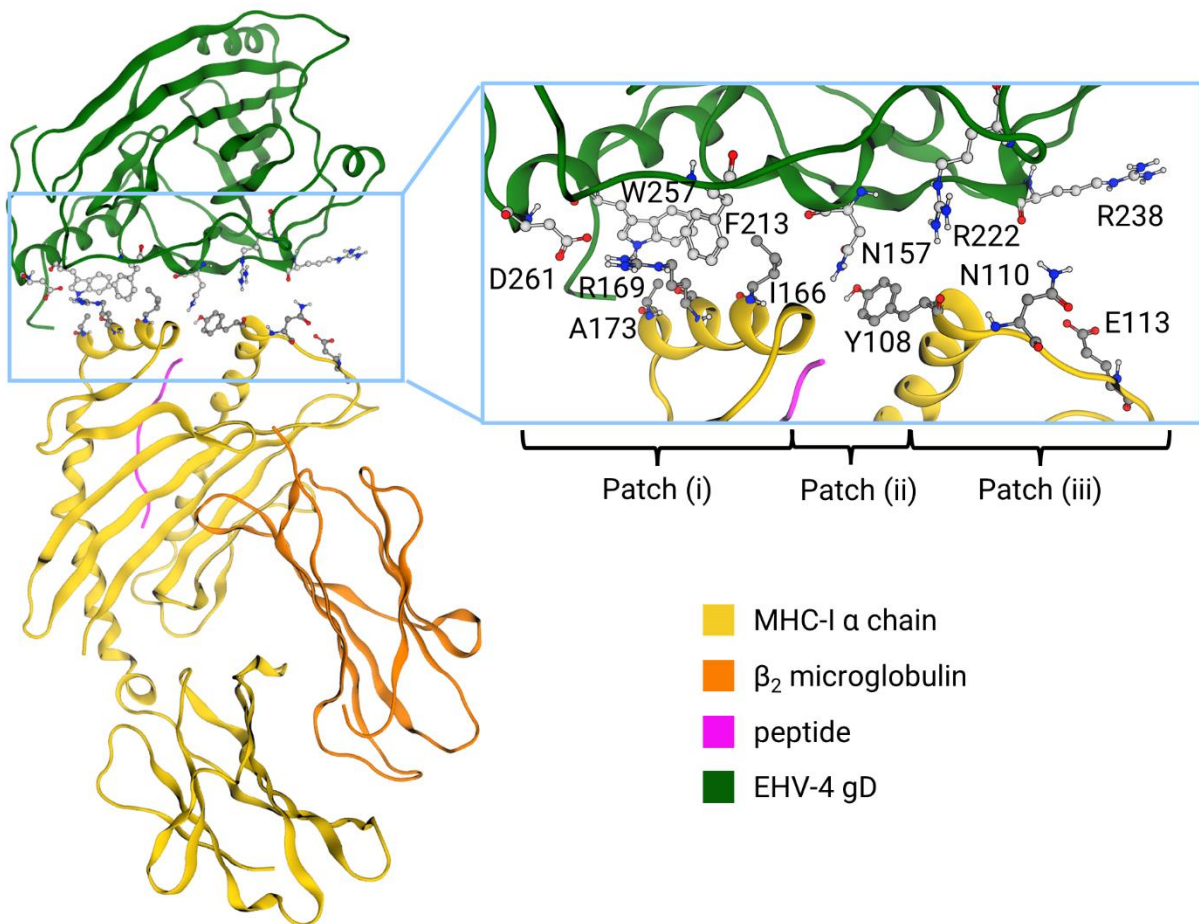
Figure 32. The binding hypothesis of equine herpesvirus 4 glycoprotein D (EHV-4 gD) to the equine major histocompatibility complex I (MHC-I) genotype 3.1 in its peptide-bound form. Color code: green ribbon and pale gray spheres- gD, golden, orange ribbon, grey spheres- MHC-I, and magenta ribbon- peptide.

Table 8. The frequencies of contacts between the EHV-4 gD and equine MHC-I genotype 3.1 in a peptide-bound binding hypothesis.

MHC-I residue	Contact frequency [%][a]	EHV-4 gD residue	Assumed role
Y108	92 ± 11	N157	O-ring sealing I166-F213 contact
N110	41 ± 28	R222	O-ring sealing I166- F213 contact
E113	28 ± 37	R238	Additional contact stabilizing the PPI
I166	**96 ± 4**	**F213**	**Hot spot contact**
R169	**99 ± 1**	**D261**	**Hot spot contact**
A173	86 ± 14	W257	O-ring sealing R169-D261 contact

[a] Molecular dynamics simulation was performed in triplicate, the value is given as average ± standard deviation. Abbreviations: EHV-1- equine herpesvirus 1, gD- glycoprotein D, MHC-I- major histocompatibility complex I, PPI- protein-protein interface.

3.2.4 Mutational Studies Confirm Predicted gD-F213 and D261 as Hot Spot Residues

Based on these binding hypotheses, we suggested two loss-of-function mutations of gD: F213A disrupting the van-der-Waals contacts and D261N disrupting the salt bridge but preserving the side chain size.

The proposed gD-mutants were successfully produced and characterized in a multi-step growth kinetics assay by Viviane Kremling under the supervision of Walid Azab in Osterrieder Lab. As predicted by our computational models, both gD-mutants F213A and D261N led to growth defects in the assay.[183] The EHV-1 gD$_{F213A}$ showed a reduction of growth by two logarithmic units and reduced virus titers in the cell supernatant. The EHV-1 gD$_{D261N}$, EHV-4 gD$_{F213A}$, and EHV-4 gD$_{D261N}$ displayed no growth in the cell culture, hence the growth kinetic assay could not be performed. Notably, the mutant EHV-1 gD$_{D261N}$ reverted to the wild-type sequence in the cell culture. This indicates that gD-residue D261 is vital for the EHV-1 replication. As a control experiment, the EHV gD-mutants were reverted to their wild-type sequences, fully restoring growth in the cell culture.

3.3 Predicting the Susceptibility of Animal Species to COVID-19

Due to the ongoing coronavirus disease 2019 (COVID-19) pandemic starting in early 2020, research on the causative agent, severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2), was urgently needed. Early reports[203] suggest that the susceptibility of animal species to SARS-CoV-2 contributes to the dynamics of the COVID-19 pandemic. Our close collaboration with the Osterrieder Lab in Berlin (Germany), which is specialized in the virology of animal infections, allowed us to focus on animal susceptibility to COVID-19 in our last study.

It has been shown for a closely related severe acute respiratory syndrome-related coronavirus 1 (SARS-CoV-1) that polymorphisms in angiotensin-converting enzyme 2 (ACE2) correlate with species susceptibility to SARS.[180] Analogously, we compared ACE2 orthologs from different animal species to identify polymorphic mutations contributing to animal susceptibility to COVID-19. This led to molecular descriptors predicting species susceptibility to this novel disease.

3.3.1 Sequence Analysis Reveals no Relevant Differences in the Spike-ACE2 Interface for Different Animal Species

In order to get an overview of animal susceptibility to SARS-CoV-2, we performed a literature search for reports describing COVID-19 infections in animals (Table 9). All case studies investigated domesticated animals and livestock with a high potential to transmit the virus to humans.

At the beginning of the project, we compared the primary structures of ACE2 orthologs in the most prominent domesticated animal species: dog (*Canis lupus familiaris*) and cat (*Felis catus*) with the human sequence as a positive control (Figure 33A, Appendix Figure 12). Firstly, we investigated the residues reported as hot spots for the binding of the receptor-binding domain (RBD) of SARS-CoV-2 (31, 34, 35, 41, 79, 82, and 83, referring to the human sequence).[181] We identified two polymorphic mutations in the canine sequence; H34/33Y and M82/81T (canine residues are offset by one value compared to other sequences). Mutation M82T is also present in cats, which are susceptible to COVID-19. Hence, we focused on the polymorphism H34/33Y only present in dogs, a virtually asymptomatic species. To determine whether the H34/33Y mutation is unique for dogs, we compared the ACE2 sequences of other animals with reported COVID-19 susceptibility (Table 9, Figure 33B). We found this polymorphism in

the ACE2 sequence of the ferret (*Mustela putorius*), a species prone to COVID-19. Therefore, we assumed that the mutation H34/33Y does not hinder viral entry. The sole comparison of ACE2 sequences did not provide any correlation between hot spot polymorphisms and COVID-19 susceptibility of animal species.

Table 9. Reported species susceptibility to COVID-19 included in this study.

Animal species	Reported susceptibility to COVID-19	Reference
Cat (*Felis catus*)	Yes	Shi et al.[203]
Dog (*Canis lupus familiaris*)	Virtually asymptomatic, unable to transfer SARS-CoV-2	Sit et al.[204]
Ferret (*Mustela putorius*)	Yes	Shi et al.[203]
Mouse (*Mus musculus*)	No	Munoz-Fontela et al.[205]
Rat (*Rattus norvegicus*)	No	Munoz-Fontela et al.[205]
Syrian hamster (*Mesocricetus auratus*)	Yes	Munoz-Fontela et al.[205]

Abbreviations: COVID-19- coronavirus disease 2019, SARS-CoV-2- severe acute respiratory syndrome-related coronavirus 2.

To investigate the influence of ACE2-mutations outside the binding hot spots, we decided to compare the dynamics of three-dimensional models of ACE2-orthologs in complex with the spike (S) RBD of SARS-CoV-2. At the beginning of the project, no animal ACE2 structures were available, eventually leading to the construction of homology models.

All homology models were built based on the first available human ACE2 X-ray structure in complex with S RBD from SARS-CoV-2 (PDB-ID: 6M0J[181]). The high sequence identity of ACE2-orthologs to the human sequence (ranging from 83% for the ferret to 87% for the Syrian hamster, Appendix Figure 12) allowed us to build high-quality models with maximally one Ramachandran outlier[89] in the dog and mouse ACE2 (Appendix Figure 13). The single Ramachandran outlier results from a polymorphic mutation (G337/336S in dog and G337A in mouse) and is located distally to the S RBD binding site. The backbone RMSD of the homology models compared to

the human ACE2 template were in the range from 0.6 to 0.9 Å, suggesting only negligible deviations from the global fold.

Figure 33. Sequence comparison of the epitope residues of (A) human, dog, and cat ACE2 orthologs and (B) extended comparison with ferret, mouse, rat, and Syrian hamster ACE2 orthologs. (C) Sequences of ACE2 orthologs in species, coronavirus disease 2019 susceptibility being predicted (Chinese hamster, Campbell's dwarf hamster, and red squirrel). Residues marked in orange are reported as hot spots[181] for the binding of the residue binding domain of severe acute respiratory syndrome-related coronavirus 2.

All ACE2 homology models were subsequently assembled together with the S RBD from the template structure (PDB-ID: 6M0J). We observed only isolated side chain clashes. They were manually relaxed using the rotamer tool integrated into MOE and energy minimized using the OPLS-AA force field.[142]

The prepared homology complexes of human and animal ACE2-orthologs in complex with SARS-CoV-2 S RBD were simulated over 100 ns in five replicates each, generating in total 3 µs of molecular dynamics (MD) trajectory. The visual inspection of all trajectories revealed that the RBD-loop coordinating the binding pocket C in the dog and rat ACE2-complexes performs a movement to a larger extent than in other species (Figure 34).
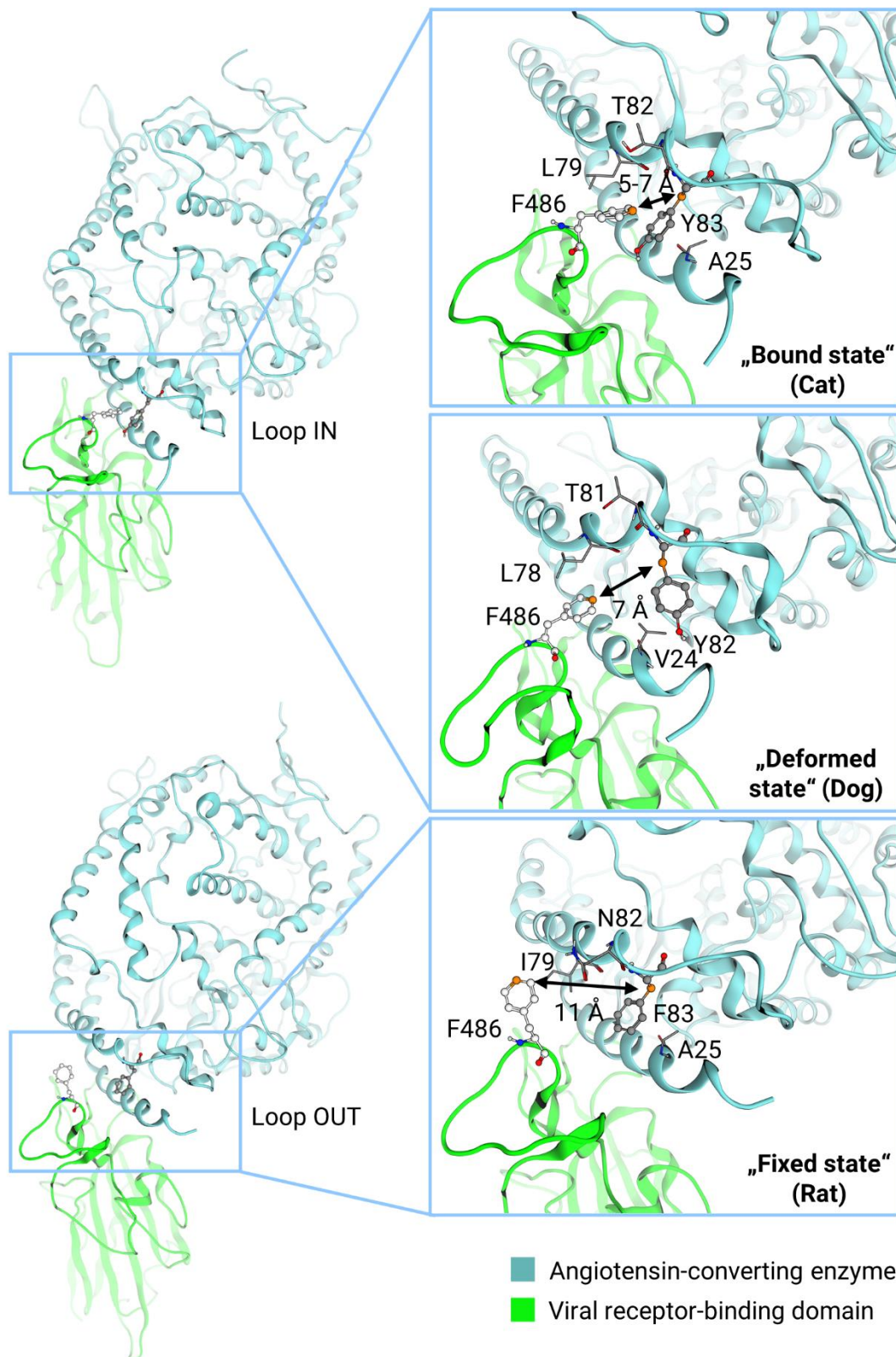
Figure 34. Conformations of the receptor-binding domain (RBD) of severe acute respiratory syndrome-related coronavirus 2 in complex with angiotensin-converting enzyme 2 (ACE2) orthologs observed in molecular dynamics simulations. Color code: green ribbon and pale grey atoms- RBD, cyan ribbon and grey atoms- ACE2.

3.3.2 The Conformations of the Lipophilic Binding Pocket C of ACE2 Restrict Lipophilic Contacts to the RBD-Loop of Spike

To find a plausible explanation for the recurring dissociation events of the RBD-loop, we analyzed the distance between the RBD hot spot residue F486 and canine ACE2-Y82. We chose the distance between Cζ of RBD-F486 and Cβ of ACE2-Y82 as a surrogate parameter for lipophilic contacts in the binding pocket C (Figure 34). This is the closest possible distance between those two residues. The MDAnalysis[201, 202] scripts used for the calculation of all molecular descriptors were provided by Trung Ngoc Nguyen from our research group.

In Syrian hamster and human ACE2 simulations, we observed a single peak at around 5-7 Å in the RBD-F486 − ACE2-Y83 distance distribution (Figure 35). This suggests frequent and close contacts between RBD-F486 and ACE2-Y83, as expected for the susceptible species. We call this conformation the "bound state". Surprisingly, the close RBD-F486 − ACE2-F83 contacts in the binding pocket C were also detected in mouse, a non-susceptible species. We assume that murine ACE2 does not allow RBD binding in binding pockets A or B.

In cat, ferret, and dog simulations, we observed two peaks at around 5 and 7 Å suggesting the existence of two possible RBD-F486 conformations. Both peaks are identical, with the "bound state" indicating frequent and close lipophilic contacts between RBD-F486 and ACE2-Y83 (82 in dog). This is surprising for the dog since this species is virtually asymptomatic. A careful manual analysis of the binding pocket C conformations revealed that ACE2-Y82 rotates around the χ1 bond, displaying two rotamers (Figure 34). The first rotamer (χ1 angle around -70°, Figure 35) is present in all species and represents the binding rotamer with the phenyl ring rotated towards the RBD. The second rotamer can be only found in the dog (χ1 angle around -170°), with the phenyl ring rotated outwards of the binding pocket C.

Figure 35. Distribution of the distance between residues F486 (receptor-binding domain) and F/Y83/82 (angiotensin-converting enzyme 2) as a surrogate parameter for a lipophilic contact between these residues (x-axis). Distribution of the χ1 angle in residue F/Y83/82 of angiotensin-converting enzyme 2 (y-axis).

We term the state with the outwards rotated ACE2-residue Y83 the "deformed state" due to a distortion of the binding pocket C. The extent of binding pocket C deformation can be seen in the distance between Cβ of ACE2-Y82 in the upper and Cβ of ACE2-F27 in the lower helix, which limits binding pocket C (Figure 36). The canine simulations show maximal distances between the upper and lower helix above 9 Å not observed in other species, confirming our observations.
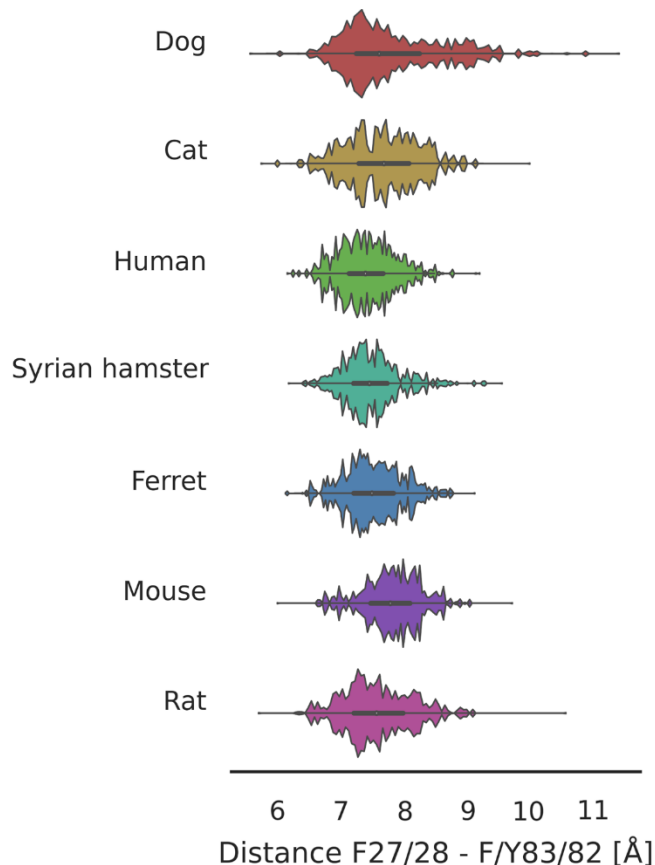
Figure 36. Distribution of the distance between angiotensin-converting enzyme 2 residues F28/27 and F/Y83/82 as a surrogate parameter for deformation of binding pocket C.

The outward rotation of the ACE2-Y82 side chain suggests that the lipophilic contact to RBD-F486 is interrupted in the "deformed state". This correlates nicely with the observation that dogs are virtually asymptomatic after a SARS-CoV-2 infection.

In the next step, we searched for the reason for the deformation of binding pocket C in dog simulations. We focused on the direct environment of ACE2-Y82. Careful visual inspection of the canine simulations showed that the residue ACE2-V24 might be responsible for the outward rotation of ACE2-Y82. Compared to other species harboring an alanine at homologous position ACE2-25, the valine side chain in canine

ACE2 is bulkier, leading to steric hindrance and pushing the Y82 side chain out of binding pocket C.

To confirm our assumption, we virtually generated a gain-of-function canine ACE2-mutant V24A and simulated it under the same conditions as the wild-type sequences. Visual inspection of the obtained trajectories and the calculated molecular descriptors (distance RBD-F486 – ACE2-Y83, χ1 angle in ACE2-Y83/82, and distance ACE2-F28/27 – ACE2-Y83/82) confirmed that the canine ACE2-V24A mutant does not display the "deformed state" of binding pocket C (Figure 37). This suggests that the polymorphism at ACE2-position 24 (in dogs, 25 in other species) might be responsible for the asymptomatic course of COVID-19 in dogs.

Subsequently, we analyzed the occupation of binding pocket C in the rat ACE2 (Figure 35). The simulations show two peaks in the distance distribution RBD-F486 – ACE2-F83: one at 7 and the second at 10 Å. The first peak is identical to the "bound state" identified in other species. The second peak displays a comparably larger distance, suggesting a dissociation event of RBD-F486. The visual inspection of rat MD trajectories revealed that the RBD-F486 leaves binding pocket C. The phenyl ring of RBD-F486 establishes lipophilic contacts to the side chains of ACE2-residues flanking the binding pocket C (24, 81, or 87). In this state, RBD-F486 is fixed outside binding pocket C, hence we call this the "fixed state" (Figure 34).

To explain the dissociation events of RBD-F486 in the rat simulation, we analyzed the previously calculated molecular descriptors. In contrast to the canine ACE2, the rat ACE2 does not display a "deformed state" of binding pocket C (Figure 35, Figure 36). The visual inspection of the rat simulations revealed that binding pocket C is slightly flatter than in other species. In order to characterize the depth of binding pocket C, we calculated distributions of the shortest distance per frame between Cβ of ACE2-F/Y83/82 as the deepest point of binding pocket C and all side chain atoms flanking the binding pocket C (ACE2-residues 24, 78, and 81, Figure 38).
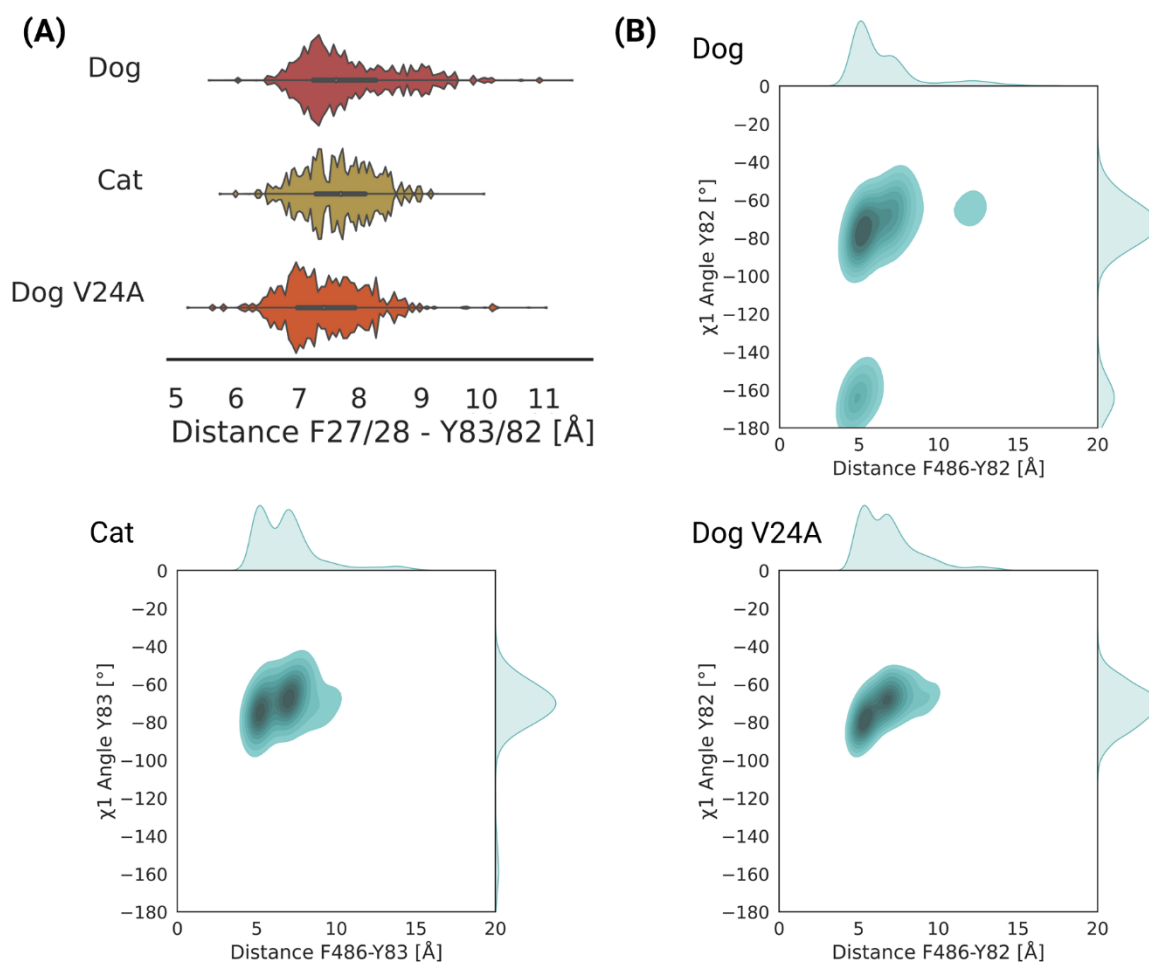
Figure 37. The molecular descriptors calculated for cat, dog, and dog ACE2-V24A mutant: (A) deformation of binding pocket C (distribution of the χ1 angle in ACE2-Y83/82 and distribution of the distance between residues F28/27 and Y83/82), (B) lipophilic contacts in binding pock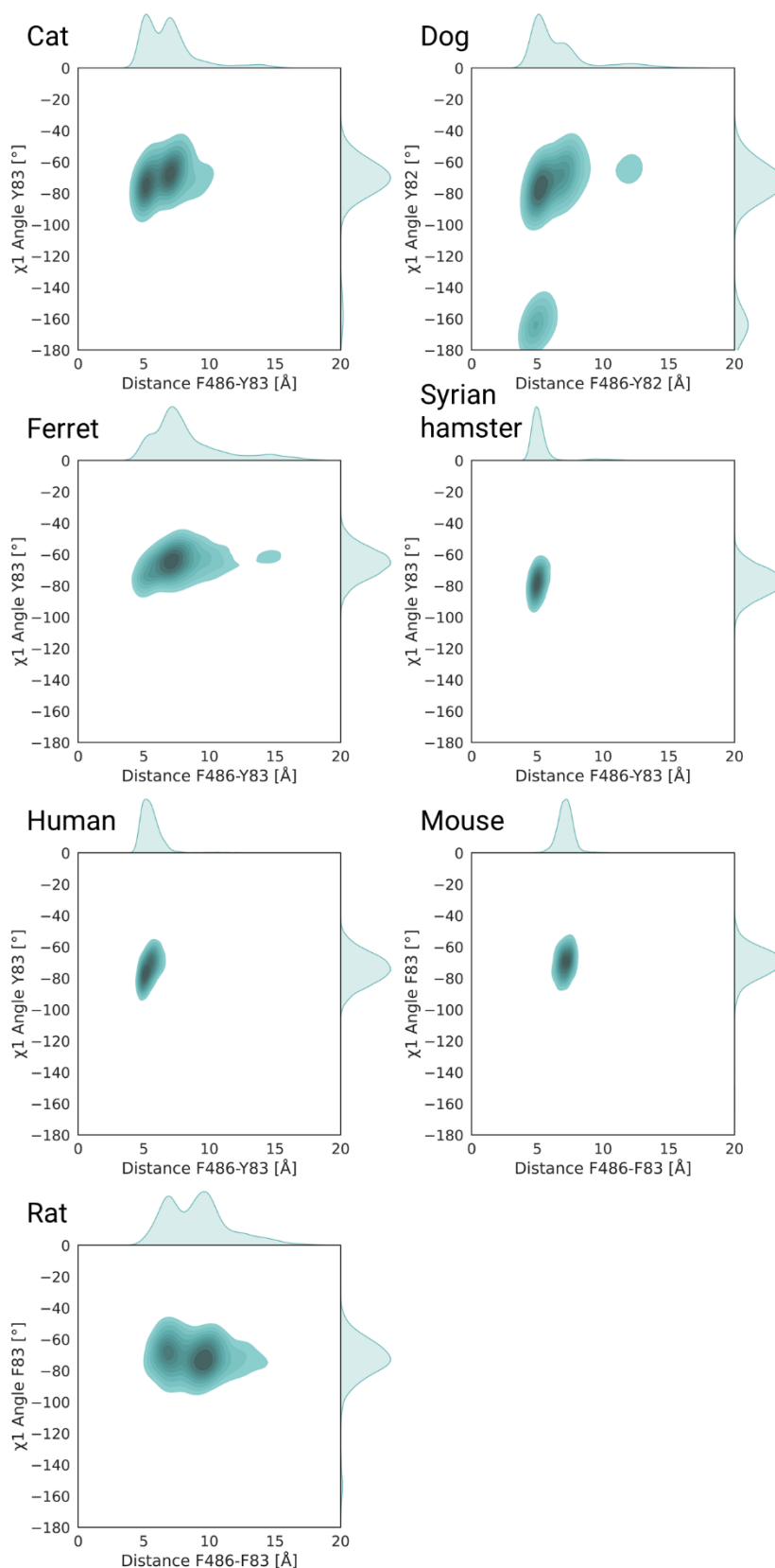et C (distribution of the distance between residues F486 in the receptor-binding domain and Y83/82 in angiotensin-converting enzyme 2, ACE2).

The analysis of the calculated "depth" descriptors revealed two trends. Firstly, the distance between ACE2-F83 and ACE2-I79 in rat simulations is distributed around 6 Å, in contrast to other species, which adopt higher distance values around 7-8 Å. This is in line with our observation of the flat binding pocket C in the rat. The rat is the only species displaying isoleucine at ACE2-position 79. We assume that the branched chain of ACE2-I79 causes a steric hindrance and repels RBD-F486 from close contact to ACE2-F83 in the rat simulations. Interestingly, the mouse also displays a branched side chain at position 79 (threonine), but this does not flatten binding pocket C. We rationalize the greater impact of isoleucine on the flattening of binding pocket C through a bulkier side chain of this residue compared to threonine.
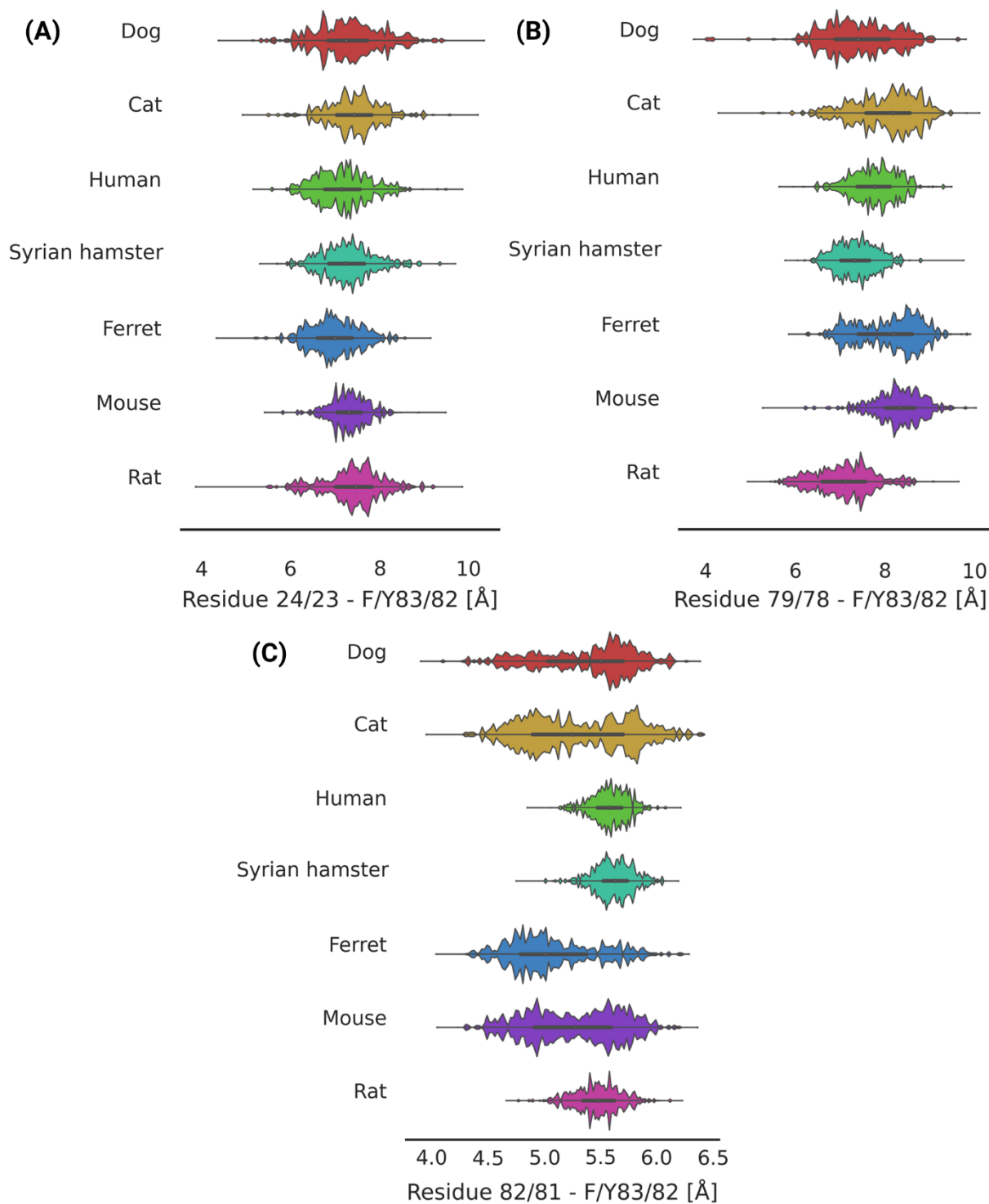
Figure 38. Distributions of shortest distance per frame between Cβ of angiotensin-converting enzyme 2 (ACE2) residue F/Y83/82 as the deepest point of binding pocket C and all side chain atoms flanking the binding pocket C (ACE2-residues (A) 24, (B) 78, and (C) 81).

The second trend shows that the distance between ACE2-F/Y83/82 and ACE2-residue 82 (or 81 in dog) displays a wider distribution in the dog, cat, ferret, and mouse simulations. All these species expose a residue with a short side chain at ACE2-

position 82/81 (threonine in dogs, cats, ferrets, and serine in mice). This suggests that a long ACE2-residue 82/81 (methionine in humans or asparagine in Syrian hamsters) might support RBD-F486 binding by trapping the side chain in binding pocket C.

To explain the non-susceptibility of the mouse, we analyzed the interactions in binding pockets A and B. In the first step, we focused on binding pocket A harboring a polymorphism K353H only present in the non-susceptible species mouse and rat.

3.3.3 Spike Protein Binds to the Binding Pockets A and B of ACE2 via Extensive Hydrogen-Bond Networks

In order to investigate the role of ACE2-polymorphism K353H, we visually inspected binding pocket A in the MD simulation of rodent ACE2 (mouse, rat, Syrian hamster). In the simulations of Syrian hamster ACE2, residue K353 establishes a salt bridge to ACE2-D38 that is coordinated by hydrogen bonds from RBD-Q498. This presumed binding hot spot is sealed by hydrogen bonding established by ACE2-residues 37, 41, 42, and RBD-residues 449, 496, 501, and 505. In the simulations of mouse and rat ACE2-orthologs, H353 cannot establish a salt bridge to the counterpart ACE2-38 due to less basic properties and a shorter side chain. Hence, we assumed that the polymorphism K353H might lower the total number of hydrogen bonds in the binding pocket A, leading to weaker RBD binding. To validate this hypothesis, we counted the hydrogen bonds in the binding pocket A (Figure 39). The measurement confirmed our expectations, showing seven hydrogen bonds on average in the binding pocket A of the Syrian hamster compared to three and four hydrogen bonds in rats and mice, respectively.

Finally, we analyzed binding pocket B showing two polymorphic mutations in ACE2 of the mouse (E/D30N, K31N) and the rat (T27S, E/D30N). We hypothesized that these mutations might also decrease the COVID-19 susceptibility of mice and rats.

Initially, we investigated the mutation E/D30N present in both non-susceptible species. Visual inspection of simulation trajectories of mice, rats, and Syrian hamsters for comparison, revealed that acidic residues at ACE2-position 30 establish a salt bridge to RBD-K417 to support the binding between these proteins. To confirm this hypothesis, we measured the shortest distance per frame between the atoms of ACE2-residue 30 and RBD-K417 (Figure 40). As expected, all species carrying an acidic residue at position 30/29 show a frequent distance of around 3 Å, indicating the

existence of a salt bridge present over the whole simulation time. Rats and mice with a neutral residue at ACE2-position 30 establish no interactions to the RBD-K417, with a dominant distance range of 6-8 Å.
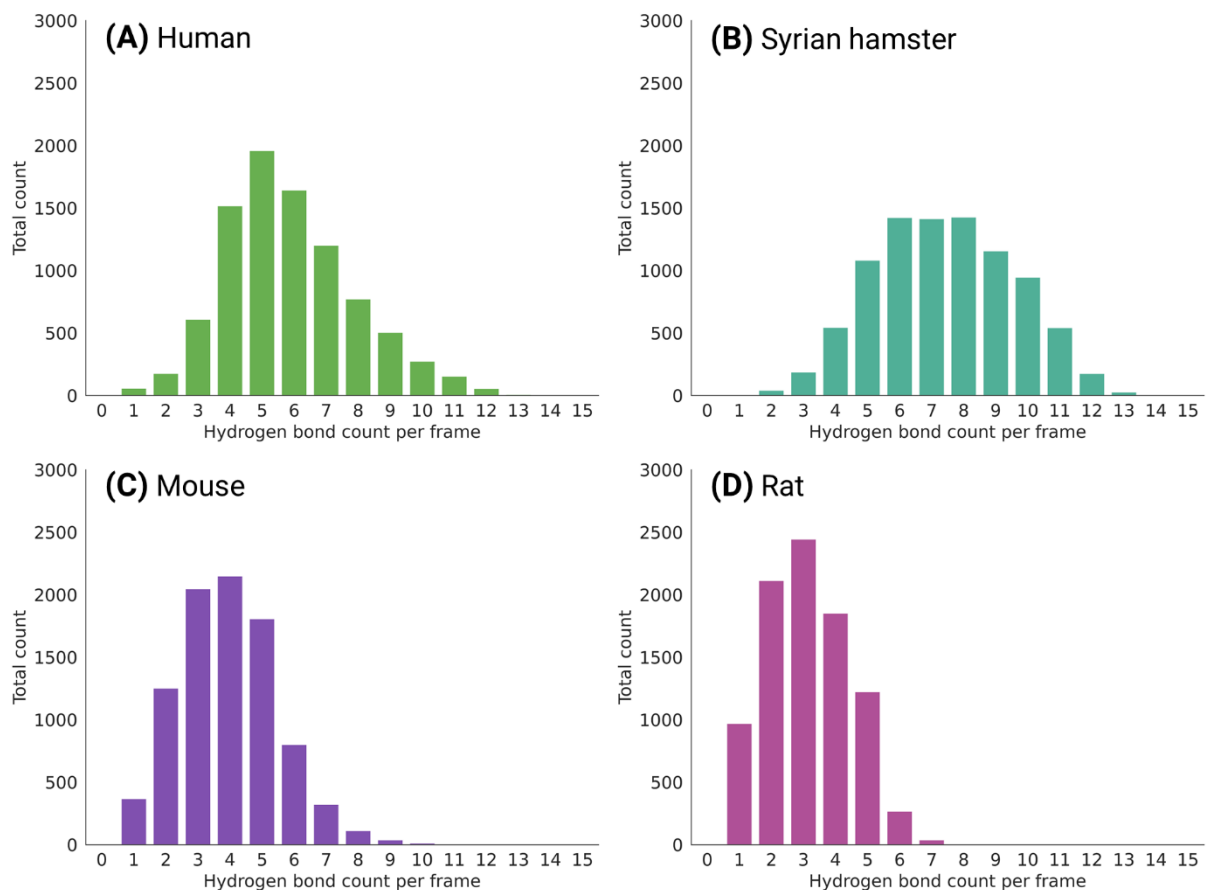
Figure 39. Hydrogen bond count in binding pocket A between angiotensin-converting enzyme 2 (ACE2) residues 37, 38, 41, 42, 353, and viral receptor-binding domain residues 449, 496, 498, 501, and 505 in the simulations of (A) human, (B) Syrian hamster, (C) mouse, and (D) rat ACE2.

In the next step, we investigated the neighboring K31N mutation present in mice only. We observed that ACE2-K31 in Syrian hamsters establishes a salt bridge to ACE2-E35. This ACE2-residue pair is coordinated by RBD-Q493, similar to the hydrogen bond network observed in the binding pocket A (ACE2-D38 – ACE2-K353 – RBD-Q498). In murine simulations, no building of a salt bridge between ACE2-N31 and ACE2-E35 is observed, hence no coordination of RBD-Q493 is possible.
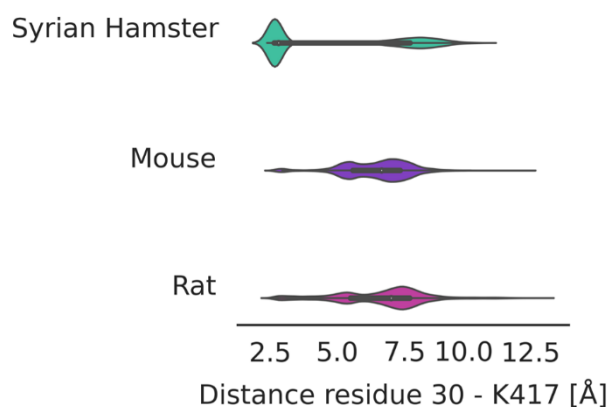
Figure 40. Distributions of shortest distance per frame between residue 30 of angiotensin-converting enzyme 2 (D30 in Syrian hamsters, N30 in mice, and rats) and residue K417 of the viral receptor-binding domain.

We calculated the shortest distance per frame for the residue pairs (i) ACE2-31 – ACE-35 and (ii) ACE2-31/30 – RBD-Q493 as surrogate parameters for the presence of hydrogen bonds (Figure 41). The measurements confirmed our observations, showing a dominant distance between the ACE2-residues N31-E35 and ACE2-N31 – RBD-Q493 around 5-6 Å in murine simulations. This suggests that no hydrogen bonding sealing binding pocket B is present in the mouse simulations.

The last polymorphic mutation T27S present in rats is buried in the lipophilic ACE2-interface between binding pockets B and C. Visual inspection of the rat ACE2 simulations did not reveal any large-scale conformational changes. Hence, we could not calculate any molecular descriptor characterizing the influence of the T27S mutation on RBD binding. The residue ACE2-S27 is surrounded by lipophilic residues RBD-F456, Y473, and Y489. We assume that the polymorphism T27S leads to a lower RBD binding affinity since a serine residue is less likely to establish lipophilic contacts to the RBD interface.
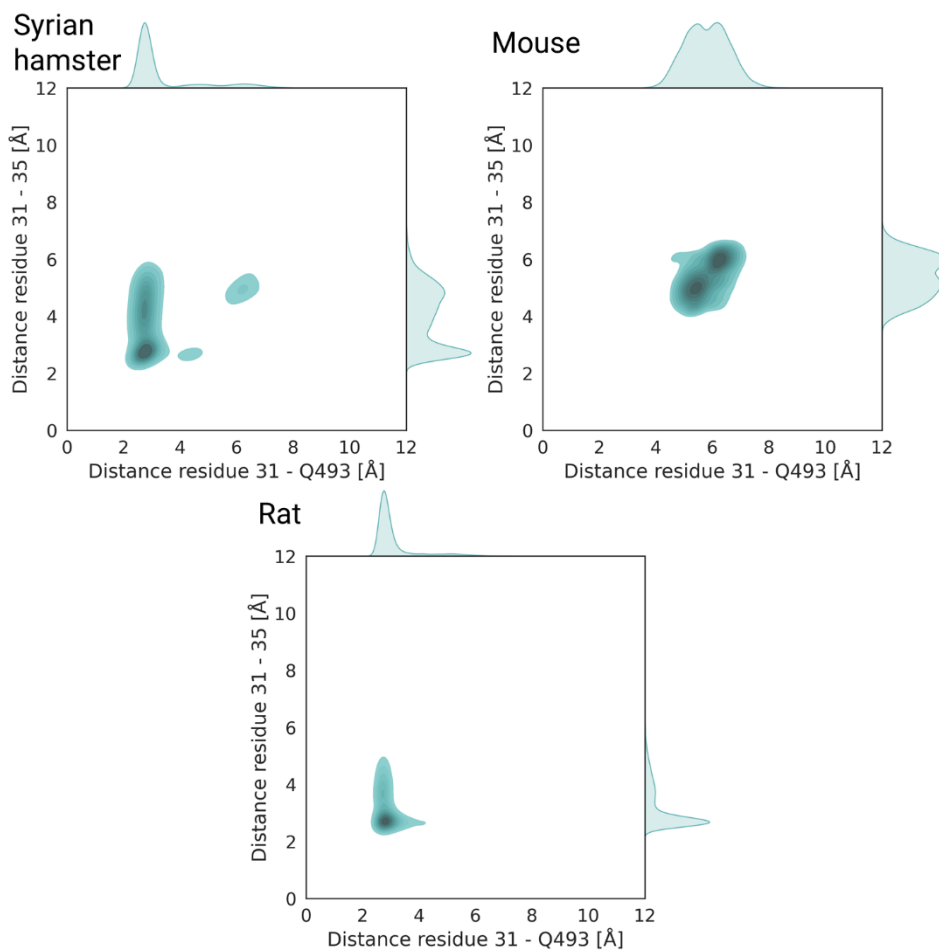
Figure 41. Distributions of shortest distance per frame (x-axis) between residues 31 and 35 of angiotensin-converting enzyme 2 (ACE2) in Syrian hamsters, mice, rats, and (y-axis) ACE2-31 and residue Q493 of the viral receptor-binding domain.

To confirm our observations, we virtually mutated ACE2 of mice (N30D, N31K, H353K-3-fold mutant) and rats (S27T, N30D, I79L, H353K- 4-fold mutant). The ACE2 mutants were simulated under the same conditions as the wild-type sequences to achieve comparability of the results. Subsequently, we calculated the molecular descriptors characterizing the depth of binding pocket C, hydrogen bond network in the binding pocket A and B (Figure 42, Figure 43). The measurements met our expectations. Both ACE2-mutants showed recurring interaction patterns present in the simulations of COVID-19-susceptible species.
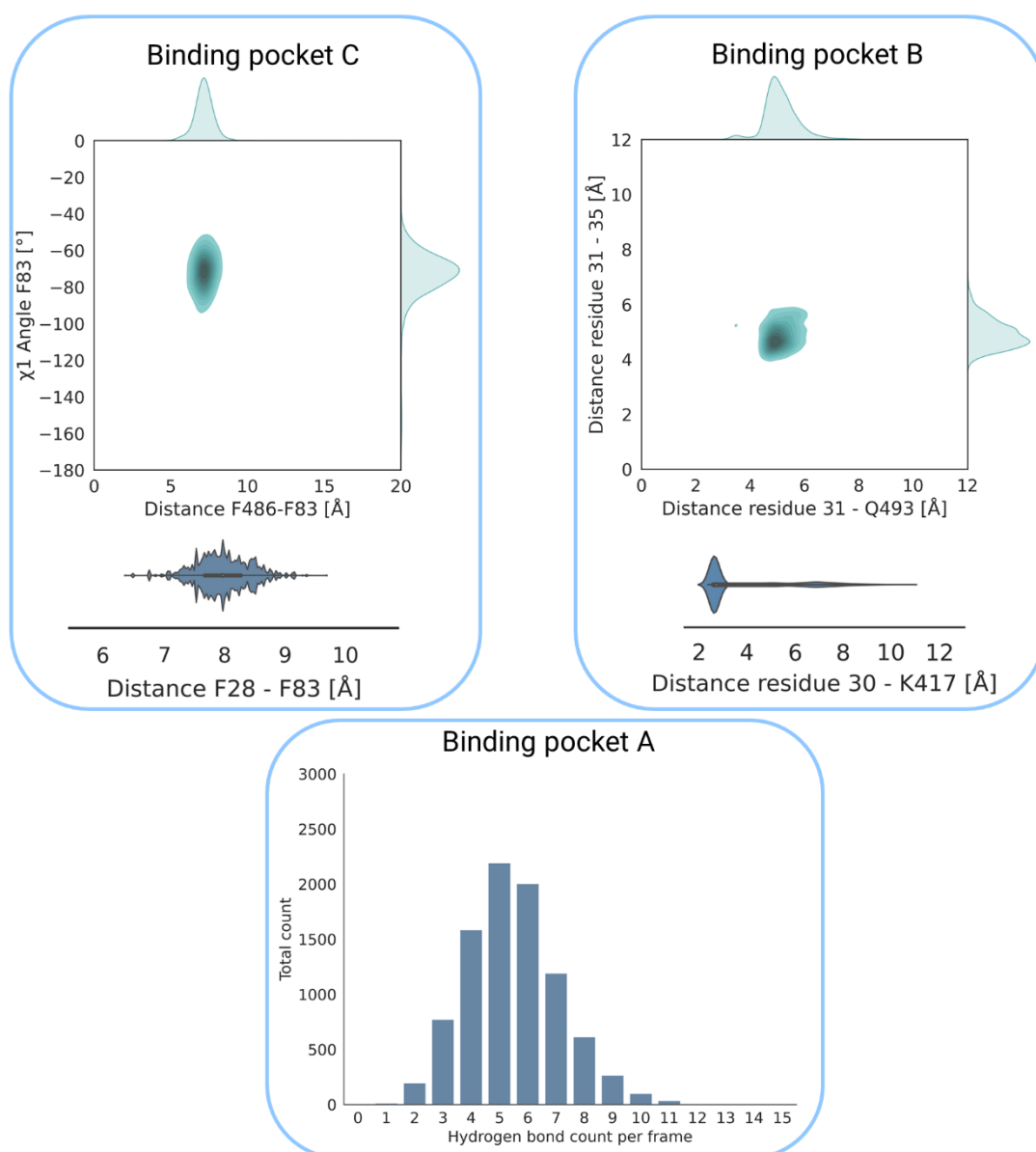
Figure 42. Molecular descriptors characterizing the depth of and lipophilic contacts in binding pocket C and the hydrogen bond network in binding pockets A and B of mouse angiotensin-converting enzyme 2 (ACE2) 3-fold mutant (N30D, N31K, H353K).
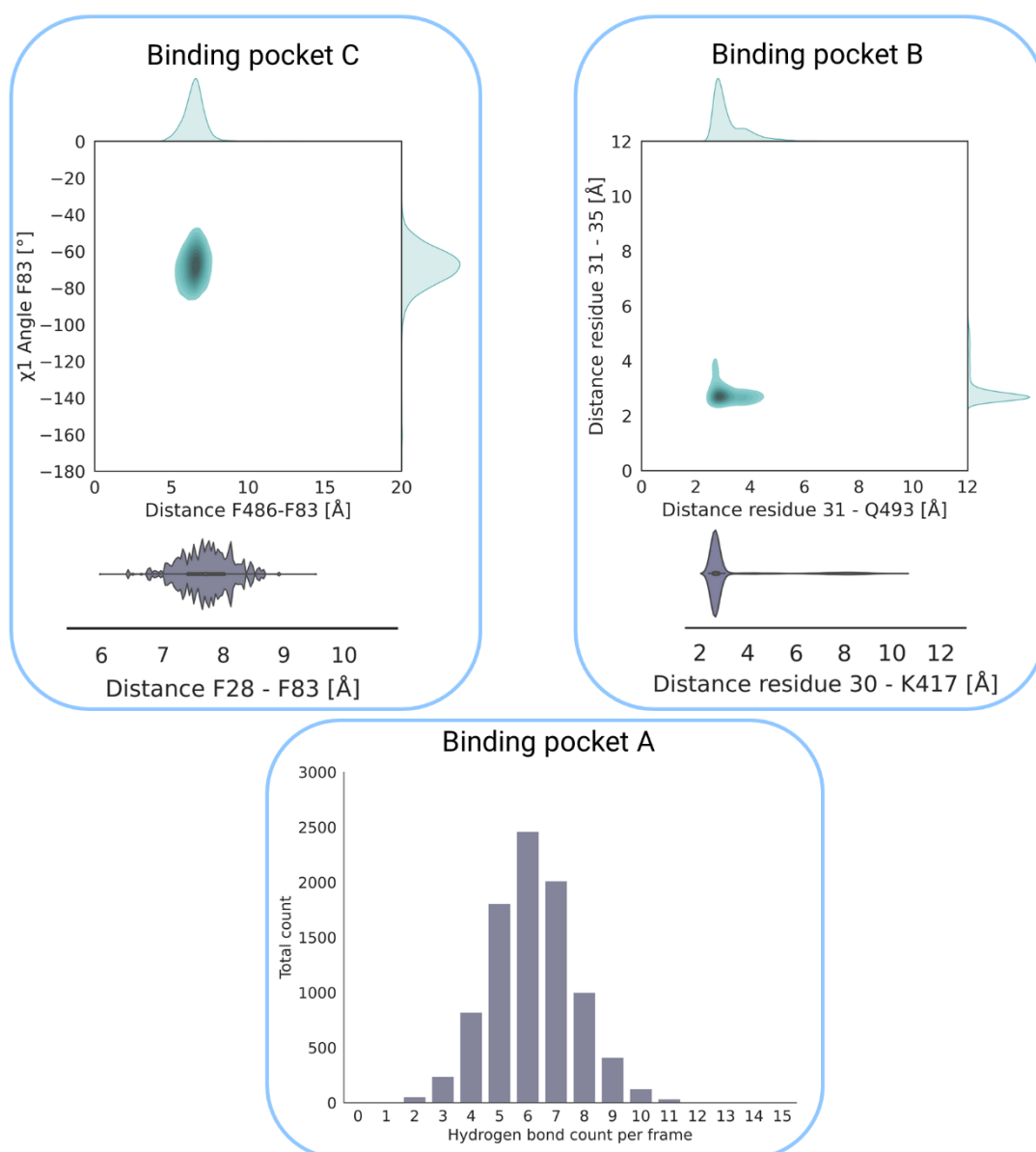
Figure 43. Molecular descriptors characterizing the depth of and lipophilic contacts in binding pocket C and the hydrogen bond network in binding pockets A and B of rat ACE2 4-fold mutant (S27T, N30D, I79L, H353K).

According to our analysis, we propose gain-of-function mutations in ACE2 of non-susceptible species as a validation for our models (Table 10).

In the next step, we investigated, which wild-type animal species could be prone to COVID-19. We chose three species with unknown reported COVID-19 susceptibility (as of April 2020): red squirrel (*Sciurus vulgaris*), Chinese hamster (*Cricteulus griseus*), and Campbell's dwarf hamster (*Phodopus campbelli*). Red squirrels are broadly present in an urban environment representing a potential SARS-CoV-2 reservoir. The Chinese and

Campbell's dwarf hamsters could be used as animal models to investigate the COVID-19 course experimentally.

Table 10. Proposed gain-of-function mutations in ACE2 of non-COVID-19-susceptible species.

Non-susceptible animal species	ACE2 mutation (in the binding pocket …)	Comparable susceptible animal species
Dog (*Canis lupus familiaris*)	V24A (C)	Cat (*Felis catus*)
	T81M (C)	Human
Mouse (*Mus musculus*)	N30D (B)	Syrian hamster (*Mesocricetus auratus*)
	N31K (B)	
	H353K (A)	
Rat (*Rattus norvegicus*)	S27T (B)	Syrian hamster (*Mesocricetus auratus*)
	N30D (B)	
	I79L (C)	
	H353K (A)	

Abbreviations: ACE2- angiotensin-converting enzyme 2, COVID-19- coronavirus disease 2019.

3.3.4 Geometric Interaction Descriptors Predict Susceptibility of Dwarf Hamster Species and Red Squirrel to COVID-19

Similar to other investigated species, we built the homology models of ACE2 from red squirrel, Chinese, and Campbell's dwarf hamsters based on the first available human ACE2 X-ray structure in complex with S RBD from SARS-CoV-2 (PDB-ID: 6M0J[181]). The high sequence similarity and identity of ACE2-orthologs to the human sequence (93% and 86%, respectively, for the Campbell's dwarf hamster, 93% and 87%, respectively, for the Chinese hamster, and 93% and 88%, respectively, for the red squirrel, Appendix Figure 12) allowed us to build high-quality models. Solely the red squirrel model showed one Ramachandran outlier[89] (Appendix Figure 13) resulting from a polymorphic mutation G337A located distally to the S RBD binding site. The backbone

root mean square deviation of the homology models compared to the human ACE2 template amounted to 0.6-0.8 Å, suggesting only negligible deviations from the global fold. The obtained models were assembled with the SARS-CoV-2 RBD and simulated under the same conditions as the other ACE2-RBD complexes. Finally, to predict COVID-19 susceptibility of the red squirrel, Chinese, and Campbell's dwarf hamsters, we calculated the molecular descriptors characterizing the depth and lipophilic contacts in the binding pocket C, hydrogen bond networks in the binding pockets A and B (Figure 44, Figure 45, Figure 46).
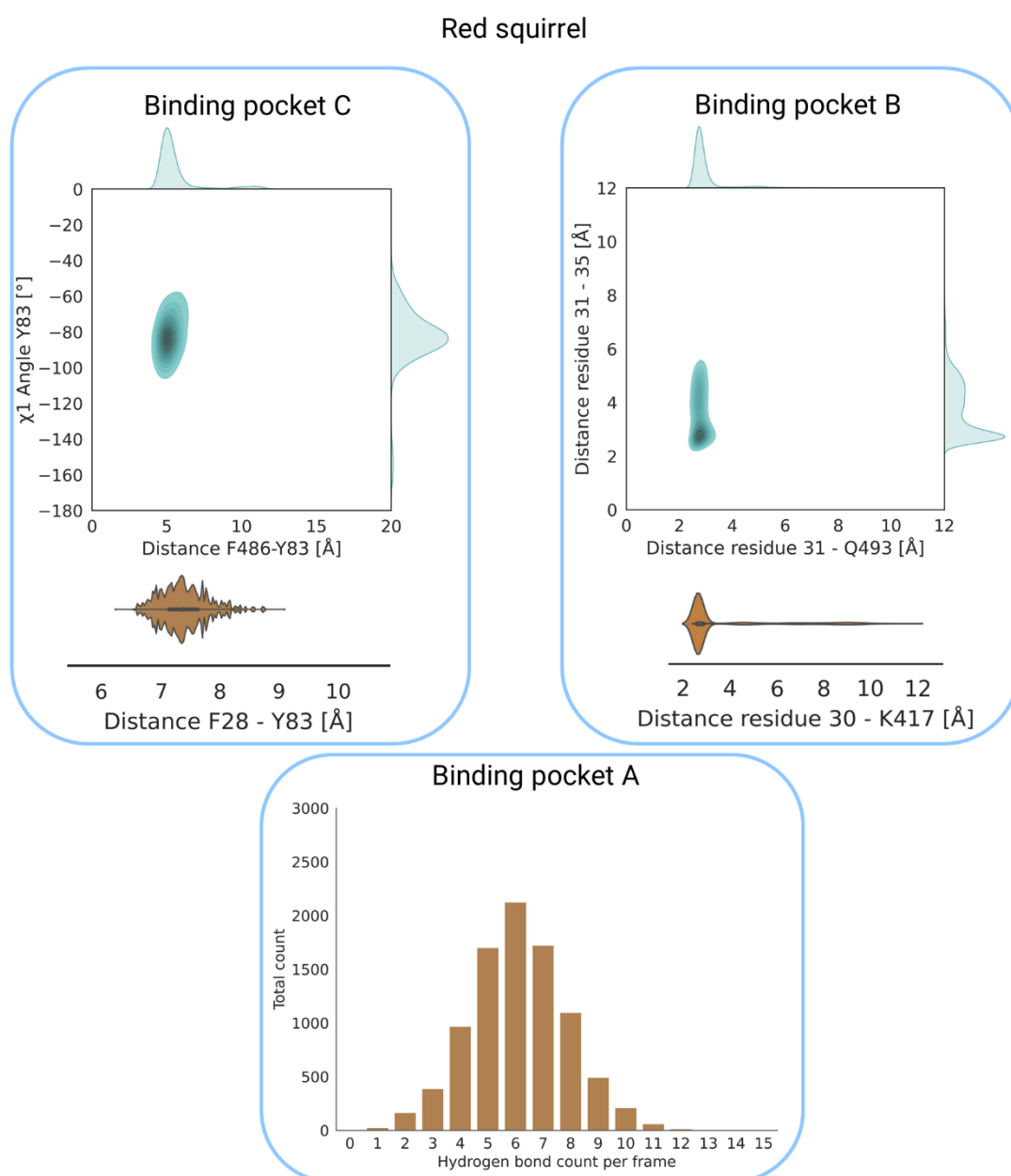
Figure 44. Descriptors characterizing binding pockets A, B, and C of angiotensin-converting enzyme 2 of red squirrel in complex with the viral receptor-binding domain.
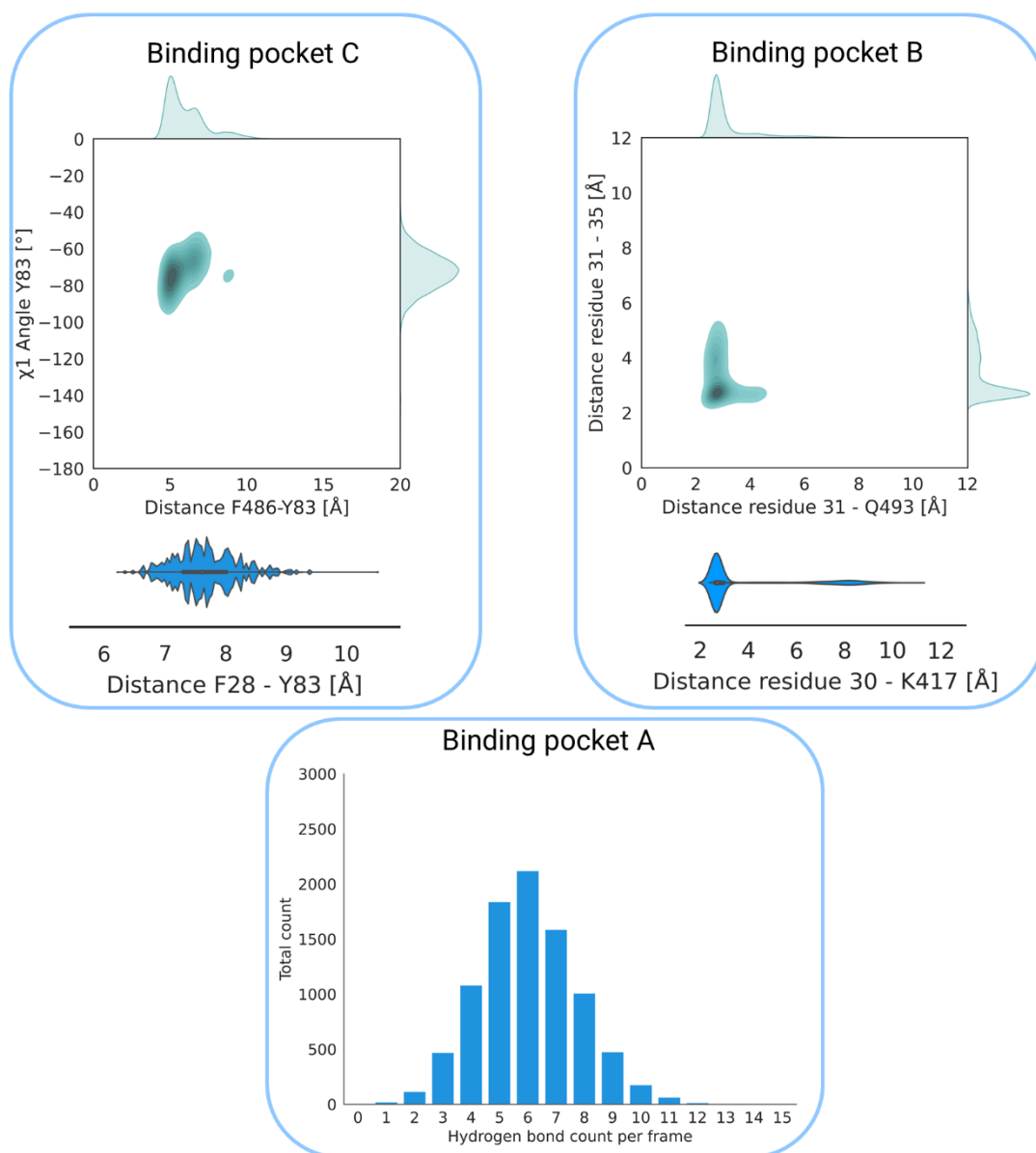
Figure 45. Descriptors characterizing binding pockets A, B, and C of angiotensin-converting enzyme 2 of Chinese hamster in complex with the viral receptor-binding domain.
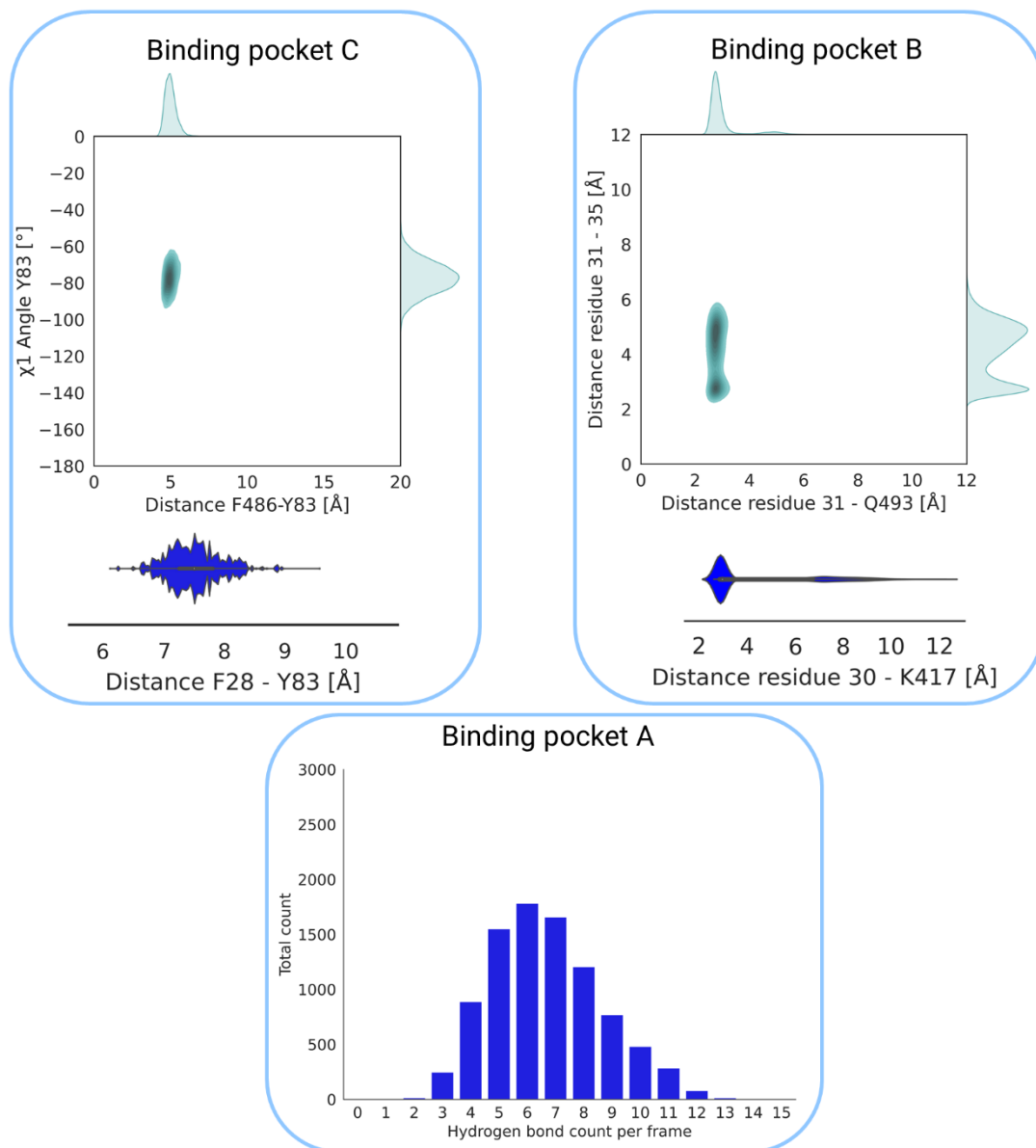
Figure 46. Descriptors characterizing binding pockets A, B, and C of angiotensin-converting enzyme 2 of Campbell's dwarf hamster in complex with the viral receptor-binding domain.

The calculated molecular descriptors for ACE2 of the red squirrel, Chinese, and Campbell's dwarf hamsters suggest that all these species should be susceptible to COVID-19. Our findings were confirmed for the Chinese and Campbell's dwarf hamsters by our collaborators from the lab of Nikolaus Osterrieder.[206, 207]

4.   Discussion

Viral infections represent a global health concern that can emerge at any time, as the example of coronavirus disease 2019 (COVID-19) has shown. Systematic research on viruses independent from epidemics is needed to control existing infections and prevent the emergence of novel viruses.

In this thesis, an overview and successful application of computational methods to target emerging viral infections were presented. Explicitly, substrate-protease and protein-protein interactions were investigated to mechanistically explain the recognition of virus receptors and identify small molecular inhibitors of viral proteases. Due to our focus on emerging and neglected viral species, we were working in the scenario of scarce data availability. We have shown that computational modeling is possible, even if there are only a few experimental starting points. Moreover, we show that computational modeling allows a quick generation of valid results and offers a platform for iterative knowledge exchange with experimental collaborators. In the following paragraphs, the novelty of each project, the gain of information for the experimental collaborators, and the limitations of the used approaches will be discussed in detail.

Development of Small Molecules Targeting NS2B-NS3 Proteases

Infections with emerging viruses as Zika or West Nile virus lead to thousands of fatalities annually.[14] Despite intensive research, no vaccines or antiviral substances against these infections are available. The development of inhibitors of flaviviral NS2B-NS3 proteases represents a viable strategy to combat infections with Zika and West Nile virus.

In this study, we have discovered two completely novel non-covalent flaviviral protease inhibitors identified in a virtual screening campaign. To overcome the challenges of protease targeting, we used a novel software, *PyRod*.[128] *PyRod* derives pharmacophore models by tracing water molecules in molecular dynamics (MD) simulations and takes into account protein flexibility. Despite using water as a molecular probe for interactions, *PyRod* identifies lipophilic contacts needed for binding to shallow binding pockets. Both novel inhibitors can serve as molecular references for protease assays and promising starting points for optimization campaigns.

Our compounds show several advantages rendering them suitable for optimization: activity in the low micromolar concentration range, molecular weight lower than 500 Da, and lack of pan-assay interfering substructures. Recent reviews by our colleagues, Voss and Nitsche, show, how unique these properties are in the field of flaviviral protease inhibitors.[208, 209] In contrast to our ligands, several inhibitors are either peptide derivatives with undesired pharmacokinetic properties, molecules with molecular masses above 500 Da (dideoxystreptamine ethers,[169] diaminoisoquinolones,[168] or novobiocin[210]), molecules that could potentially interfere with assay readout due to fluorophore moieties, or structures that contain reactive substructures (Michael acceptors, thiazolidinones, catechols, and quinones). Furthermore, we show a hypothesized binding mode, allowing the prediction of derivatives for a structure-activity relationship study and facilitating an optimization campaign.

Flaviviral proteases recognize positively charged proteins as substrates,[159] hence the identified inhibitors of NS2B-NS3 are also positively charged. This presumably leads to low activity on Dengue 2 virus (DENV2) *in-vitro*. The measured EC_{50} value of 26.8 µM for **427_1** is in a different order of magnitude than the K_i value of 0.09 µM. The reason might be low cellular permeability due to the double positive charge on both aliphatic amino groups. Hence, we propose cellular studies combined with a hit optimization campaign that prioritizes derivatives with higher lipophilicity.

Surprisingly, compound **397_2** shows an EC_{50} value of 47.8 µM, which is superior to the inhibition seen in the cell-free assay (no measured inhibition below 50 µM). This compound showed slight cellular toxicity at concentrations above 10 µM. Hence, we cannot exclude that **397_2** interferes with cellular protein biosynthesis mechanisms leading to apparent DENV2[Pro] inhibition.

According to our dynamic binding hypotheses of inhibitor **427_1**, it would be possible to develop a more lipophilic ligand by extending the structure towards the S4 binding pocket. Hence, we propose to design a T-shaped molecule preserving the contacts in the S1 and S2 subpockets and profiting from interactions in the S4 subpocket as described for the alternative binding mode of **427_1** in ZIKV[Pro] (Figure 26).

Equine Herpesvirus: Targeting the MHC-I − Glycoprotein D Interface for Vaccine Development

Viral infections in animals are both an important topic to improve animal welfare, and also an economic factor, as caused by canceling sports events.[26] Equine herpesviruses 1 and 4 (EHV-1, -4) represent an important threat by infecting horses and other wildlife species, like bears.[27] Currently available vaccines used for preventing EHV infections show low efficacy and do not prevent naturally occurring outbreaks.[6] In order to improve vaccines, more research on understanding viral entry into the host cells is needed. Here, studies on EHV-1 and 4 glycoprotein D (gD) essential for viral entry might support vaccine design. The gD binds to the equine major histocompatibility complex I (MHC-I), allowing the virus to evade the immune system. The mechanism of gD binding to MHC-I remains elusive.

We present the first gD-MHC-I binding model and provide insights into the binding interface. Furthermore, we predict potential hot spot residues and prove their biological importance. The characterization of the predicted gD-MHC-I binding mode contributes to the understanding of how EHV evades the immune system and supports, therefore, the development of more effective and immunogenic vaccines. Moreover, our research might trigger the development of specific therapies, such as small molecular entry inhibitors.

The main limitation of our study is the only partial validation of the model. Out of the whole binding interface, only two residues were subjected to a mutagenesis study. Furthermore, the structure of the gD transmembranous domain and the mechanism of interaction with other herpesviral glycoproteins are unknown, hence the impact of MHC-I binding on the viral entry could not be mechanistically investigated. Therefore, we suggest additional experiments to obtain a full gD-MHC-I structure as the ultimate validation of the binding hypothesis. Preferably, cryoscopic electron microscopy imaging should be used to obtain a holistic picture of all glycoproteins pivotal for viral entry.

Predicting the Susceptibility of Animal Species to COVID-19

COVID-19 pandemic is still ongoing resulting in several million casualties.[5] Despite dynamic vaccine development, several people cannot be vaccinated, furthermore, potential escape variants might occur.[24] COVID-19 is a new disease and more research

is needed for understanding aspects such as virus transmission or immune reaction to the virus. Here, novel animal models might provide additional insights. A rational approach for the identification of susceptible species would improve COVID-19 research substantially.

In this thesis, we have shown that the polymorphisms of angiotensin-converting enzyme 2 (ACE2) orthologs could explain the reported susceptibility differences of animal species to COVID-19. We describe an *in-silico* workflow identifying molecular descriptors in MD simulations of severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) spike protein (S) in complex with ACE2 orthologs to predict animal susceptibility to COVID-19. The workflow enables fast and rational identification of animal species that could serve as potential models of COVID-19 imitating the human course of the disease. On the other hand, we propose gain-of-function mutations in the ACE2 domain that could lead to new susceptible animal models (rat and mouse models). Our work highlights, however, the complexity of the mutational patterns needed for the enhancement of susceptibility. For example, in rats, four mutations (S27T, N30D, I79L, and H353K) are required to enhance the binding to S. We hypothesize that the detailed characterization of the protein-protein binding interface might contribute to the prediction of novel, more infectious SARS-CoV-2 variants.

Due to the highly dynamic nature of coronaviral research, several groups were working in parallel on similar approaches aiming to predict animal susceptibility to COVID-19.[211-216] Most of these approaches are based on a simple sequence comparison of ACE2 orthologs between susceptible and non-susceptible species. Some of the works[212, 213, 216] additionally presented static structure-based approaches using homology modeling. As a result, differences in the S-ACE2 binding interface were shown. The interactions between both proteins, however, have not been analyzed. We believe that an in-depth understanding of S-ACE2 interactions contributes to a more rational selection of COVID-19 susceptible species. Our work was the first to describe an extensive comparison of three-dimensional ACE2 models and to analyze the dynamics of the receptor-spike interface.[217] This study contributes to a mechanistic understanding of ACE2 recognition by SARS-CoV-2 S. More than a half year after our publication, Rodrigues and colleagues independently published an analysis[216] similar

to our work. Their results are comparable with our findings and appear to confirm our results.

Mutations in the receptor-binding domain (RBD) of S remain a limiting factor for our study since they can change the interaction patterns in the ACE2-binding interface. This can lead to a reduction of the predictive power of our model. For example, the N501Y substitution in binding pocket A reported in novel SARS-CoV-2 mutants[218] leads to an additional lipophilic contact between the phenyl rest of RBD-Y501 and the ACE2 interface. We assume that this change diminishes the predictive power of the hydrogen bond count in binding pocket A as a descriptor for COVID-19 susceptibility. Hence, the descriptors should be constantly updated depending on novel reported SARS-CoV-2 mutants to retain the predictive power of our approach. Ideally, our models should be biophysically validated on ACE2 mutants to confirm our predictions.

5.   Conclusions

In this thesis, we provide in-depth insights into different stages of the development of antiviral substances: from the discovery of small molecular protease inhibitors, through the characterization of an epitope for a potential vaccine, to the prediction of suitable animal infection models. We have established easy-to-use and rapid computational workflows targeting biomolecules of emerging viruses in a scenario of scarce data availability.

The importance of protein-protein interactions for the development of antiviral substances is emphasized in the introduced literature. Notably, the inhibition of viral proteases and fusion proteins bears potential for the discovery of highly active antiviral therapeutics. The substrate-protease and protein-protein interfaces selected in this thesis can be considered challenging targets, due to their high conformational flexibility, shallowness of the binding pocket, spatial separation of hot spot regions, and complexity of binding residue networks. In order to overcome these properties, we combined novel approaches such as *PyRod*, *Dynophores*, and *PyContact* with well-established methods as pharmacophore modeling, protein-protein docking, and molecular dynamics simulations.

The novel competitive NS2B-NS3 protease inhibitors presented in this work are the first well-characterized compounds targeting Zika virus protease discovered by pharmacophore modeling. The comparative modeling of closely related proteases of West Nile and Dengue virus 2 enabled us to identify inhibitors with broad antiflaviviral activity. Furthermore, these ligands represent promising starting points for an optimization campaign due to their drug-like properties such as low molecular weight and low micromolar activity.

In the second part of this work, we established the first binding hypothesis of the equine herpesvirus 1 and 4 (EHV-1, 4) glycoprotein D (gD) and its receptor, equine major histocompatibility complex I (MHC-I). To our knowledge, this is the first computational study on this target. The derived binding epitope was extensively characterized and led to the identification of two hot spot residues of the EHV-1 and 4 gD.

In the last part of this thesis, we presented an extensive analysis of severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) spike protein (S) binding interface in complex with the angiotensin-converting enzyme 2 (ACE2) orthologs from

different animal species. The established molecular descriptors enable the prediction of species susceptibility to coronavirus disease 2019 (COVID-19). Our analysis of the S-ACE2 binding epitope represents the first structure-based study on this target. Our methodology is superior to other reported approaches that are based solely on the comparison of the sequences of ACE2 orthologs or static homology models neglecting the dynamic nature of S-ACE2 binding. The advantage of the dynamic approach was shown in the example of canine ACE2; we were able to explain how a buried ACE2 polymorphism A24V (seen in dogs) influences the S binding, rendering dogs non-susceptible to COVID-19 in contrast to ferrets and cats.

We have shown that *in-silico* models support the development of antiviral substances, vaccines, and animal models needed for the research. We are confident that the gained knowledge can be transferred to other viral species, allowing efficient combat against existing and emerging infections. Furthermore, due to the universality of the employed methods, we encourage the application of our workflows to other pathogenic microorganisms, like bacteria or protozoa.

6. Experimental Details

If not else stated, the *in-silico* experiments were performed on the local computational cluster in the Wolber Lab in Berlin (Germany). The molecular dynamics simulations were conducted on water-cooled GeForce RTX 2080 Ti graphics processing units (NVIDIA Corporation, Santa Clara, USA). All other calculations were performed on either AMD Ryzen Threadripper 1950X (Advanced Micro Devices, Santa Clara, USA) or Intel Core i5-8600K (Intel Corporation, Santa Clara, USA) central processing units.

6.1 Development of Small Molecules Targeting NS2B-NS3 Proteases

Homology Modeling and Protein Preparation

Lacking a Dengue 2 virus protease (DENV2[Pro]) structure in its closed state, we constructed a homology model of this protein. The sequence of DENV2[Pro] consistent with the sequence used by our collaborators in the biochemical assay was obtained from UniProt[219] (UniProt-ID: P29991).

The homology model was built based on the DENV3[Pro] X-ray structure with the best resolution (PDB[62]-ID: 3U1J[167]) as a template. The sequence similarity was calculated with MOE 2018.0101 (Chemical Computing Group ULC, Montreal, Canada, Appendix Figure 2). The DENV2[Pro] homology model was constructed using MOE 2018.0101 with integrated GB/VI[88] scoring and a maximum of ten main chain models.

The X-ray crystal structures of the West Nile (WNV[Pro], PDB-ID: 5IDK[157]), Zika (ZIKV[Pro], PDB-ID: 5YOF[156]) proteases were selected for our workflow due to their closed conformation, best resolution, and PDB validation scores.

All proteases were subsequently prepared for further experiments using MOE 2018.0101. The termini of the proteases were capped, and non-resolved residues, co-crystallized water, salt, and ligand molecules were deleted. The proteases were protonated at a pH of 7 and a temperature of 300 K using the Protonate3D[220] function integrated into MOE 2018.0101.

Calculation of Pocket Area

The area of the binding pocket was calculated using POVME 3[190] within a sphere of radius 15 Å centered on the point above the Cα atom of NS3-G153 (-1, 11, 0) to cover the whole competitive binding site of the WNV[Pro].

Molecular Dynamics Simulations

The MD simulations of the proteases were prepared using Maestro 11.7 (Schrödinger, LLC: New York, USA) by solving with the TIP4P water[221] model in a cubic box with 15 Å padding, neutralizing the net system charge with sodium ions, and adjusting the osmotic pressure with 0.15 M sodium chloride to achieve physiological conditions. The MD simulations were parameterized using the OPLS2005 force field[222] and performed with Desmond 5.5[138] on GeForce GTX 1080 Ti graphics processing units (NVIDIA Corporation, Santa Clara, USA). The systems were equilibrated according to the default Desmond protocol. The main simulation runs were performed under periodic boundary conditions as an NPT ensemble (constant particle number, pressure, and temperature). The constant temperature of 300 K was held using the Nose-Hoover thermostat[223, 224] and the constant pressure of 1.01325 bar using Martyna-Tobias-Klein barostat.[225]

The MD simulations were performed over 10 ns yielding 2000 time steps for each replicate. Each protease system was simulated in ten replicates. The trajectories were wrapped and aligned on the first simulation frame and the heavy atoms of the protein backbone using VMD 1.9.3.[226]

Pharmacophore Generation (PyRod)

The pharmacophore models were generated using *PyRod*[128] (in version 0.7.1), a novel software developed in-house and described in detail in the Theoretical Background section. The first 5 ns of the MD simulations were used for the equilibration of the protease-conformations. For the interaction sampling, the last 5 ns (corresponding to 1000 time steps) of the MD simulations were analyzed. The interactions were sampled in a rectangular box of size 22 x 17 x 30 Å with the center on the Oγ atom of the catalytic NS3-S135 to sufficiently cover the competitive binding pocket. The distance of grid points was set to 0.5 Å and the settings for the generation of exclusion volume features representing the binding pocket surface were kept default. The dynamic molecular interaction fields (dMIFs) were automatically transformed to a general pharmacophore model with 20 interaction points per feature type.

The general pharmacophore was subsequently refined by removing interaction points outside the S1 and S2 subpockets and prioritizing the remaining features according to the dMIFs to yield the focused pharmacophore. Finally, a combinatorial

pharmacophore library was created, based on the focused pharmacophore. The combinatorial pharmacophores contained 3-6 independent features (including maximal 3 hydrophobic contacts, 3 aromatic interactions, 3 ionic interactions, and 4 hydrogen bonds). To reduce the number of combinatorial pharmacophores, the cationic interaction to NS3-D129 was preserved in each model. All combinatorial pharmacophores were automatically validated using a Python-based[227] script developed in-house.

Preparation of Libraries Used for Pharmacophore Validation

In order to validate the obtained pharmacophore models, a manually curated library of NS2B-NS3 inhibitors reported in the literature was prepared. The inhibitors included in the library were selected according to their activity (IC_{50} or K_D value lower than 50 μM[11]), competitive binding mode, and ability to establish a cationic interaction with the selectivity residue NS3-D129.[157] All peptides and derivatives were excluded from the library due to lacking drug-likeness of such structures. In total, 17 inhibitors of WNV[Pro] and no ZIKV[Pro] inhibitors were identified.

The library of presumably inactive molecules (decoys) was generated using the DUD-E web application[129] based on the structures of found active molecules. This yielded 667 decoy molecules.

All molecules were protonated using MOE 2018.0101. The starting conformations were generated using Corina 3 (Molecular Networks GmbH Computerchemie, Erlangen, Germany). Based on these starting conformations, the conformation libraries for pharmacophore validation were generated using iCon on "best settings" (maximally 200 conformations per structure, root mean square threshold of 0.8, and energy window of 20.0) implemented in LigandScout 4.2[112, 113]

Virtual Screening and Hit Selection

The validated and selected combinatorial pharmacophores **C_65**, **C_397**, and **C_427** were used for a virtual screening campaign of commercially available libraries containing around 7.6 million compounds (Asinex, Chembridge, ChemDiv, Enamine, Key Organics, LifeChemicals, Maybridge, SPECS, VitasM, each in version from the year 2018). The virtual screening was performed using LigandScout 4.2. The obtained ligand candidates (hits) were subsequently filtered using molecular docking to obtain

plausible binding hypotheses. All hits were docked into the WNV[Pro] and ZIKV[Pro] (PDB-IDs as listed above) using GOLD 5.6.3[97] to generate ten diverse docking poses per compound (root mean square deviation between the docking poses of more than 1.5 Å) with 100% search efficiency. The ASP[101] function was used for scoring and Chemscore[102] for rescoring. The center of the binding sphere with a radius of 10 Å was placed on the Oγ atom of the catalytic NS3-S135. The generated docking poses were subsequently energy minimized in the presence of the protease using the Merck molecular force field (MMFF94)[228-232] implemented in LigandScout 4.2. The minimized structures were scored according to the pharmacophore fit used for the virtual screening. The best-scored docking poses were visually inspected applying the rules described in the Theoretical Background section. The remaining structures were finally simulated using the settings described above in a single 20 ns MD simulation, each. Finally, the ligands showing no global conformational changes upon the visual inspection of the trajectories were selected for the biochemical characterization.

Dynamic Pharmacophores (Dynophores)

To extensively characterize the binding hypothesis of the most active inhibitor **427_1**, a series of MD simulations of **427_1** in complex with WNV[Pro], ZIKV[Pro], and DENV2[Pro] was performed. The MD simulations were calculated with the settings described above over 50 ns in five replicates generating in total 25,000 ligand-protein conformations.

The protease-inhibitor interactions were analyzed with *Dynophore*[73, 127], a software developed in-house and described in the Theoretical Background section, to obtain the spatiotemporal interaction probability clouds and interaction distance.

Compound Purity, Protein Expression, and Protease Assays

The selected compounds were purchased from Enamine, Ukraine (**397_12**) and Chembridge, USA (**397_2**, **397_6**, **427_1**, **427_2**). The purity of the compounds (>94%) was confirmed with an in-house developed method on an HPLC-MS by Lukas Harps in Parr Lab in Berlin (Germany).

The biochemical characterization of the ligands obtained from the virtual screening was performed by our collaborators (ZIKV[Pro] assay- Tim M. Sarter in the Nitsche Lab in Canberra (Australia); WNV[Pro] assay- Rafe Yousef in the Rademann Lab in Berlin

(Germany); DENV2^{Pro} assay- Silke Bergmann and Christoph Arkona in the Rademann Lab in Berlin (Germany). The ZIKV^{Pro} was expressed by Christoph Nitsche as described previously.[233] The WNV^{Pro} and DENV2^{Pro} were expressed by Christoph Arkona similar to the method described in the literature.[234]

All measurements were conducted in triplicates.

The ZIKV^{Pro} assay was performed by pipetting various concentrations of the selected compounds (10 mM stock solutions in DMSO) to the buffer mixture (10 mM Tris-HCl, pH 8.5, 20% v/v glycerol, 1 mM CHAPS) and incubating with ZIKV^{Pro} at a final concentration of 1 nM for 10 min. The measurement was started by adding the fluorescent substrate Bz-Nle-Lys-Lys-Arg-AMC (Biosyntan, Berlin, Germany). The release of the fluorescent label AMC (7-amino-4-methylcoumarin) was observed for 70 s at 460 nm and an excitation wavelength of 360 nm using the fluorophotometer Spectramax M2e (Molecular Devices, San Jose, USA).

The WNV^{Pro} and DENV2^{Pro} assays were performed by pipetting various concentrations of the selected compounds (15 mM stock solutions in DMSO) to the buffer mixture (10 mM MOPS, pH 8, 20% v/v glycerol for WNV^{Pro} and 50 mM Tris-HCl, pH 8.5, 20% v/v glycerol for DENV2^{Pro}) and adding the proteases at a final concentration of 20 nM (WNV^{Pro}) or 800 nM (DENV2^{Pro}). The measurement was initiated by adding the fluorescent substrate Boc-Gly-Lys-Arg-AMC (Bachem, Bubendorf, Switzerland). The release of the fluorescent label AMC (7-amino-4-methylcoumarin) was observed for 600 s at 465 nm and an excitation wavelength of 360 nm using the fluorophotometer infinite M1000 (Tecan, Männedorf, Switzerland).

The initial velocities obtained from the protease assays were obtained from the linear part of the curve slope as a variation of the relative fluorescence per time unit. The IC$_{50}$ values were calculated with Prism 8.2 (GraphPad Prism, La Jolla, USA) using at least seven different inhibitor concentrations. To obtain the K$_i$ values, the IC$_{50}$ values were plotted against the substrate concentrations according to the Cheng-Prusoff equation,[173] returning coefficients of determination (R^2) of: 0.999, 0.951 for the inhibitor **397_2** on ZIKV^{Pro} and WNV^{Pro}, respectively, and 0.980, 0.961, and 0.931 for the inhibitor **427_1** on ZIKV^{Pro}, WNV^{Pro}, and DENV2^{Pro}, respectively.

6.2 Equine Herpesvirus: Targeting the MHC-I – Glycoprotein D Interface for Vaccine Development

Homology Modeling and Protein Preparation

Lacking a structure of the equine major histocompatibility complex I (MHC-I) genotype 3.1, we constructed a homology model of this protein. The sequence of the MHC-I genotype 3.1 consistent with the sequence used by our collaborators in the cell culture was obtained from UniProt (UniProt-ID: Q30483 for the α-chain and UniProt-ID: P30441 for the equine β_2-microglobulin).

The homology model was built based on the X-ray structure of chimeric equine α-chain MHC-I genotype 1.18.7-6 and murine β_2-microglobulin with the best resolution (PDB-ID: 4ZUU[198]) as a template. The sequence similarity was calculated with MOE 2018.0101 (Appendix Figure 6). The equine MHC-I genotype 3.1 homology model was constructed using MOE 2018.0101 with integrated GB/VI scoring and a maximum of ten main chain models. The α-chain and β_2-microglobulin models were assembled into the final homology model in MOE 2018.0101.

The final homology model was subsequently prepared for further experiments using MOE 2018.0101. The termini were capped, and the homology model was protonated at a pH of 7 and a temperature of 300 K using the Protonate3D function integrated into MOE 2018.0101. The final homology model of MHC-I genotype 3.1 was relaxed in a molecular dynamics (MD) simulation as described below.

In order to obtain a physiological state of the MHC-I homology model, a nonapeptide of sequence SDYVKVSNI used by our collaborators was manually fitted into the peptide-binding cleft of MHC-I. To yield the peptide-bound structure, the nonapeptide CTSEEMNAF co-crystallized with the template MHC-I structure (PDB-ID: 4ZUU) was superposed on the coordinates of the relaxed final MHC-I genotype 3.1 homology model. The residues in the template peptide were manually mutated using MOE 2018.0101. The conformations of the side chains were relaxed using the rotamer tool implemented into MOE and energy minimized using the OPLS-AA force field.

Protein-Protein Docking

In the first step, initial protein-protein complexes of glycoprotein D from the equine herpesviruses 1 and 4 (gD EHV-1, PDB-ID: 6SQJ[183] and gD EHV-4, PDB-ID: 6TM8[183]) with peptide-free and peptide-bound MHC-I genotype 3.1 homology model were

prepared using MOE 2018.0101. The initial protein-protein complexes were subsequently used for the protein-protein docking performed with Rosetta 3 suite in version 2018-33.[107, 235] At the beginning of the docking process, the positions of gD and MHC-I were randomized (flags -randomize 1 -randomize 2) and spun (flag -spin). The docking run was performed with default perturbation settings 3 Å for translation and 8° for rotation (flag -dock_pert 3 8).[236] The side chains of both docking partners could rotate around the χ1 and χ2 angles (flags -ex1 -ex2). In order to sufficiently sample the conformational space,[104] 10,000 docking runs per protein-protein complex were performed (flag -nstruct 10000). To reduce the number of possible docking solutions, a flat harmonic distance constraint between the Cα of the key residue MHC-I A173[34, 197] and the gD backbone was applied. The distance between both proteins was set to a default[104] of 0 Å with a standard deviation of 1 Å and a tolerance of 5 Å to force the closest possible contact between both docking partners.

Molecular Dynamics Simulations

The MD simulations of the final MHC-I genotype 3.1 homology model and the docking poses were prepared using Maestro 11.7 (Schrödinger, LLC: New York, USA) by solvating with the SPC water model[237] in a cubic box with 12 Å padding, neutralizing the net system charge with sodium ions, and adjusting the osmotic pressure with 0.15 M sodium chloride to achieve physiological conditions. The MD simulations were parameterized using the OPLS2005 force field and performed with Desmond 5.5. The systems were equilibrated according to the default Desmond protocol. The main simulation runs were performed under periodic boundary conditions as an NPT ensemble (constant particle number, pressure, and temperature). The constant temperature of 300 K was held using the Nose-Hoover thermostat and the constant pressure of 1.01325 bar using Martyna-Tobias-Klein barostat.

The single MD simulation of the final MHC-I genotype 3.1 homology model was performed over 100 ns. The relaxed MHC-I structure was obtained after the root mean square deviation (RMSD) of the protein backbone had reached a stable plateau of 3 Å, suggesting equilibration of the protein.

The docking poses of the gD from EHV-1 and -4 to the MHC-I homology model in the peptide-free and peptide-bound states were simulated over 100 ns, yielding 5000 time steps for each docking pose. A single simulation was performed for each pose for

efficient filtering. The finally selected peptide-bound EHV-1 and -4 gD docking poses were simulated in triplicate to sufficiently characterize the binding interface as described below.

The trajectories of all simulations were wrapped and aligned on the first simulation frame and the heavy atoms of the protein backbone using VMD 1.9.3.

Analysis of Protein-Protein Interactions

The dynamics of protein-protein interfaces of the final gD − MHC-I binding hypotheses were analyzed using *PyContact* 1.0.1[200] in a Python 2.7[227] environment. The detection of the interactions was set to default with a distance cutoff of 5.0 Å, angle of 120.0 Å, and hydrogen bond distance cutoff of 2.5 Å.

The final peptide-bound binding hypotheses were filtered out using an MDAnalysis-based[201, 202] script (in version 0.19.2) to identify docking poses containing the hypothesized crucial contact between gD-D261 and MHC-I-R169. The script was run in a Python 3.6[227] environment. The salt bridge was indicated as a distance of maximal 4.5 Å between the Cγ atom of gD-D261 and the Cζ atom of MHC-I-R169. The script was provided by David Machalz, a colleague from our group.

Mutagenesis and Virus Growth Kinetics Assay

The biological validation of the predicted binding modes was performed by Viviane Kremling under the supervision of Walid Azab in the Osterrieder Lab in Berlin (Germany). The proposed EHV-1 and -4 gD-mutants gD$_{F213A}$ and gD$_{D261N}$ were prepared using the bacterial artificial chromosome (BAC) mutagenesis similar to the described method.[238]

The viral growth kinetics assay was performed as described previously.[239] Briefly, the equine dermal cells were grown at 37°C, infected with the wild-type and gD-mutant variants, and incubated for one hour. The viral particles that did not enter the cells were washed out using citrate treatment and buffer (pH 3, 40 mM citric acid, 10 mM potassium chloride, 135 mM sodium chloride). Subsequently, the buffer was neutralized. The equine dermal cells were washed with phosphate-buffered saline. The cells and the supernatant were collected at 0, 6, 12, 24, 30, and 48 hours after incubation. Until the collection was finalized, the first samples were frozen at -80°C. The determination of the virus titers was performed by diluting the supernatant

samples, infecting the equine dermal cells, and counting plaques after one or two days of incubation.

6.3   Predicting the Susceptibility of Animal Species to COVID-19

Homology Modeling and Protein Preparation

Lacking the structures of the animal orthologs of the angiotensin-converting enzyme 2 (ACE2), we constructed homology models of these proteins. The sequences of the ACE2 orthologs were kindly provided by our collaborators from the Osterrieder Lab in Berlin (Germany) and can be found in the RefSeq[240] or GenBank[241] database (Table 11, Appendix Figure 12).

Table 11. The UniProt-IDs of the protein sequences used for the homology modeling.

Animal species	RefSeq- or GenBank-ID
Cat (*Felis catus*)	XP_023104564.1 (RefSeq)
Dog (*Canis lupus familiaris*)	NP_001158732.1 (RefSeq)
Ferret (*Mustela putorius*)	NP_001297119.1 (RefSeq)
Mouse (*Mus musculus*)	NP_081562.2 (RefSeq)
Rat (*Rattus norvegicus*)	NP_001012006.1 (RefSeq)
Syrian hamster (*Mesocricetus auratus*)	XP_005074266.1 (RefSeq)
Red squirrel (*Sciurus vulgaris*)	-*
Chinese hamster (*Cricteulus griseus*)	XP_003503283.1 (RefSeq)
Campbell's dwarf hamster (*Phodopus campbelli*)	ACT66274.1 (GenBank)

* The sequence marked with an asterisk is neither listed in the RefSeq nor GenBank database.

The homology models were built based on the first available X-ray structure of the human ACE2 in complex with the residue binding domain (RBD) of severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) (PDB-ID: 6M0J[181]) as a template. The sequence similarity was calculated with MOE 2018.0101 (Appendix Figure 12). All homology models were constructed using MOE 2018.0101 with integrated GB/VI scoring and a maximum of ten main chain models. The ACE2 models were assembled with the RBD in MOE 2018.0101 to generate the final homology models. The catalytic center of ACE2 and atomic clashes between the ACE2 homology

models and RBD were relaxed using the rotamer tool integrated into MOE and energy minimized with the OPLS-AA force field.

The final homology models were subsequently prepared for further experiments using MOE 2018.0101. The termini of the RBD were capped and the homology models were protonated at a pH of 7 and a temperature of 300 K using the Protonate3D function integrated into MOE 2018.0101.

Molecular Dynamics Simulations

The MD simulations of the ACE2 homology models and the ACE2-RBD complexes were prepared using Maestro 11.7 (Schrödinger, LLC: New York, USA) by solvating with the SPC water model in a cubic box with 12 Å padding, neutralizing the net system charge with sodium ions, and adjusting the osmotic pressure with 0.15 M sodium chloride to achieve physiological conditions. The MD simulations were parameterized using the OPLS2005 force field and performed with Desmond 5.5. The systems were equilibrated according to the default Desmond protocol. The main simulation runs were performed under periodic boundary conditions as an NPT ensemble (constant particle number, pressure, and temperature). The constant temperature of 300 K was held using the Nose-Hoover thermostat and the constant pressure of 1.01325 bar using Martyna-Tobias-Klein barostat.

All MD simulations were performed in five replicates, over 100 ns each, yielding 2000 time steps per replicate. The trajectories of all simulations were wrapped and aligned on the first simulation frame and heavy atoms of the protein backbone using VMD 1.9.3.

Trajectory Analysis

In the first step, all trajectories of the ACE2-RBD complexes were visually inspected to identify global conformational changes (e.g. backbone movements or dissociation events). Subsequently, all trajectories were analyzed using MDAnalysis-based scripts in a Python 3.7 environment provided by Trung Ngoc Nguyen, a colleague from our group. The first 10 ns of the simulations were used for the equilibration of the system. For the analysis, the last 90 ns of the simulations were processed using pandas 0.25.3.[242] The data was plotted using matplotlib 3.1.1[243] and seaborn 0.9.0.[244]

7.   Summary

Viral infections represent an old threat to global health, with multiple epidemics and pandemics in the history of mankind. Despite several advances in the development of antiviral substances and vaccines, many viral species are still not targeted. Additionally, new viral species emerge, posing a menace without precedent to humans and animals and causing fatalities, disabilities, environmental harm, and economic losses.

In this thesis, we present rational modeling approaches for targeting specific protease-substrate and protein-protein interactions pivotal for the viral replication cycle. Over the course of this work, antiviral research is supported beginning with the development of small molecular antiviral substances, going through the modeling of a potential immunogenic epitope for vaccine development, towards the establishment of descriptors for susceptibility of animals to a viral infection (Figure 47). Notably, all the research was done under scarce data availability, highlighting the predictive power of computational methods and complementarity between *in-silico* and *in-vitro* or *in-vivo* methods.

In the first part of this thesis, the discovery of two novel chemotypes of small molecular competitive flaviviral NS2B-NS3 protease inhibitors is described. Despite several issues, such as shallowness, hydrophilicity, and large size of the binding pocket, it was possible to develop pharmacophore models suitable for a virtual screening campaign. To overcome these protease-specific challenges, we used a novel software, *PyRod*, to benefit from tracking interactions of water molecules over the course of molecular dynamics simulations. The inhibitory activity of suggested ligands was shown in a biochemical assay and the cell-based replicon model.

In the next step, we characterized the binding epitope of equine herpesvirus 1 and 4 (EHV-1, 4) glycoprotein D (gD) and equine major histocompatibility complex I (MHC-I). We provided the first insights into the binding interface and characterized two hot spot residues. It was shown by our collaborators that the predicted loss of function mutations hinder viral replication in a cell-based model. The identification of the pivotal gD epitope could contribute to the development of novel vaccines.

Finally, we established molecular descriptors derived from molecular dynamics simulations of angiotensin-converting enzyme 2 (ACE2) orthologs from different animal species in complex with the spike protein (S) of a novel severe acute respiratory

syndrome-related coronavirus 2 (SARS-CoV-2). Due to a unique structure-based approach, we were able to rationally explain differences in the susceptibility of animal species to coronavirus disease 2019 (COVID-19) and successfully predict which animal species can serve as suitable models to study the infection.
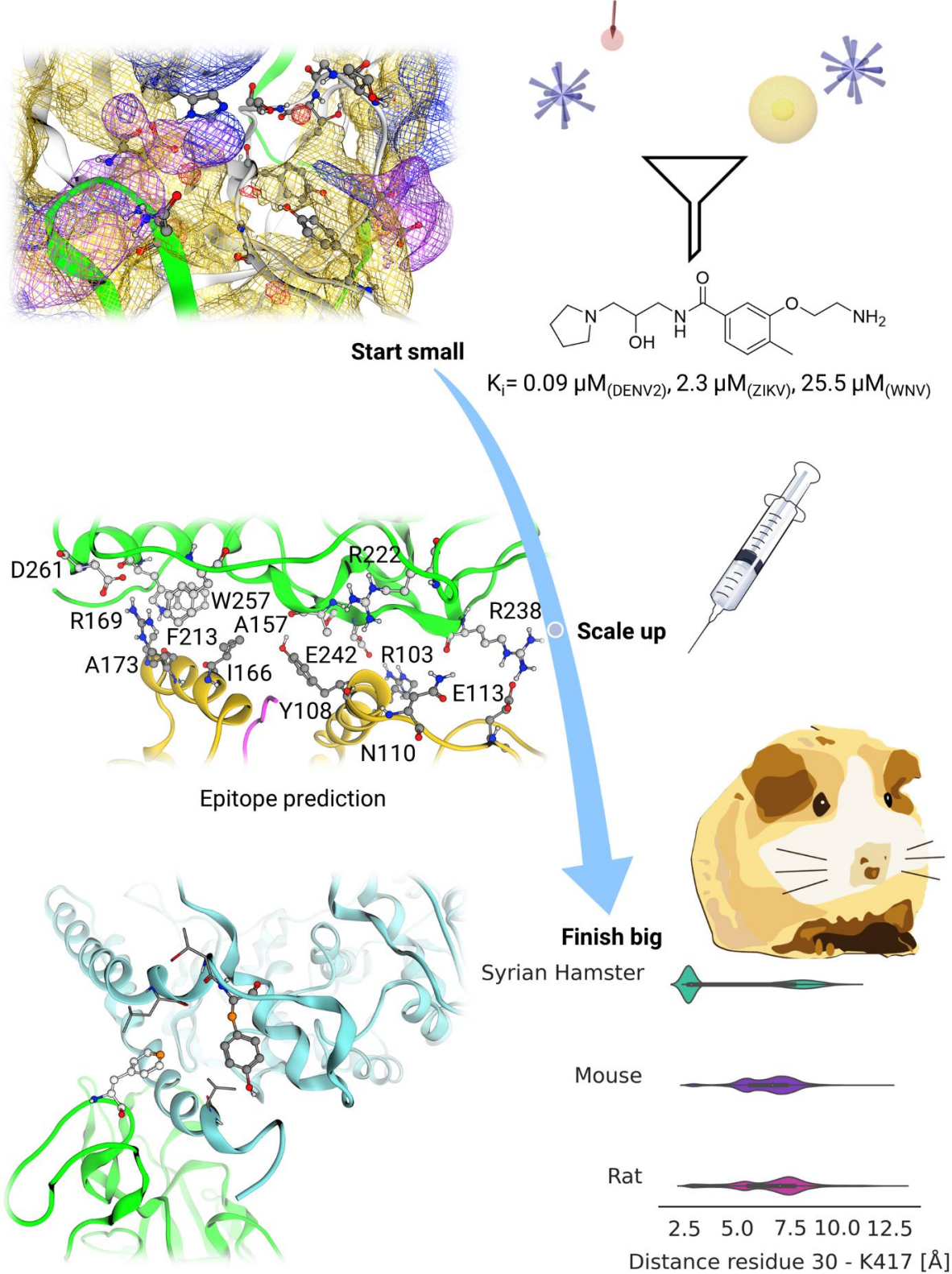
Figure 47. Graphical summary of the thesis.

Zusammenfassung

Virusinfektionen stellen eine alte Bedrohung für die globale Gesundheit dar, mit unzähligen Epidemien und Pandemien in der Geschichte der Menschheit. Trotz zahlreicher Fortschritte bei der Entwicklung von antiviralen Substanzen und Impfstoffen werden viele Virusarten noch immer nicht bekämpft. Darüber hinaus tauchen immer wieder neue Virusarten auf, die eine noch nie dagewesene Bedrohung für Mensch und Tier darstellen und zu Todesfällen, Behinderungen, Umweltschäden und wirtschaftlichen Verlusten führen.

In dieser Arbeit stellen wir rationale Modellierungsansätze für die gezielte Beeinflussung spezifischer Substrat-Protease und Protein-Protein-Interaktionen vor, die für den viralen Replikationszyklus entscheidend sind. Im Verlauf dieser Arbeit wird die antivirale Forschung unterstützt, beginnend mit der Entwicklung kleinmolekularer antiviraler Substanzen, über die Charakterisierung von einem potenziell immunogenen Epitop für die Entwicklung neuen Impfstoffen bis hin zur Etablierung von Deskriptoren für die Anfälligkeit von Tieren für eine virale Erkrankung (Abbildung 47). Dies betont die Vorhersagekraft von Berechnungsmethoden und die Komplementarität zwischen *In-silico-* und *In-vitro-* oder *In-vivo*-Methoden.

Im ersten Teil dieser Arbeit wird die Entdeckung zweier neuartiger Chemotypen von kleinmolekularen, kompetitiven flaviviralen NS2B-NS3-Proteaseinhibitoren beschrieben. Trotz verschiedener Herausforderungen wie flacher Geometrie, Hydrophilie und Größe der Bindungstasche war es möglich, dreidimensionale Pharmakophormodelle zu entwickeln, die für virtuelles Screening geeignet waren. Um die spezifischen Eigenschaften von Proteasen adäquat zu adressieren, wurde eine neuartige Software, *PyRod*, verwendet, um die Solvatisierung im Verlauf der Molekulardynamiksimulationen zu verfolgen. Die inhibitorische Wirkung der vorgeschlagenen Liganden wurde in einem biochemischen Assay und im zellbasierten Replikonmodell nachgewiesen.

Im nächsten Schritt charakterisierten wir das Bindungsepitop des Glykoproteins D (gD) des equinen Herpesvirus 1 und 4 (EHV-1, 4) und des equinen Haupthistokompatibilitätskomplexes I (MHC-I). Plausible Hypothesen zum Bindungsinterface zwischen den beiden Proteinen konnten gewonnen werden. Zwei wichtige Aminosäuren für diese Bindungen wurden vorhergesagt. Die daraus folgenden von uns vorgeschlagenen loss-of-function Mutationen verhindern die virale Replikation in einem zellbasierten Modell

verhindern, wie experimentell von unseren Kooperationspartnern gezeigt werden konnte. Die Identifizierung des entscheidenden gD-Epitops könnte zur Entwicklung neuer Impfstoffe beitragen.

Schließlich erstellten wir molekulare Deskriptoren, die aus Molekular-dynamiksimulationen von Angiotensin-konvertierendem Enzym 2 (ACE2) -Orthologen aus verschiedenen Tierarten im Komplex mit dem Spike-Protein (S) eines neuartigen schweren akuten respiratorischen Syndroms – Coronavirus 2 (SARS-CoV-2) stammen. Dank eines neu entwickelten, strukturbasierten Ansatzes konnten wir Unterschiede in der Anfälligkeit von Tierarten für die Coronavirus-Krankheit 2019 (COVID-19) rational erklären und erfolgreich vorhersagen, welche Tierarten als geeignete Modelle zur Untersuchung der Infektion dienen können.

Von einer kleinmolekularen Substanz

K_i= 0.09 μM$_{(DENV2)}$, 2.3 μM$_{(ZIKV)}$, 25.5 μM$_{(WNV)}$

D261
R222
R169
W257
A157
R238
F213
A173
E242 R103
I166
Y108 E113
N110

Über Epitopvorhersage

Zu Vorhersage von Tiermodellen

Syrian Hamster

Mouse

Rat

2.5   5.0   7.5 10.0 12.5

Distance residue 30 - K417 [Å]
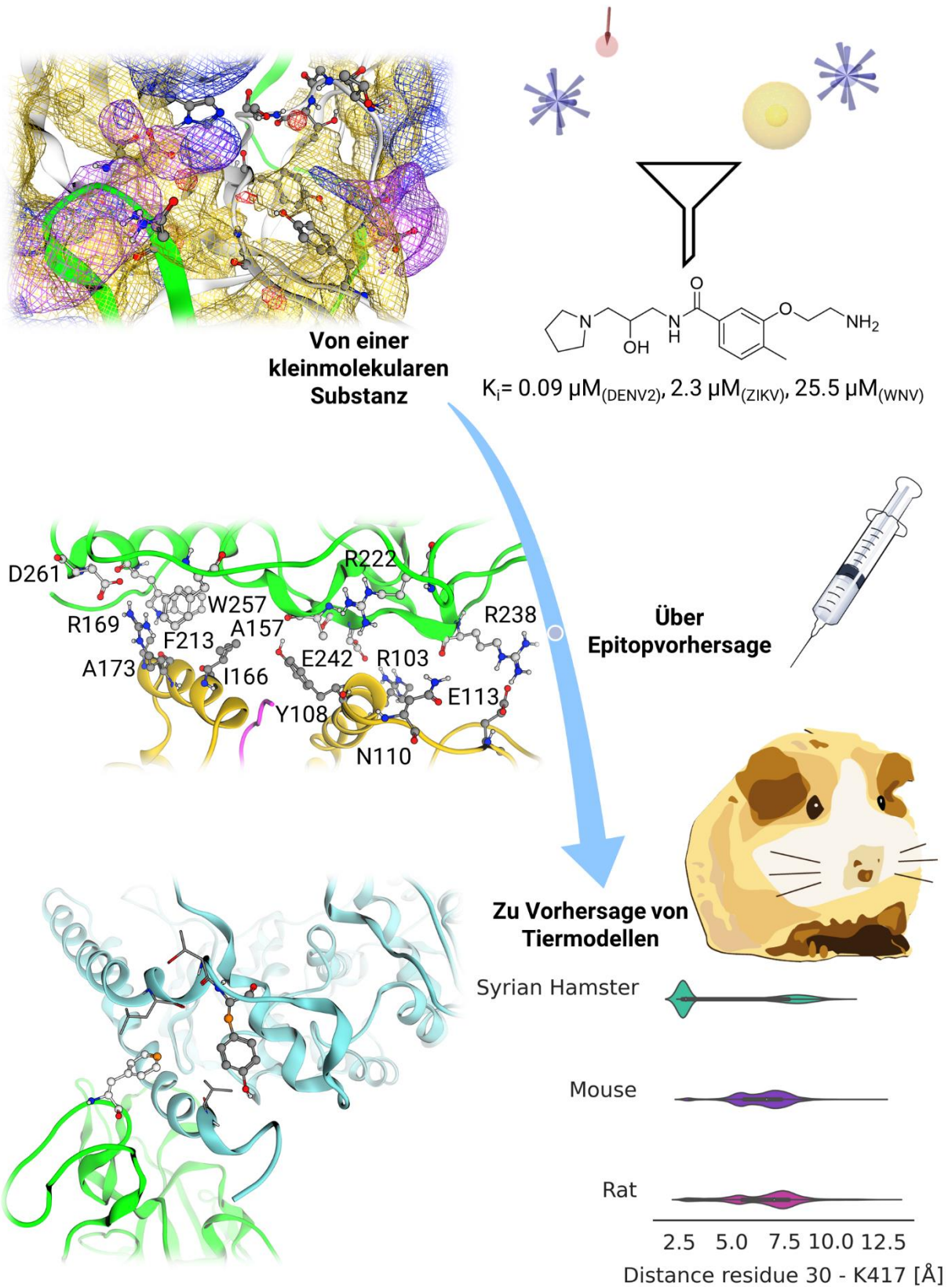
Abbildung 47. Grafische Zusammenfassung der Doktorarbeit.

8. References

1.   Campos, G. S.; Bandeira, A. C.; Sardi, S. I., Zika Virus Outbreak, Bahia, Brazil. *Emerging Infect. Dis.* **2015,** *21* (10), 1885-1886.

2.   Oliveira Melo, A. S.; Malinger, G.; Ximenes, R.; Szejnfeld, P. O.; Alves Sampaio, S.; Bispo de Filippis, A. M., Zika Virus Intrauterine Infection Causes Fetal Brain Abnormality and Microcephaly: Tip of the Iceberg? *Ultrasound Obstet. Gynecol.* **2016,** *47* (1), 6-7.

3.   Attaran, A., Zika Virus and the 2016 Olympic Games. *Lancet Infect. Dis.* **2016,** *16* (9), 1001-1003.

4.   Dick, G. W. A.; Kitchen, S. F.; Haddow, A. J., Zika Virus (I). Isolations and Serological Specificity. *Trans. R. Soc. Trop. Med. Hyg.* **1952,** *46* (5), 509-520.

5.   Dong, E.; Du, H.; Gardner, L., An Interactive Web-Based Dashboard to Track COVID-19 in Real Time. *Lancet Infect. Dis.* **2020,** *20* (5), 533-534.

6.   Lunn, D. P.; Davis-Poynter, N.; Flaminio, M. J.; Horohov, D. W.; Osterrieder, K.; Pusterla, N.; Townsend, H. G., Equine Herpesvirus-1 Consensus Statement. *J. Vet. Intern. Med.* **2009,** *23* (3), 450-61.

7.   Huremović, D., Brief History of Pandemics (Pandemics Throughout History). In *Psychiatry of Pandemics*, Huremović, D., Ed. Springer: Cham, **2019**; pp 7−35.

8.   Zhou, P.; Yang, X. L.; Wang, X. G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H. R.; Zhu, Y.; Li, B.; Huang, C. L.; Chen, H. D.; Chen, J.; Luo, Y.; Guo, H.; Jiang, R. D.; Liu, M. Q.; Chen, Y.; Shen, X. R.; Wang, X.; Zheng, X. S.; Zhao, K.; Chen, Q. J.; Deng, F.; Liu, L. L.; Yan, B.; Zhan, F. X.; Wang, Y. Y.; Xiao, G. F.; Shi, Z. L., A Pneumonia Outbreak Associated with a New Coronavirus of Probable Bat Origin. *Nature* **2020,** *579* (7798), 270-273.

9.   Iwanowski, D., Über die Mosaikkrankheit der Tabakspflanze. *Bull. Acad. Imp. Sci. Saint-Pétersbourg* **1892,** *35*, 67−70.

10.   Iwanowski, D., Über die Mosaikkrankheit der Tabakspflanze. *Zeitschrift für Pflanzenkrankheiten* **1903,** *13* (1), 1-41.

11.   Boldescu, V.; Behnam, M. A. M.; Vasilakis, N.; Klein, C. D., Broad-Spectrum Agents for Flaviviral Infections: Dengue, Zika and Beyond. *Nat. Rev. Drug Discov.* **2017,** *16* (8), 565-586.

12.   Collins, M. H.; Metz, S. W., Progress and Works in Progress: Update on Flavivirus Vaccine Development. *Clin. Ther.* **2017,** *39* (8), 1519-1536.

13.   Paules, C. I.; Fauci, A. S., Yellow Fever - Once Again on the Radar Screen in the Americas. *N. Engl. J. Med.* **2017,** *376* (15), 1397-1399.

14.   Leyssen, P.; De Clercq, E.; Neyts, J., Perspectives for the Treatment of Infections with Flaviviridae. *Clin. Microbiol. Rev.* **2000,** *13* (1), 67-82.

15.   Cunha, A. J.; de Magalhaes-Barbosa, M. C.; Lima-Setta, F.; Medronho, R. A.; Prata-Barbosa, A., Microcephaly Case Fatality Rate Associated with Zika Virus Infection in Brazil: Current Estimates. *Pediatr. Infect. Dis. J.* **2017,** *36* (5), 528-530.

16.   Theiler, M.; Smith, H. H., The Use of Yellow Fever Virus Modified by In Vitro Cultivation for Human Immunization. *J. Exp. Med.* **1937,** *65* (6), 787-800.

17.   Srivastava, A. K.; Putnak, J. R.; Lee, S. H.; Hong, S. P.; Moon, S. B.; Barvir, D. A.; Zhao, B.; Olson, R. A.; Kim, S.-O.; Yoo, W.-D.; Towle, A. C.; Vaughn, D. W.; Innis, B. L.; Eckels, K. H., A Purified Inactivated Japanese Encephalitis Virus Vaccine Made in Vero Cells. *Vaccine* **2001,** *19* (31), 4557-4565.

18.   Zent, O.; Banzhoff, A.; Hilbert, A. K.; Meriste, S.; Słuzewski, W.; Wittermann, C., Safety, Immunogenicity and Tolerability of a New Pediatric Tick-Borne Encephalitis (TBE) Vaccine, Free of Protein-Derived Stabilizer. *Vaccine* **2003,** *21* (25-26), 3584-3592.

19.   Guy, B.; Barrere, B.; Malinowski, C.; Saville, M.; Teyssou, R.; Lang, J., From Research to Phase III: Preclinical, Industrial and Clinical Development of the Sanofi Pasteur Tetravalent Dengue Vaccine. *Vaccine* **2011,** *29* (42), 7229-41.

20.   Hadinegoro, S. R.; Arredondo-Garcia, J. L.; Capeding, M. R.; Deseda, C.; Chotpitayasunondh, T.; Dietze, R.; Muhammad Ismail, H. I.; Reynales, H.; Limkittikul, K.; Rivera-Medina, D. M.; Tran, H. N.; Bouckenooghe, A.; Chansinghakul, D.; Cortes, M.; Fanouillere, K.; Forrat, R.; Frago, C.; Gailhardou, S.; Jackson, N.; Noriega, F.; Plennevaux, E.; Wartel, T. A.; Zambrano, B.; Saville, M.; Group, C.-T. D. V. W., Efficacy and Long-Term Safety of a Dengue Vaccine in Regions of Endemic Disease. *N. Engl. J. Med.* **2015,** *373* (13), 1195-206.

21.   Li, L. Q.; Huang, T.; Wang, Y. Q.; Wang, Z. P.; Liang, Y.; Huang, T. B.; Zhang, H. Y.; Sun, W.; Wang, Y., COVID-19 Patients' Clinical Characteristics, Discharge Rate, and Fatality Rate of Meta-Analysis. *J. Med. Virol.* **2020,** *92* (6), 577-583.

22.   Imam, Z.; Odish, F.; Gill, I.; O'Connor, D.; Armstrong, J.; Vanood, A.; Ibironke, O.; Hanna, A.; Ranski, A.; Halalau, A., Older Age and Comorbidity Are Independent Mortality Predictors in a Large Cohort of 1305 COVID-19 Patients in Michigan, United States. *J. Intern. Med.* **2020,** *288* (4), 469-476.

23.   Yanez, N. D.; Weiss, N. S.; Romand, J. A.; Treggiari, M. M., COVID-19 Mortality Risk for Older Men and Women. *BMC Public Health* **2020,** *20* (1), 1742.

24.   Thompson, R. N.; Hill, E. M.; Gog, J. R., SARS-CoV-2 Incidence and Vaccine Escape. *Lancet Infect. Dis.* **2021,** *21* (7), 913-914.

25.   Osterrieder, N.; Bertzbach, L. D.; Dietert, K.; Abdelgawad, A.; Vladimirova, D.; Kunec, D.; Hoffmann, D.; Beer, M.; Gruber, A. D.; Trimpert, J., Age-Dependent Progression of SARS-CoV-2 Infection in Syrian Hamsters. *Viruses* **2020,** *12* (7), 779.

26.   Fédération Equestre Internationale, Fei Cancels European Mainland Events Due to EHV-1 (Neurological Form). Online press release (accessed 2021-08-02): inside.fei.org/content/fei-cancels-european-mainland-events-due-ehv-1-neurological-form-0, **2021**.

27.   Azab, W.; Dayaram, A.; Greenwood, A. D.; Osterrieder, N., How Host Specific Are Herpesviruses? Lessons from Herpesviruses Infecting Wild and Endangered Mammals. *Annu. Rev. Virol.* **2018,** *5* (1), 53-68.

28.   Stanley, W. M.; Lauffer, M. A., Disintegration of Tobacco Mosaic Virus in Urea Solutions. *Science* **1939,** *89* (2311), 345-7.

29.   Moreira, D.; Lopez-Garcia, P., Ten Reasons to Exclude Viruses from the Tree of Life. *Nat. Rev. Microbiol.* **2009,** *7* (4), 306-11.

30.   International Committee on Taxonomy of Viruses Executive Committee, The New Scope of Virus Taxonomy: Partitioning the Virosphere Into 15 Hierarchical Ranks. *Nat. Microbiol.* **2020,** *5* (5), 668-674.

31.   Lefkowitz, E. J.; Dempsey, D. M.; Hendrickson, R. C.; Orton, R. J.; Siddell, S. G.; Smith, D. B., Virus Taxonomy: the Database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res.* **2018,** *46* (D1), D708-D717.

32.   Baltimore, D., Expression of Animal Virus Genomes. *Bacteriol. Rev.* **1971,** *35* (3), 235-241.

33.   Sasaki, M.; Hasebe, R.; Makino, Y.; Suzuki, T.; Fukushi, H.; Okamoto, M.; Matsuda, K.; Taniyama, H.; Sawa, H.; Kimura, T., Equine Major Histocompatibility Complex Class I Molecules Act as Entry Receptors That Bind to Equine Herpesvirus-1 Glycoprotein D. *Genes Cells* **2011,** *16* (4), 343-57.

34.   Azab, W.; Harman, R.; Miller, D.; Tallmadge, R.; Frampton, A. R., Jr.; Antczak, D. F.; Osterrieder, N., Equid Herpesvirus Type 4 Uses a Restricted Set of Equine Major

Histocompatibility Complex Class I Proteins as Entry Receptors. *J. Gen. Virol.* **2014,** *95* (Pt 7), 1554-63.

35.   Goodman, L. B.; Loregian, A.; Perkins, G. A.; Nugent, J.; Buckles, E. L.; Mercorelli, B.; Kydd, J. H.; Palu, G.; Smith, K. C.; Osterrieder, N.; Davis-Poynter, N., A Point Mutation in a Herpesvirus Polymerase Determines Neuropathogenicity. *PLoS Pathog.* **2007,** *3* (11), e160.

36.   Waxman, L.; Darke, P. L., The Herpesvirus Proteases as Targets for Antiviral Chemotherapy. *Antivir. Chem. Chemother.* **2000,** *11* (1), 1-22.

37.   Granzow, H.; Klupp, B. G.; Fuchs, W.; Veits, J.; Osterrieder, N.; Mettenleiter, T. C., Egress of Alphaherpesviruses: Comparative Ultrastructural Study. *J. Virol.* **2001,** *75* (8), 3675-84.

38.   Simmonds, P.; Becher, P.; Bukh, J.; Gould, E. A.; Meyers, G.; Monath, T.; Muerhoff, S.; Pletnev, A.; Rico-Hesse, R.; Smith, D. B.; Stapleton, J. T.; Ictv Report, C., ICTV Virus Taxonomy Profile: Flaviviridae. *J. Gen. Virol.* **2017,** *98* (1), 2-3.

39.   Hasan, S. S.; Sevvana, M.; Kuhn, R. J.; Rossmann, M. G., Structural Biology of Zika Virus and Other Flaviviruses. *Nat. Struct. Mol. Biol.* **2018,** *25* (1), 13-20.

40.   Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R., Crystal Structure of SARS-CoV-2 Main Protease Provides a Basis for Design of Improved Alpha-Ketoamide Inhibitors. *Science* **2020,** *368* (6489), 409-412.

41.   Bestle, D.; Heindl, M. R.; Limburg, H.; Van Lam van, T.; Pilgram, O.; Moulton, H.; Stein, D. A.; Hardes, K.; Eickmann, M.; Dolnik, O.; Rohde, C.; Klenk, H. D.; Garten, W.; Steinmetzer, T.; Bottcher-Friebertshauser, E., TMPRSS2 and Furin Are Both Essential for Proteolytic Activation of SARS-CoV-2 in Human Airway Cells. *Life Sci. Alliance* **2020,** *3* (9).

42.   Mukhopadhyay, S.; Kuhn, R. J.; Rossmann, M. G., A Structural Perspective of the Flavivirus Life Cycle. *Nat. Rev. Microbiol.* **2005,** *3* (1), 13-22.

43.   De Clercq, E., Strategies in the Design of Antiviral Drugs. *Nat. Rev. Drug. Discov.* **2002,** *1* (1), 13-25.

44.   Hamre, D.; Bernstein, J.; Donovick, R., Activity of p-Aminobenzaldehyde, 3-Thiosemicarbazone on Vaccinia Virus in the Chick Embryo and in the Mouse. *Exp. Biol. Med.* **1950,** *73* (2), 275-278.

45.   Herrmann, E. C., Plaque Inhibition Test for Detection of Specific Inhibitors of DNA Containing Viruses. *Exp. Biol. Med.* **1961,** *107* (1), 142-145.

46.   Herrmann, E. C., Jr.; Gabliks, J.; Engle, C.; Perlman, P. L., Agar Diffusion Method for Detection and Bioassay of Antiviral Antibiotics. *Proc. Soc. Exp. Biol. Med.* **1960,** *103*, 625-8.

47.   Elion, G. B.; Furman, P. A.; Fyfe, J. A.; de Miranda, P.; Beauchamp, L.; Schaeffer, H. J., Selectivity of Action of an Antiherpetic Agent, 9-(2-Hydroxyethoxymethyl) Guanine. *Proc. Natl. Acad. Sci. USA* **1977,** *74* (12), 5716-20.

48.   Jensen, E. M.; Force, E. E.; Unger, J. B., Inhibitory Effect of Ammonium Ions on Influenza Virus in Tissue Culture. *Proc. Soc. Exp. Biol. Med.* **1961,** *107*, 447-51.

49.   Jensen, E. M.; Liu, O. C., Inhibitory Effect of Simple Aliphatic Amines on Influenza Virus in Tissue Culture. *Proc. Soc. Exp. Biol. Med.* **1963,** *112*, 456-9.

50.   Davies, W. L.; Grunert, R. R.; Haff, R. F.; McGahen, J. W.; Neumayer, E. M.; Paulshock, M.; Watts, J. C.; Wood, T. R.; Hermann, E. C.; Hoffmann, C. E., Antiviral Activity of 1-Adamantanamine (Amantadine). *Science* **1964,** *144* (3620), 862-3.

51.   Meindl, P.; Bodo, G.; Palese, P.; Schulman, J.; Tuppy, H., Inhibition of Neuraminidase Activity by Derivatives of 2-Deoxy-2,3-dehydro-N-acetylneuraminic Acid. *Virology* **1974,** *58* (2), 457-463.

52.   von Itzstein, M.; Wu, W. Y.; Kok, G. B.; Pegg, M. S.; Dyason, J. C.; Jin, B.; Van Phan, T.; Smythe, M. L.; White, H. F.; Oliver, S. W.; et al., Rational Design of Potent Sialidase-Based Inhibitors of Influenza Virus Replication. *Nature* **1993,** *363* (6428), 418-23.

53.   Barre-Sinoussi, F.; Chermann, J. C.; Rey, F.; Nugeyre, M. T.; Chamaret, S.; Gruest, J.; Dauguet, C.; Axler-Blin, C.; Vezinet-Brun, F.; Rouzioux, C.; Rozenbaum, W.; Montagnier, L., Isolation of a T-Lymphotropic Retrovirus From a Patient at Risk for Acquired Immune Deficiency Syndrome (AIDS). *Science* **1983,** *220* (4599), 868-71.

54.   Mitsuya, H.; Weinhold, K. J.; Furman, P. A.; St Clair, M. H.; Lehrman, S. N.; Gallo, R. C.; Bolognesi, D.; Barry, D. W.; Broder, S., 3'-Azido-3'-deoxythymidine (BW A509U): An Antiviral Agent That Inhibits the Infectivity and Cytopathic Effect of Human T-Lymphotropic Virus Type III/Lymphadenopathy-Associated Virus In Vitro. *Proc. Natl. Acad. Sci. USA* **1985,** *82* (20), 7096-100.

55.   Roberts, N. A.; Martin, J. A.; Kinchington, D.; Broadhurst, A. V.; Craig, J. C.; Duncan, I. B.; Galpin, S. A.; Handa, B. K.; Kay, J.; Krohn, A.; et al., Rational Design of Peptide-Based HIV Proteinase Inhibitors. *Science* **1990,** *248* (4953), 358-61.

56.   Lai, Y. T.; Wang, T.; O'Dell, S.; Louder, M. K.; Schon, A.; Cheung, C. S. F.; Chuang, G. Y.; Druz, A.; Lin, B.; McKee, K.; Peng, D.; Yang, Y.; Zhang, B.; Herschhorn, A.; Sodroski,

J.; Bailer, R. T.; Doria-Rose, N. A.; Mascola, J. R.; Langley, D. R.; Kwong, P. D., Lattice Engineering Enables Definition of Molecular Features Allowing for Potent Small-molecule Inhibition of HIV-1 Entry. *Nat. Commun.* **2019,** *10* (1), 47.

57.   Choo, Q. L.; Kuo, G.; Weiner, A. J.; Overby, L. R.; Bradley, D. W.; Houghton, M., Isolation of a cDNA Clone Derived From a Blood-Borne Non-A, Non-B Viral Hepatitis Genome. *Science* **1989,** *244* (4902), 359-62.

58.   Malcolm, B. A.; Liu, R.; Lahser, F.; Agrawal, S.; Belanger, B.; Butkiewicz, N.; Chase, R.; Gheyas, F.; Hart, A.; Hesk, D.; Ingravallo, P.; Jiang, C.; Kong, R.; Lu, J.; Pichardo, J.; Prongay, A.; Skelton, A.; Tong, X.; Venkatraman, S.; Xia, E.; Girijavallabhan, V.; Njoroge, F. G., SCH 503034, A Mechanism-Based Inhibitor of Hepatitis C Virus NS3 Protease, Suppresses Polyprotein Maturation and Enhances the Antiviral Activity of Alpha Interferon in Replicon Cells. *Antimicrob. Agents Chemother.* **2006,** *50* (3), 1013-20.

59.   Perni, R. B.; Almquist, S. J.; Byrn, R. A.; Chandorkar, G.; Chaturvedi, P. R.; Courtney, L. F.; Decker, C. J.; Dinehart, K.; Gates, C. A.; Harbeson, S. L.; Heiser, A.; Kalkeri, G.; Kolaczkowski, E.; Lin, K.; Luong, Y. P.; Rao, B. G.; Taylor, W. P.; Thomson, J. A.; Tung, R. D.; Wei, Y.; Kwong, A. D.; Lin, C., Preclinical Profile of VX-950, a Potent, Selective, and Orally Bioavailable Inhibitor of Hepatitis C Virus NS3-4A Serine Protease. *Antimicrob. Agents Chemother.* **2006,** *50* (3), 899-909.

60.   Sofia, M. J.; Bao, D.; Chang, W.; Du, J.; Nagarathnam, D.; Rachakonda, S.; Reddy, P. G.; Ross, B. S.; Wang, P.; Zhang, H. R.; Bansal, S.; Espiritu, C.; Keilman, M.; Lam, A. M.; Steuer, H. M.; Niu, C.; Otto, M. J.; Furman, P. A., Discovery of a β-d-2'-Deoxy-2'-α-fluoro-2'-β-C-methyluridine Nucleotide Prodrug (PSI-7977) for the Treatment of Hepatitis C Virus. *J. Med. Chem.* **2010,** *53* (19), 7202-18.

61.   Callaway, E., 'It Will Change Everything': Deepmind's AI Makes Gigantic Leap in Solving Protein Structures. *Nature* **2020,** *588* (7837), 203-204.

62.   Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res.* **2000,** *28* (1), 235-42.

63.   Lawson, C. L.; Patwardhan, A.; Baker, M. L.; Hryc, C.; Garcia, E. S.; Hudson, B. P.; Lagerstedt, I.; Ludtke, S. J.; Pintilie, G.; Sala, R.; Westbrook, J. D.; Berman, H. M.; Kleywegt, G. J.; Chiu, W., EMDataBank Unified Data Resource for 3DEM. *Nucleic Acids Res.* **2016,** *44* (D1), D396-403.

64.   Bunkoczi, G.; Echols, N.; McCoy, A. J.; Oeffner, R. D.; Adams, P. D.; Read, R. J., Phaser.MRage: Automated Molecular Replacement. *Acta Crystallogr. D Biol. Crystallogr.* **2013,** *69* (Pt 11), 2276-86.

65.   Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallogr. D Biol. Crystallogr.* **2010,** *66* (Pt 2), 213-21.

66.   Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and Development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **2010,** *66* (Pt 4), 486-501.

67.   Wang, H. W.; Wang, J. W., How Cryo-Electron Microscopy and X-Ray Crystallography Complement Each Other. *Protein Sci.* **2017,** *26* (1), 32-39.

68.   Shi, Y., A Glimpse of Structural Biology Through X-ray Crystallography. *Cell* **2014,** *159* (5), 995-1014.

69.   Wlodawer, A.; Minor, W.; Dauter, Z.; Jaskolski, M., Protein Crystallography for Aspiring Crystallographers or How to Avoid Pitfalls and Traps in Macromolecular Structure Determination. *FEBS J.* **2013,** *280* (22), 5705-36.

70.   Acharya, K. R.; Lloyd, M. D., The Advantages and Limitations of Protein Crystal Structures. *Trends Pharmacol. Sci.* **2005,** *26* (1), 10-4.

71.   Zheng, H.; Handing, K. B.; Zimmerman, M. D.; Shabalin, I. G.; Almo, S. C.; Minor, W., X-Ray Crystallography over the Past Decade for Novel Drug Discovery - Where Are We Heading Next? *Expert Opin. Drug Discov.* **2015,** *10* (9), 975-89.

72.   Liebeschuetz, J.; Hennemann, J.; Olsson, T.; Groom, C. R., The Good, the Bad and the Twisted: a Survey of Ligand Geometry in Protein Crystal Structures. *J. Comput. Aided Mol. Des.* **2012,** *26* (2), 169-83.

73.   Bock, A.; Bermudez, M.; Krebs, F.; Matera, C.; Chirinda, B.; Sydow, D.; Dallanoce, C.; Holzgrabe, U.; De Amici, M.; Lohse, M. J.; Wolber, G.; Mohr, K., Ligand Binding Ensembles Determine Graded Agonist Efficacies at a G Protein-Coupled Receptor. *J. Biol. Chem.* **2016,** *291* (31), 16375-89.

74.   Jorgensen, W. L., The Many Roles of Computation in Drug Discovery. *Science* **2004,** *303* (5665), 1813-8.

75.   Talele, T. T.; Khedkar, S. A.; Rigby, A. C., Successful Applications of Computer Aided Drug Discovery: Moving Drugs from Concept to the Clinic. *Curr. Top. Med. Chem.* **2010,** *10* (1), 127-41.

76.   Kuhlman, B.; Bradley, P., Advances in Protein Structure Prediction and Design. *Nat. Rev. Mol. Cell Biol.* **2019,** *20* (11), 681-697.

77.   Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Soding, J.; Thompson, J. D.; Higgins, D. G., Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* **2011,** *7*, 539.

78.   Notredame, C.; Higgins, D. G.; Heringa, J., T-Coffee: A Novel Method for Fast and Accurate Multiple Sequence Alignment. *J. Mol. Biol.* **2000,** *302* (1), 205-17.

79.   Needleman, S. B.; Wunsch, C. D., A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins. *J. Mol. Biol.* **1970,** *48* (3), 443-453.

80.   Smith, T. F.; Waterman, M. S., Identification of Common Molecular Subsequences. *J. Mol. Biol.* **1981,** *147* (1), 195-197.

81.   Lipman, D. J.; Pearson, W. R., Rapid and Sensitive Protein Similarity Searches. *Science* **1985,** *227* (4693), 1435-41.

82.   Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J., Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990,** *215* (3), 403-410.

83.   Henikoff, S.; Henikoff, J. G., Amino Acid Substitution Matrices from Protein Blocks. *Proc. Natl. Acad. Sci. USA* **1992,** *89* (22), 10915-9.

84.   Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J., The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis. *Nat. Protoc.* **2015,** *10* (6), 845-58.

85.   Roy, A.; Kucukural, A.; Zhang, Y., I-TASSER: A Unified Platform for Automated Protein Structure and Function Prediction. *Nat. Protoc.* **2010,** *5* (4), 725-38.

86.   Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T., SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Res.* **2018,** *46* (W1), W296-W303.

87.   Nayeem, A.; Sitkoff, D.; Krystek, S., Jr., A Comparative Study of Available Software for High-Accuracy Homology Modeling: From Sequence Alignments to Structural Models. *Protein Sci.* **2006,** *15* (4), 808-24.

88.   Labute, P., The Generalized Born/Volume Integral Implicit Solvent Model: Estimation of the Free Energy of Hydration Using London Dispersion Instead of Atomic Surface Area. *J. Comput. Chem.* **2008,** *29* (10), 1693-8.

89.   Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V., Stereochemistry of Polypeptide Chain Configurations. *J. Mol. Biol.* **1963,** *7*, 95-9.

90.   Du, W.; Machalz, D.; Yan, Q.; Sorensen, E. J.; Wolber, G.; Bureik, M., Importance of Asparagine-381 and Arginine-487 for Substrate Recognition in CYP4Z1. *Biochem. Pharmacol.* **2020,** *174*, 113850.

91.   Murgueitio, M. S.; Bermudez, M.; Mortier, J.; Wolber, G., In Silico Virtual Screening Approaches for Anti-Viral Drug Discovery. *Drug Discov. Today Technol.* **2012,** *9* (3), e219-e225.

92.   Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G., A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *J. Mol. Biol.* **1996,** *261* (3), 470-89.

93.   Moustakas, D. T.; Lang, P. T.; Pegg, S.; Pettersen, E.; Kuntz, I. D.; Brooijmans, N.; Rizzo, R. C., Development and Validation of a Modular, Extensible Docking Program: DOCK 5. *J. Comput. Aided Mol. Des.* **2006,** *20* (10-11), 601-19.

94.   Jain, A. N., Surflex: Fully Automatic Flexible Molecular Docking Using a Molecular Similarity-Based Search Engine. *J. Med. Chem.* **2003,** *46* (4), 499-511.

95.   Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., AutoDock4 and AutoDockTools4: Automated Docking With Selective Receptor Flexibility. *J. Comput. Chem.* **2009,** *30* (16), 2785-91.

96.   Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S., Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004,** *47* (7), 1739-49.

97.   Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R., Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997,** *267* (3), 727-48.

98.   Zhou, Z.; Felts, A. K.; Friesner, R. A.; Levy, R. M., Comparative Performance of Several Flexible Docking Programs and Scoring Functions: Enrichment Studies for a

Diverse Set of Pharmaceutically Relevant Targets. *J. Chem. Inf. Model.* **2007,** *47* (4), 1599-608.

99.   Wang, Z.; Sun, H.; Yao, X.; Li, D.; Xu, L.; Li, Y.; Tian, S.; Hou, T., Comprehensive Evaluation of Ten Docking Programs on a Diverse Set of Protein-Ligand Complexes: The Prediction Accuracy of Sampling Power and Scoring Power. *Phys. Chem. Chem. Phys.* **2016,** *18* (18), 12964-75.

100.   Jones, G.; Willett, P.; Glen, R. C., Molecular Recognition of Receptor Sites Using a Genetic Algorithm With a Description of Desolvation. *J. Mol. Biol.* **1995,** *245* (1), 43-53.

101.   Mooij, W. T.; Verdonk, M. L., General and Targeted Statistical Potentials for Protein-Ligand Interactions. *Proteins* **2005,** *61* (2), 272-87.

102.   Baxter, C. A.; Murray, C. W.; Clark, D. E.; Westhead, D. R.; Eldridge, M. D., Flexible Docking Using Tabu Search and an Empirical Estimate of Binding Affinity. *Proteins* **1998,** *33* (3), 367-82.

103.   Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S., A Critical Assessment of Docking Programs and Scoring Functions. *J. Med. Chem.* **2006,** *49* (20), 5912-31.

104.   Weitzner, B. B., M.; Kilambi, K.; Thottungal, R.; Chaudhury, S.; Wang, C.; Gray, J., Docking Protocol (RosettaDock). (online, accessed 2019-04-12) **2017,** (rosettacommons.org/docs/latest/application_documentation/docking/docking-protocol#tips).

105.   Rakers, C.; Bermudez, M.; Keller, B. G.; Mortier, J.; Wolber, G., Computational Close Up on Protein-Protein Interactions: How to Unravel the Invisible Using Molecular Dynamics Simulations? *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2015,** *5* (5), 345-359.

106.   Dominguez, C.; Boelens, R.; Bonvin, A. M., HADDOCK: A Protein-Protein Docking Approach Based on Biochemical or Biophysical Information. *J. Am. Chem. Soc.* **2003,** *125* (7), 1731-7.

107.   Gray, J. J.; Moughon, S.; Wang, C.; Schueler-Furman, O.; Kuhlman, B.; Rohl, C. A.; Baker, D., Protein-Protein Docking with Simultaneous Optimization of Rigid-Body Displacement and Side-Chain Conformations. *J. Mol. Biol.* **2003,** *331* (1), 281-99.

108.   Chaudhury, S.; Berrondo, M.; Weitzner, B. D.; Muthu, P.; Bergman, H.; Gray, J. J., Benchmarking and Analysis of Protein Docking Performance in Rosetta v3.2. *PLoS One* **2011,** *6* (8), e22477.

109.   Marze, N. A.; Jeliazkov, J. R.; Roy Burman, S. S.; Boyken, S. E.; DiMaio, F.; Gray, J. J., Modeling Oblong Proteins and Water-Mediated Interfaces with RosettaDock in CAPRI Rounds 28-35. *Proteins* **2017,** *85* (3), 479-486.

110.   Wermuth, C. G.; Ganellin, C. R.; Lindberg, P.; Mitscher, L. A., Glossary of Terms Used in Medicinal Chemistry (IUPAC Recommendations 1998). *Pure Appl. Chem.* **1998,** *70* (5), 1129-1143.

111.   Greene, J.; Kahn, S.; Savoj, H.; Sprague, P.; Teig, S., Chemical Function Queries for 3D Database Search. *J. Chem. Inf. Comput. Sci.* **2002,** *34* (6), 1297-1308.

112.   Wolber, G.; Dornhofer, A. A.; Langer, T., Efficient Overlay of Small Organic Molecules Using 3D Pharmacophores. *J. Comput. Aided Mol. Des.* **2006,** *20* (12), 773-88.

113.   Wolber, G.; Langer, T., LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters. *J. Chem. Inf. Model.* **2005,** *45* (1), 160-9.

114.   Barnum, D.; Greene, J.; Smellie, A.; Sprague, P., Identification of Common Functional Configurations among Molecules. *J. Chem. Inf. Comput. Sci.* **1996,** *36* (3), 563-71.

115.   Koes, D. R.; Camacho, C. J., Pharmer: Efficient and Exact Pharmacophore Search. *J. Chem. Inf. Model.* **2011,** *51* (6), 1307-14.

116.   Dixon, S. L.; Smondyrev, A. M.; Rao, S. N., PHASE: A Novel Approach to Pharmacophore Modeling and 3D Database Searching. *Chem. Biol. Drug Des.* **2006,** *67* (5), 370-2.

117.   Schaller, D.; Šribar, D.; Noonan, T.; Deng, L.; Nguyen, T. N.; Pach, S.; Machalz, D.; Bermudez, M.; Wolber, G., Next Generation 3D Pharmacophore Modeling. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2020,** *10* (4), e1468.

118.   Bissantz, C.; Kuhn, B.; Stahl, M., A Medicinal Chemist's Guide to Molecular Interactions. *J. Med. Chem.* **2010,** *53* (14), 5061-84.

119.   Taylor, R.; Kennard, O.; Versichel, W., Geometry of the Imino-Carbonyl (N-H...O:C) Hydrogen Bond. 1. Lone-Pair Directionality. *J. Am. Chem. Soc.* **1983,** *105* (18), 5761-5766.

120. Norberg, J.; Nilsson, L., On the Truncation of Long-Range Electrostatic Interactions in DNA. *Biophys. J.* **2000,** *79* (3), 1537-1553.

121. Honig, B.; Nicholls, A., Classical Electrostatics in Biology and Chemistry. *Science* **1995,** *268* (5214), 1144-9.

122. Sribar, D.; Grabowski, M.; Murgueitio, M. S.; Bermudez, M.; Weindl, G.; Wolber, G., Identification and Characterization of a Novel Chemotype for Human TLR8 Inhibitors. *Eur. J. Med. Chem.* **2019,** *179*, 744-752.

123. Machalz, D.; Li, H.; Du, W.; Sharma, S.; Liu, S.; Bureik, M.; Wolber, G., Discovery of a Novel Potent Cytochrome P450 CYP4Z1 Inhibitor. *Eur. J. Med. Chem.* **2021,** *215*, 113255.

124. Pach, S.; Sarter, T. M.; Yousef, R.; Schaller, D.; Bergemann, S.; Arkona, C.; Rademann, J.; Nitsche, C.; Wolber, G., Catching a Moving Target: Comparative Modeling of Flaviviral NS2B-NS3 Reveals Small Molecule Zika Protease Inhibitors. *ACS Med. Chem. Lett.* **2020,** *11* (4), 514-520.

125. Denzinger, K.; Nguyen, T. N.; Noonan, T.; Wolber, G.; Bermudez, M., Biased Ligands Differentially Shape the Conformation of the Extracellular Loop Region in 5-HT2B Receptors. *Int. J. Mol. Sci.* **2020,** *21* (24), 9728-9742.

126. Schaller, D.; Hagenow, S.; Stark, H.; Wolber, G., Ligand-Guided Homology Modeling Drives Identification of Novel Histamine H3 Receptor Ligands. *PLoS One* **2019,** *14* (6), e0218820.

127. Sydow, D. Dynophores: Novel Dynamic Pharmacophores. Implementation of Pharmacophore Generation Based on Molecular Dynamics Trajectories and Their Graphical Representation. Master Thesis, Humboldt University, Berlin, **2015**.

128. Schaller, D.; Pach, S.; Wolber, G., PyRod: Tracing Water Molecules in Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2019,** *59* (6), 2818-2829.

129. Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K., Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **2012,** *55* (14), 6582-94.

130. Braga, R. C.; Andrade, C. H., Assessing the Performance of 3D Pharmacophore Models in Virtual Screening: How Good Are They? *Curr. Top. Med. Chem.* **2013,** *13* (9), 1127-38.

131. Fischer, A.; Smiesko, M.; Sellner, M.; Lill, M. A., Decision Making in Structure-Based Drug Discovery: Visual Inspection of Docking Results. *J. Med. Chem.* **2021,** *64* (5), 2489-2500.

132. Baell, J. B.; Holloway, G. A., New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) From Screening Libraries and for Their Exclusion in Bioassays. *J. Med. Chem.* **2010,** *53* (7), 2719-40.

133. Stork, C.; Embruch, G.; Sicho, M.; de Bruyn Kops, C.; Chen, Y.; Svozil, D.; Kirchmair, J., NERDD: A Web Portal Providing Access to in Silico Tools for Drug Discovery. *Bioinformatics* **2020,** *36* (4), 1291-1292.

134. Nelson, K. M.; Dahlin, J. L.; Bisson, J.; Graham, J.; Pauli, G. F.; Walters, M. A., The Essential Medicinal Chemistry of Curcumin. *J. Med. Chem.* **2017,** *60* (5), 1620-1637.

135. Riniker, S., Fixed-Charge Atomistic Force Fields for Molecular Dynamics Simulations in the Condensed Phase: An Overview. *J. Chem. Inf. Model.* **2018,** *58* (3), 565-578.

136. Walton, E. B.; Vanvliet, K. J., Equilibration of Experimentally Determined Protein Structures for Molecular Dynamics Simulation. *Phys. Rev. E Stat. Nonlin. Soft. Matter Phys.* **2006,** *74* (6 Pt 1), 061901.

137. Knapp, B.; Ospina, L.; Deane, C. M., Avoiding False Positive Conclusions in Molecular Simulation: The Importance of Replicas. *J. Chem. Theory Comput.* **2018,** *14* (12), 6127-6138.

138. Bowers, K. J. C., E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D.; Salmon, J. K.; Shan, Y.; Shaw, D. E., Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06)* **2006**, 43.

139. Salomon-Ferrer, R.; Case, D. A.; Walker, R. C., An Overview of the Amber Biomolecular Simulation Package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013,** *3* (2), 198-210.

140. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R., GROMACS: A Message-passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995,** *91* (1-3), 43-56.

141. Hollingsworth, S. A.; Dror, R. O., Molecular Dynamics Simulation for All. *Neuron* **2018,** *99* (6), 1129-1143.

142.   Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L., Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001,** *105* (28), 6474-6487.

143.   Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004,** *25* (9), 1157-74.

144.   Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr., CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2010,** *31* (4), 671-90.

145.   Levitt, M.; Lifson, S., Refinement of Protein Conformations Using a Macromolecular Energy Minimization Procedure. *J. Mol. Biol.* **1969,** *46* (2), 269-279.

146.   Rouhani, M.; Khodabakhsh, F.; Norouzian, D.; Cohan, R. A.; Valizadeh, V., Molecular Dynamics Simulation for Rational Protein Engineering: Present and Future Prospectus. *J. Mol. Graph. Model.* **2018,** *84*, 43-53.

147.   Ghosh, A. K.; Osswald, H. L.; Prato, G., Recent Progress in the Development of HIV-1 Protease Inhibitors for the Treatment of HIV/AIDS. *J. Med. Chem.* **2016,** *59* (11), 5172-208.

148.   Asselah, T.; Boyer, N.; Saadoun, D.; Martinot-Peignoux, M.; Marcellin, P., Direct-Acting Antivirals for the Treatment of Hepatitis C Virus Infection: Optimizing Current IFN-Free Treatment and Future Perspectives. *Liver Int.* **2016,** *36 Suppl 1*, 47-57.

149.   Tong, L., Viral Proteases. *Chem. Rev.* **2002,** *102* (12), 4609-26.

150.   Lockbaum, G. J.; Reyes, A. C.; Lee, J. M.; Tilvawala, R.; Nalivaika, E. A.; Ali, A.; Kurt Yilmaz, N.; Thompson, P. R.; Schiffer, C. A., Crystal Structure of SARS-CoV-2 Main Protease in Complex with the Non-Covalent Inhibitor ML188. *Viruses* **2021,** *13* (2).

151.   Rawlings, N. D.; Barrett, A. J., Evolutionary Families of Peptidases. *Biochem. J.* **1993,** *290 (Pt 1)*, 205-18.

152.   Rawlings, N. D.; Barrett, A. J.; Thomas, P. D.; Huang, X.; Bateman, A.; Finn, R. D., The MEROPS Database of Proteolytic Enzymes, Their Substrates and Inhibitors in 2017 and a Comparison With Peptidases in the PANTHER Database. *Nucleic Acids Res.* **2018,** *46* (D1), D624-D632.

153.   Lei, J.; Hansen, G.; Nitsche, C.; Klein, C. D.; Zhang, L.; Hilgenfeld, R., Crystal Structure of Zika Virus NS2B-NS3 Protease in Complex with a Boronate Inhibitor. *Science* **2016,** *353* (6298), 503-5.

154.   Hedstrom, L., Serine Protease Mechanism and Specificity. *Chem. Rev.* **2002,** *102* (12), 4501-4524.

155.   Brynda, J.; Rezacova, P.; Fabry, M.; Horejsi, M.; Stouracova, R.; Sedlacek, J.; Soucek, M.; Hradilek, M.; Lepsik, M.; Konvalinka, J., A Phenylnorstatine Inhibitor Binding to HIV-1 Protease: Geometry, Protonation, and Subsite-Pocket Interactions Analyzed at Atomic Resolution. *J. Med. Chem.* **2004,** *47* (8), 2030-6.

156.   Li, Y.; Zhang, Z.; Phoo, W. W.; Loh, Y. R.; Li, R.; Yang, H. Y.; Jansson, A. E.; Hill, J.; Keller, T. H.; Nacro, K.; Luo, D.; Kang, C., Structural Insights into the Inhibition of Zika Virus NS2B-NS3 Protease by a Small-Molecule Inhibitor. *Structure* **2018,** *26* (4), 555-564 e3.

157.   Nitsche, C.; Zhang, L.; Weigel, L. F.; Schilz, J.; Graf, D.; Bartenschlager, R.; Hilgenfeld, R.; Klein, C. D., Peptide-Boronic Acid Inhibitors of Flaviviral Proteases: Medicinal Chemistry and Structural Biology. *J. Med. Chem.* **2017,** *60* (1), 511-516.

158.   Berger, A.; Schechter, I., Mapping the Active Site of Papain With the Aid of Peptide Substrates and Inhibitor. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **1970,** *257* (813), 249-64.

159.   Chappell, K. J.; Stoermer, M. J.; Fairlie, D. P.; Young, P. R., Insights to Substrate Binding and Processing by West Nile Virus NS3 Protease through Combined Modeling, Protease Mutagenesis, and Kinetic Studies. *J. Biol. Chem.* **2006,** *281* (50), 38448-58.

160.   Yamashita, T.; Unno, H.; Mori, Y.; Tani, H.; Moriishi, K.; Takamizawa, A.; Agoh, M.; Tsukihara, T.; Matsuura, Y., Crystal Structure of the Catalytic Domain of Japanese Encephalitis Virus NS3 Helicase/Nucleoside Triphosphatase at a Resolution of 1.8 A. *Virology* **2008,** *373* (2), 426-36.

161.   Falgout, B.; Pethel, M.; Zhang, Y. M.; Lai, C. J., Both Nonstructural Proteins NS2B and NS3 Are Required for the Proteolytic Processing of Dengue Virus Nonstructural Proteins. *J. Virol.* **1991,** *65* (5), 2467-75.

162.   Erbel, P.; Schiering, N.; D'Arcy, A.; Renatus, M.; Kroemer, M.; Lim, S. P.; Yin, Z.; Keller, T. H.; Vasudevan, S. G.; Hommel, U., Structural Basis for the Activation of Flaviviral NS3 Proteases From Dengue and West Nile Virus. *Nat. Struct. Mol. Biol.* **2006,** *13* (4), 372-3.

163. Nitsche, C., Proteases From Dengue, West Nile and Zika Viruses as Drug Targets. *Biophys. Rev.* **2019,** *11* (2), 157-165.

164. Braun, N. J.; Quek, J. P.; Huber, S.; Kouretova, J.; Rogge, D.; Lang-Henkel, H.; Cheong, E. Z. K.; Chew, B. L. A.; Heine, A.; Luo, D.; Steinmetzer, T., Structure-Based Macrocyclization of Substrate Analogue NS2B-NS3 Protease Inhibitors of Zika, West Nile and Dengue Viruses. *ChemMedChem* **2020,** *15* (15), 1439-1452.

165. Lin, K. H.; Ali, A.; Rusere, L.; Soumana, D. I.; Kurt Yilmaz, N.; Schiffer, C. A., Dengue Virus NS2B/NS3 Protease Inhibitors Exploiting the Prime Side. *J. Virol.* **2017,** *91* (10), e00045-17.

166. Bazan, J. F.; Fletterick, R. J., Detection of a Trypsin-Like Serine Protease Domain in Flaviviruses and Pestviruses. *Virology* **1989,** *171* (2), 637-639.

167. Noble, C. G.; Seh, C. C.; Chao, A. T.; Shi, P. Y., Ligand-Bound Structures of the Dengue Virus Protease Reveal the Active Conformation. *J. Virol.* **2012,** *86* (1), 438-46.

168. Bodenreider, C.; Beer, D.; Keller, T. H.; Sonntag, S.; Wen, D.; Yap, L.; Yau, Y. H.; Shochat, S. G.; Huang, D.; Zhou, T.; Caflisch, A.; Su, X. C.; Ozawa, K.; Otting, G.; Vasudevan, S. G.; Lescar, J.; Lim, S. P., A Fluorescence Quenching Assay to Discriminate between Specific and Nonspecific Inhibitors of Dengue Virus Protease. *Anal. Biochem.* **2009,** *395* (2), 195-204.

169. Cregar-Hernandez, L.; Jiao, G. S.; Johnson, A. T.; Lehrer, A. T.; Wong, T. A.; Margosiak, S. A., Small Molecule Pan-Dengue and West Nile Virus NS3 Protease Inhibitors. *Antivir. Chem. Chemother.* **2011,** *21* (5), 209-17.

170. Ong, I. L. H.; Yang, K. L., Recent Developments in Protease Activity Assays and Sensors. *Analyst* **2017,** *142* (11), 1867-1881.

171. Michaelis, L.; Menten, M. L., Die Kinetik der Invertinwirkung. *Biochem. Z.* **1913,** *49,* 333-369.

172. Michaelis, L.; Menten, M. L.; Johnson, K. A.; Goody, R. S., The Original Michaelis Constant: Translation of the 1913 Michaelis-Menten Paper. *Biochemistry* **2011,** *50* (39), 8264-9.

173. Cheng, Y.; Prusoff, W. H., Relationship Between the Inhibition Constant (Ki) and the Concentration of Inhibitor Which Causes 50 Per Cent Inhibition (I50) of an Enzymatic Reaction. *Biochem. Pharmacol.* **1973,** *22* (23), 3099-3108.

174.   White, J. M.; Delos, S. E.; Brecher, M.; Schornberg, K., Structures and Mechanisms of Viral Membrane Fusion Proteins: Multiple Variations on a Common Theme. *Crit. Rev. Biochem. Mol. Biol.* **2008,** *43* (3), 189-219.

175.   Kilby, J. M.; Hopkins, S.; Venetta, T. M.; DiMassimo, B.; Cloud, G. A.; Lee, J. Y.; Alldredge, L.; Hunter, E.; Lambert, D.; Bolognesi, D.; Matthews, T.; Johnson, M. R.; Nowak, M. A.; Shaw, G. M.; Saag, M. S., Potent Suppression of HIV-1 Replication in Humans by T-20, a Peptide Inhibitor of gp41-mediated Virus Entry. *Nat. Med.* **1998,** *4* (11), 1302-7.

176.   Volz, T.; Allweiss, L.; Ben, M. M.; Warlich, M.; Lohse, A. W.; Pollok, J. M.; Alexandrov, A.; Urban, S.; Petersen, J.; Lutgehetmann, M.; Dandri, M., The Entry Inhibitor Myrcludex-B Efficiently Blocks Intrahepatic Virus Spreading in Humanized Mice Previously Infected with Hepatitis B Virus. *J. Hepatol.* **2013,** *58* (5), 861-7.

177.   Rey, F. A.; Lok, S. M., Common Features of Enveloped Viruses and Implications for Immunogen Design for Next-Generation Vaccines. *Cell* **2018,** *172* (6), 1319-1334.

178.   Pach, S.; Nguyen, T. N.; Trimpert, J.; Kunec, D.; Osterrieder, N.; Wolber, G., ACE2-Variants Indicate Potential SARS-CoV-2-Susceptibility in Animals: A Molecular Dynamics Study. *Mol. Inform.* **2021,** *40* (9), 2100031.

179.   Harrison, S. C., Viral Membrane Fusion. *Virology* **2015,** *479-480*, 498-507.

180.   Li, F.; Li, W.; Farzan, M.; Harrison, S. C., Structure of SARS Coronavirus Spike Receptor-Binding Domain Complexed with Receptor. *Science* **2005,** *309* (5742), 1864-8.

181.   Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; Wang, X., Structure of the SARS-CoV-2 Spike Receptor-Binding Domain Bound to the ACE2 Receptor. *Nature* **2020,** *581* (7807), 215-220.

182.   Walls, A. C.; Park, Y. J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D., Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020,** *181* (2), 281-292 e6.

183.   Kremling, V. Structural Insights Into Ligand-Receptor Interaction of Equine Herpesvirus Type 1 and 4. Dissertation, Freie Universität Berlin, Berlin, **2020**.

184.   Liedberg, B.; Nylander, C.; Lunström, I., Surface Plasmon Resonance for Gas Detection and Biosensing. *Sens. Actuators* **1983,** *4*, 299-304.

185.   Keskin, O.; Tuncbag, N.; Gursoy, A., Predicting Protein-Protein Interactions from the Molecular to the Proteome Level. *Chem. Rev.* **2016,** *116* (8), 4884-909.

186. Bogan, A. A.; Thorn, K. S., Anatomy of Hot Spots in Protein Interfaces. *J. Mol. Biol.* **1998,** *280* (1), 1-9.

187. Liu, Q.; Li, J., Protein Binding Hot Spots and the Residue-Residue Pairing Preference: A Water Exclusion Perspective. *BMC Bioinform.* **2010,** *11*, 244.

188. Kurt Yilmaz, N.; Swanstrom, R.; Schiffer, C. A., Improving Viral Protease Inhibitors to Counter Drug Resistance. *Trends Microbiol.* **2016,** *24* (7), 547-557.

189. Li, Z.; Brecher, M.; Deng, Y. Q.; Zhang, J.; Sakamuru, S.; Liu, B.; Huang, R.; Koetzner, C. A.; Allen, C. A.; Jones, S. A.; Chen, H.; Zhang, N. N.; Tian, M.; Gao, F.; Lin, Q.; Banavali, N.; Zhou, J.; Boles, N.; Xia, M.; Kramer, L. D.; Qin, C. F.; Li, H., Existing Drugs as Broad-Spectrum and Potent Inhibitors for Zika Virus by Targeting NS2B-NS3 Interaction. *Cell Res.* **2017,** *27* (8), 1046-1064.

190. Wagner, J. R.; Sorensen, J.; Hensley, N.; Wong, C.; Zhu, C.; Perison, T.; Amaro, R. E., POVME 3.0: Software for Mapping Binding Pocket Flexibility. *J. Chem. Theory Comput.* **2017,** *13* (9), 4584-4592.

191. Nall, T. A.; Chappell, K. J.; Stoermer, M. J.; Fang, N. X.; Tyndall, J. D.; Young, P. R.; Fairlie, D. P., Enzymatic Characterization and Homology Model of a Catalytically Active Recombinant West Nile Virus NS3 Protease. *J. Biol. Chem.* **2004,** *279* (47), 48535-42.

192. Ekonomiuk, D.; Su, X. C.; Ozawa, K.; Bodenreider, C.; Lim, S. P.; Otting, G.; Huang, D.; Caflisch, A., Flaviviral Protease Inhibitors Identified by Fragment-Based Library Docking Into a Structure Generated by Molecular Dynamics. *J. Med. Chem.* **2009,** *52* (15), 4860-8.

193. Ekonomiuk, D.; Su, X. C.; Ozawa, K.; Bodenreider, C.; Lim, S. P.; Yin, Z.; Keller, T. H.; Beer, D.; Patel, V.; Otting, G.; Caflisch, A.; Huang, D., Discovery of a Non-peptidic Inhibitor of West Nile Virus NS3 Protease by High-throughput Docking. *PLoS Negl. Trop. Dis.* **2009,** *3* (1), e356.

194. Ganesh, V. K.; Muller, N.; Judge, K.; Luan, C. H.; Padmanabhan, R.; Murthy, K. H., Identification and Characterization of Nonsubstrate Based Inhibitors of the Essential Dengue and West Nile Virus Proteases. *Bioorg. Med. Chem.* **2005,** *13* (1), 257-64.

195. Schüler, C. Exploring Flaviviral Resistance: Dynamic Hotspot Analysis. Master Thesis, Freie Universität Berlin, Berlin, **2018**.

196. Kuhl, N.; Graf, D.; Bock, J.; Behnam, M. A. M.; Leuthold, M. M.; Klein, C. D., A New Class of Dengue and West Nile Virus Protease Inhibitors with Submicromolar Activity

in Reporter Gene DENV-2 Protease and Viral Replication Assays. *J. Med. Chem.* **2020,** *63* (15), 8179-8197.

197.   Sasaki, M.; Kim, E.; Igarashi, M.; Ito, K.; Hasebe, R.; Fukushi, H.; Sawa, H.; Kimura, T., Single Amino Acid Residue in the A2 Domain of Major Histocompatibility Complex Class I Is Involved in the Efficiency of Equine Herpesvirus-1 Entry. *J. Biol. Chem.* **2011,** *286* (45), 39370-8.

198.   Yao, S.; Liu, J.; Qi, J.; Chen, R.; Zhang, N.; Liu, Y.; Wang, J.; Wu, Y.; Gao, G. F.; Xia, C., Structural Illumination of Equine MHC Class I Molecules Highlights Unconventional Epitope Presentation Manner That Is Evolved in Equine Leukocyte Antigen Alleles. *J. Immunol.* **2016,** *196* (4), 1943-54.

199.   Zhang, N.; Yan, J.; Lu, G.; Guo, Z.; Fan, Z.; Wang, J.; Shi, Y.; Qi, J.; Gao, G. F., Binding of Herpes Simplex Virus Glycoprotein D to Nectin-1 Exploits Host Cell Adhesion. *Nat. Commun.* **2011,** *2*, 577.

200.   Scheurer, M.; Rodenkirch, P.; Siggel, M.; Bernardi, R. C.; Schulten, K.; Tajkhorshid, E.; Rudack, T., PyContact: Rapid, Customizable, and Visual Analysis of Noncovalent Interactions in MD Simulations. *Biophys. J.* **2018,** *114* (3), 577-583.

201.   Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O., MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **2011,** *32* (10), 2319-27.

202.   Gowers, R. J.; Linke, M.; Barnoud, J.; Reddy, T. J. E.; Melo, M. N.; Seyler, S. L.; Dotson, D. L.; Domanski, J.; Buchoux, S.; Kenney, I. M.; Beckstein, O., MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. In *Proceedings of the 15th Python in Science Conference*, Benthall, S.; Rostrup, S., Eds. Austin, Texas (USA), **2016**; pp 98-105.

203.   Shi, J.; Wen, Z.; Zhong, G.; Yang, H.; Wang, C.; Huang, B.; Liu, R.; He, X.; Shuai, L.; Sun, Z.; Zhao, Y.; Liu, P.; Liang, L.; Cui, P.; Wang, J.; Zhang, X.; Guan, Y.; Tan, W.; Wu, G.; Chen, H.; Bu, Z., Susceptibility of Ferrets, Cats, Dogs, and Other Domesticated Animals to SARS-Coronavirus 2. *Science* **2020,** *368* (6494), 1016-1020.

204.   Sit, T. H. C.; Brackman, C. J.; Ip, S. M.; Tam, K. W. S.; Law, P. Y. T.; To, E. M. W.; Yu, V. Y. T.; Sims, L. D.; Tsang, D. N. C.; Chu, D. K. W.; Perera, R.; Poon, L. L. M.; Peiris, M., Infection of Dogs with SARS-CoV-2. *Nature* **2020,** *586* (7831), 776-778.

205.   Munoz-Fontela, C.; Dowling, W. E.; Funnell, S. G. P.; Gsell, P. S.; Riveros-Balta, A. X.; Albrecht, R. A.; Andersen, H.; Baric, R. S.; Carroll, M. W.; Cavaleri, M.; Qin, C.; Crozier,

I.; Dallmeier, K.; de Waal, L.; de Wit, E.; Delang, L.; Dohm, E.; Duprex, W. P.; Falzarano, D.; Finch, C. L.; Frieman, M. B.; Graham, B. S.; Gralinski, L. E.; Guilfoyle, K.; Haagmans, B. L.; Hamilton, G. A.; Hartman, A. L.; Herfst, S.; Kaptein, S. J. F.; Klimstra, W. B.; Knezevic, I.; Krause, P. R.; Kuhn, J. H.; Le Grand, R.; Lewis, M. G.; Liu, W. C.; Maisonnasse, P.; McElroy, A. K.; Munster, V.; Oreshkova, N.; Rasmussen, A. L.; Rocha-Pereira, J.; Rockx, B.; Rodriguez, E.; Rogers, T. F.; Salguero, F. J.; Schotsaert, M.; Stittelaar, K. J.; Thibaut, H. J.; Tseng, C. T.; Vergara-Alert, J.; Beer, M.; Brasel, T.; Chan, J. F. W.; Garcia-Sastre, A.; Neyts, J.; Perlman, S.; Reed, D. S.; Richt, J. A.; Roy, C. J.; Segales, J.; Vasan, S. S.; Henao-Restrepo, A. M.; Barouch, D. H., Animal Models for COVID-19. *Nature* **2020,** *586* (7830), 509-515.

206.   Trimpert, J.; Vladimirova, D.; Dietert, K.; Abdelgawad, A.; Kunec, D.; Dokel, S.; Voss, A.; Gruber, A. D.; Bertzbach, L. D.; Osterrieder, N., The Roborovski Dwarf Hamster Is a Highly Susceptible Model for a Rapid and Fatal Course of SARS-CoV-2 Infection. *Cell Rep.* **2020,** *33* (10), 108488.

207.   Bertzbach, L. D.; Vladimirova, D.; Dietert, K.; Abdelgawad, A.; Gruber, A. D.; Osterrieder, N.; Trimpert, J., SARS-CoV-2 Infection of Chinese Hamsters (Cricetulus Griseus) Reproduces COVID-19 Pneumonia in a Well-Established Small Animal Model. *Transbound. Emerg. Dis.* **2021,** *68* (3), 1075-1079.

208.   Voss, S.; Nitsche, C., Inhibitors of the Zika Virus Protease NS2B-NS3. *Bioorg. Med. Chem. Lett.* **2020,** *30* (5), 126965.

209.   Voss, S.; Nitsche, C., Targeting the protease of West Nile virus. *RSC Med. Chem.* **2021,** *12* (8), 1262-1272.

210.   Yuan, S.; Chan, J. F.; den-Haan, H.; Chik, K. K.; Zhang, A. J.; Chan, C. C.; Poon, V. K.; Yip, C. C.; Mak, W. W.; Zhu, Z.; Zou, Z.; Tee, K. M.; Cai, J. P.; Chan, K. H.; de la Pena, J.; Perez-Sanchez, H.; Ceron-Carrasco, J. P.; Yuen, K. Y., Structure-Based Discovery of Clinically Approved Drugs as Zika Virus NS2B-NS3 Protease Inhibitors That Potently Inhibit Zika Virus Infection In Vitro and In Vivo. *Antiviral Res.* **2017,** *145*, 33-43.

211.   Procko, E., The Sequence of Human ACE2 is Suboptimal for Binding the S Spike Protein of SARS Coronavirus 2. *bioRxiv* **2020**, 10.1101/2020.03.16.994236.

212.   Lam, S. D.; Bordin, N.; Waman, V. P.; Scholes, H. M.; Ashford, P.; Sen, N.; van Dorp, L.; Rauer, C.; Dawson, N. L.; Pang, C. S. M.; Abbasian, M.; Sillitoe, I.; Edwards, S. J. L.; Fraternali, F.; Lees, J. G.; Santini, J. M.; Orengo, C. A., SARS-CoV-2 Spike Protein

Predicted to Form Complexes with Host Receptor Protein Orthologues From a Broad Range of Mammals. *Sci. Rep.* **2020,** *10* (1), 16471.

213. Wan, Y.; Shang, J.; Graham, R.; Baric, R. S.; Li, F., Receptor Recognition by the Novel Coronavirus from Wuhan: An Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. *J. Virol.* **2020,** *94* (7), e00127-20.

214. Armijos-Jaramillo, V.; Yeager, J.; Muslin, C.; Perez-Castillo, Y., SARS-CoV-2, an Evolutionary Perspective of Interaction with Human ACE2 Reveals Undiscovered Amino Acids Necessary for Complex Stability. *Evol. Appl.* **2020,** *13* (9), 2168-2178.

215. Damas, J.; Hughes, G. M.; Keough, K. C.; Painter, C. A.; Persky, N. S.; Corbo, M.; Hiller, M.; Koepfli, K. P.; Pfenning, A. R.; Zhao, H.; Genereux, D. P.; Swofford, R.; Pollard, K. S.; Ryder, O. A.; Nweeia, M. T.; Lindblad-Toh, K.; Teeling, E. C.; Karlsson, E. K.; Lewin, H. A., Broad Host Range of SARS-CoV-2 Predicted by Comparative and Structural Analysis of ACE2 in Vertebrates. *Proc. Natl. Acad. Sci. USA* **2020,** *117* (36), 22311-22322.

216. Rodrigues, J. P. G. L. M.; Barrera-Vilarmau, S.; J, M. C. T.; Sorokina, M.; Seckel, E.; Kastritis, P. L.; Levitt, M., Insights on Cross-Species Transmission of SARS-CoV-2 from Structural Modeling. *PLoS Comput. Biol.* **2020,** *16* (12), e1008449.

217. Pach, S.; Nguyen, T. N.; Trimpert, J.; Kunec, D.; Osterrieder, N.; Wolber, G., ACE2-Variants Indicate Potential SARS-CoV-2-Susceptibility in Animals: An Extensive Molecular Dynamics Study. *bioRxiv* **2020**, 10.1101/2020.05.14.092767.

218. Tegally, H.; Wilkinson, E.; Lessells, R. J.; Giandhari, J.; Pillay, S.; Msomi, N.; Mlisana, K.; Bhiman, J. N.; von Gottberg, A.; Walaza, S.; Fonseca, V.; Allam, M.; Ismail, A.; Glass, A. J.; Engelbrecht, S.; Van Zyl, G.; Preiser, W.; Williamson, C.; Petruccione, F.; Sigal, A.; Gazy, I.; Hardie, D.; Hsiao, N. Y.; Martin, D.; York, D.; Goedhals, D.; San, E. J.; Giovanetti, M.; Lourenco, J.; Alcantara, L. C. J.; de Oliveira, T., Sixteen Novel Lineages of SARS-CoV-2 in South Africa. *Nat. Med.* **2021,** *27* (3), 440-446.

219. UniProt-Consortium, UniProt: A Worldwide Hub of Protein Knowledge. *Nucleic Acids Res.* **2019,** *47* (D1), D506-D515.

220. Labute, P., Protonate3D: Assignment of Ionization States and Hydrogen Coordinates to Macromolecular Structures. *Proteins* **2009,** *75* (1), 187-205.

221. Abascal, J. L.; Vega, C., A General Purpose Model for the Condensed Phases of Water: TIP4P/2005. *J. Chem. Phys.* **2005,** *123* (23), 234505.

222.	Banks, J. L.; Beard, H. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; Halgren, T. A.; Mainz, D. T.; Maple, J. R.; Murphy, R.; Philipp, D. M.; Repasky, M. P.; Zhang, L. Y.; Berne, B. J.; Friesner, R. A.; Gallicchio, E.; Levy, R. M., Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J. Comput. Chem.* **2005,** *26* (16), 1752-80.

223.	Nosé, S., A Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984,** *52* (2), 255-268.

224.	Hoover, W. G., Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A Gen. Phys.* **1985,** *31* (3), 1695-1697.

225.	Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L., Explicit Reversible Integrators for Extended Systems Dynamics. *Mol. Phys.* **1996,** *87* (5), 1117-1157.

226.	Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996,** *14* (1), 33-8, 27-8.

227.	van Rossum, G., Python Tutorial, Technical Report CS-R9526. *Centrum voor Wiskunde en Informatica (CWI)* **1995**.

228.	Halgren, T. A., Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *J. Comput. Chem.* **1996,** *17* (5-6), 490-519.

229.	Halgren, T. A., Merck Molecular Force Field. II. MMFF94 Van Der Waals and Electrostatic Parameters for Intermolecular Interactions. *J. Comput. Chem.* **1996,** *17* (5-6), 520-552.

230.	Halgren, T. A., Merck Molecular Force Field. III. Molecular Geometries and Vibrational Frequencies for MMFF94. *J. Comput. Chem.* **1996,** *17* (5-6), 553-586.

231.	Halgren, T. A.; Nachbar, R. B., Merck Molecular Force Field. IV. Conformational Energies and Geometries for MMFF94. *J. Comput. Chem.* **1996,** *17* (5-6), 587-615.

232.	Halgren, T. A., Merck Molecular Force Field. V. Extension of MMFF94 Using Experimental Data, Additional Computational Data, and Empirical Rules. *J. Comput. Chem.* **1996,** *17* (5-6), 616-641.

233.	Zhang, Z.; Li, Y.; Loh, Y. R.; Phoo, W. W.; Hung, A. W.; Kang, C.; Luo, D., Crystal Structure of Unlinked NS2B-NS3 Protease from Zika Virus. *Science* **2016,** *354* (6319), 1597-1600.

234.	Shiryaev, S. A.; Ratnikov, B. I.; Chekanov, A. V.; Sikora, S.; Rozanov, D. V.; Godzik, A.; Wang, J.; Smith, J. W.; Huang, Z.; Lindberg, I.; Samuel, M. A.; Diamond, M. S.;

Strongin, A. Y., Cleavage Targets and the D-Arginine-Based Inhibitors of the West Nile Virus NS3 Processing Proteinase. *Biochem. J.* **2006,** *393* (Pt 2), 503-11.

235.   Sircar, A.; Chaudhury, S.; Kilambi, K. P.; Berrondo, M.; Gray, J. J., A Generalized Approach to Sampling Backbone Conformations with RosettaDock for CAPRI Rounds 13-19. *Proteins* **2010,** *78* (15), 3115-23.

236.   Chaudhury, S.; Gray, J. J., Conformer Selection and Induced Fit in Flexible Backbone Protein-Protein Docking Using Computational and NMR Ensembles. *J. Mol. Biol.* **2008,** *381* (4), 1068-87.

237.   Toukan, K.; Rahman, A., Molecular-Dynamics Study of Atomic Motions in Water. *Phys. Rev. B Condens. Matter.* **1985,** *31* (5), 2643-2648.

238.   Azab, W.; Kato, K.; Abdel-Gawad, A.; Tohya, Y.; Akashi, H., Equine Herpesvirus 4: Recent Advances Using BAC Technology. *Vet. Microbiol.* **2011,** *150* (1-2), 1-14.

239.   Azab, W.; Tsujimura, K.; Maeda, K.; Kobayashi, K.; Mohamed, Y. M.; Kato, K.; Matsumura, T.; Akashi, H., Glycoprotein C of Equine Herpesvirus 4 Plays a Role in Viral Binding to Cell Surface Heparan Sulfate. *Virus Res.* **2010,** *151* (1), 1-9.

240.   O'Leary, N. A.; Wright, M. W.; Brister, J. R.; Ciufo, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; Astashyn, A.; Badretdin, A.; Bao, Y.; Blinkova, O.; Brover, V.; Chetvernin, V.; Choi, J.; Cox, E.; Ermolaeva, O.; Farrell, C. M.; Goldfarb, T.; Gupta, T.; Haft, D.; Hatcher, E.; Hlavina, W.; Joardar, V. S.; Kodali, V. K.; Li, W.; Maglott, D.; Masterson, P.; McGarvey, K. M.; Murphy, M. R.; O'Neill, K.; Pujar, S.; Rangwala, S. H.; Rausch, D.; Riddick, L. D.; Schoch, C.; Shkeda, A.; Storz, S. S.; Sun, H.; Thibaud-Nissen, F.; Tolstoy, I.; Tully, R. E.; Vatsan, A. R.; Wallin, C.; Webb, D.; Wu, W.; Landrum, M. J.; Kimchi, A.; Tatusova, T.; DiCuccio, M.; Kitts, P.; Murphy, T. D.; Pruitt, K. D., Reference Sequence (RefSeq) Database at NCBI: Current Status, Taxonomic Expansion, and Functional Annotation. *Nucleic Acids Res.* **2016,** *44* (D1), D733-45.

241.   Benson, D. A.; Cavanaugh, M.; Clark, K.; Karsch-Mizrachi, I.; Lipman, D. J.; Ostell, J.; Sayers, E. W., GenBank. *Nucleic Acids Res.* **2013,** *41* (D1), D36-42.

242.   McKinney, W., Data Structures for Statistical Computing in Python. *Proc. of the 9th Python in Science Conf.* **2010,** *445*, 56-61.

243.   Hunter, J. D., Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007,** *9* (3), 90-95.

244.   Waskom, M., mwaskom/seaborn: v0.9.0. *https://zenodo.org/record/1313201* **2018**.

9.   List of Publications

List of Publications in Peer-Reviewed Journals

1. Pach, S.*; Nguyen, T. N.*; Trimpert, J.; Kunec, D.; Osterrieder, N.; Wolber, G., ACE2-Variants Indicate Potential SARS-CoV-2-Susceptibility in Animals: A Molecular Dynamics Study. *Mol. Inform.* **2021**, 40 (9), 2100031.
2. Machalz, D.; Pach, S.; Bermudez, M.; Bureik, M.; Wolber, G., Structural Insights into Understudied Cytochrome P450 Enzymes. *Drug Discov. Today* **2021**, 26 (10), 2456-2464.
3. Tauber, C.*; Wamser, R.*; Arkona, C.; Tugend, M.; Abdul Aziz, U. B.; Pach, S.; Schulz, R.; Jochmans, D.; Wolber, G.; Neyts, J.; Rademann, J., Chemical Evolution of Antivirals Against Enterovirus D68 through Protein-Templated Knoevenagel Reactions. *Angew. Chem. Int. Ed. Engl.* **2021**, 60 (24), 13294-13301.
4. Gabr, M. T.; Machalz, D.; Pach, S.; Wolber, G., A benzoxazole derivative as an inhibitor of anaerobic choline metabolism by human gut microbiota. *RSC Medicinal Chemistry* **2020**, 11 (12), 1402-1412.
5. Pach, S.; Sarter, T. M.; Yousef, R.; Schaller, D.; Bergemann, S.; Arkona, C.; Rademann, J.; Nitsche, C.; Wolber, G., Catching a Moving Target: Comparative Modeling of Flaviviral NS2B-NS3 Reveals Small Molecule Zika Protease Inhibitors. *ACS Med. Chem. Lett.* **2020**, 11 (4), 514-520.
6. Schaller, D.*; Šribar, D.*; Noonan, T.*; Deng, L.; Nguyen, T. N.; Pach, S.; Machalz, D.; Bermudez, M.; Wolber, G., Next generation 3D pharmacophore modeling. *WIREs Computational Molecular Science* **2020**, 10 (4), e1468.
7. Schaller, D.; Pach, S.; Wolber, G., PyRod: Tracing Water Molecules in Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2019**, 59 (6), 2818-2829.

* Indicates shared authorship.

List of Oral Presentations in Conferences

1. Prediction of Binding Epitopes for Herpesviral Glycoprotein D. *EUROPIN Retreat*, online **2020**.
2. Design of novel NS2B-NS3 flaviviral protease inhibitors. *EUROPIN Retreat*, Vienna, Austria **2019**.

3. Development of inhibitors of Flavivirus- proteases NS2B - NS3. *EUROPIN Retreat*, Vienna, Austria **2017**.

List of Poster Presentations in Conferences

1. Estimation of electrophilicity for warheads of covalent protease inhibitors. *11th International Conference on Chemical Structures,* Noordwijkerhout, Netherlands **2018**.
2. Development of small-molecular competitive inhibitors of Flavivirus proteases NS2B - NS3. *EUROPIN Retreat*, Vienna, Austria **2017**.

List of Book Chapters

1. Schaller, D.; Pach, S.; Bermudez, M.; Wolber, G., Exploiting Water Dynamics for Pharmacophore Screening. In Protein-Ligand Interactions and Drug Design, Ballante, F., Ed. Springer Nature: Heidelberg, **2021**; pp 227-238.

10. Independence Declaration

I hereby declare that the presented work has been independently completed without any unauthorized assistance. No additional aids than these stated in the text were used.

Berlin, Germany

November 11, 2021

Szymon Pach

11. Appendix

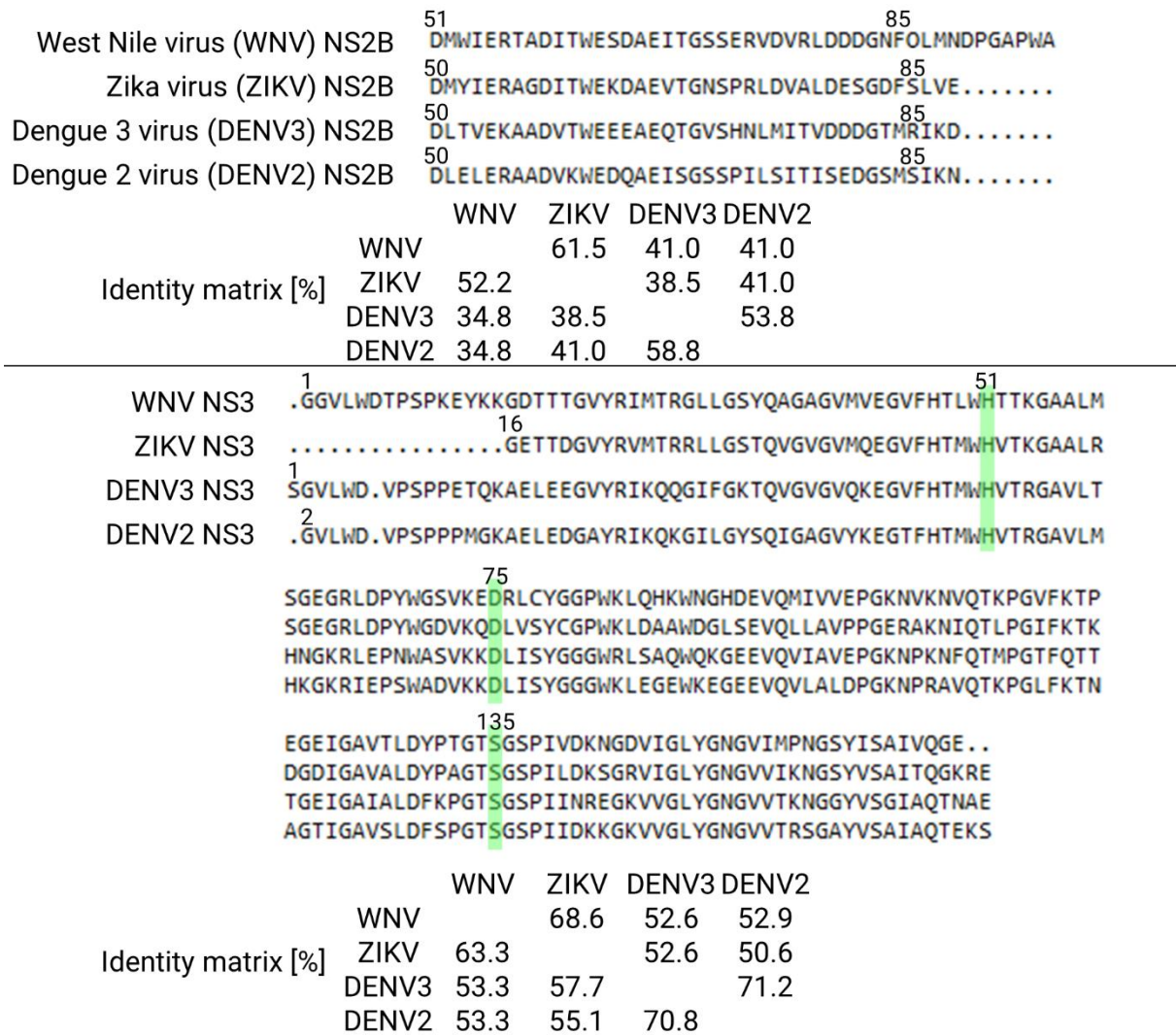## 11.1 List of Abbreviations

In alphabetical order:

ACE2- angiotensin-converting enzyme 2

AIDS- acquired immunodeficiency syndrome

AMC- 7-amino-4-methylcoumarin

Arbovirus- arthropod-borne virus

BAC- bacterial artificial chromosome

CHIKV- Chikungunya virus

COVID-19- coronavirus disease 2019

Cryo-EM- cryogenic electron microscopy

DAAD- directly acting antiviral drugs

DENV- Dengue virus

DENV2- Dengue 2 virus

dMIF- dynamic molecular interaction field

dsDNA- double-stranded deoxyribonucleic acid

dsRNA- double-stranded ribonucleic acid

E- envelope protein

EC_{50}- half-effective concentration

$EF_{1\%}$- early enrichment factor

EHV-1- equine herpesvirus 1

EHV-4- equine herpesvirus 4

EMDB- Electron Microscopy Data Bank

FP- fusion protein

FRET- Förster resonance energy transfer

gD- glycoprotein D

HCV- hepatitis C virus

HIV- human immunodeficiency virus

IC_{50}- half-inhibitory concentration

ICTV- International Committee on Taxonomy of Viruses

K_i- inhibitory equilibrium constant

k_{inact}- inactivation reaction rate

K_M- Michaelis-Menten constant

MD- molecular dynamics

MHC-I- major histocompatibility complex I

M^{Pro}- major protease

MXRA8- matrix remodeling associated 8

NMR- nuclear magnetic resonance

NS- non-structural protein

nsP- non-structural protein

PAINS- pan-assay interference compounds

PDB- Protein Data Bank

PPI- protein-protein interface

Pro- protease

RBD- residue binding domain

RMSD- root mean square deviation

ROC- receiver operating characteristic

RT- reverse transcriptase

S- spike protein

SARS-CoV-1- severe acute respiratory syndrome-related coronavirus 1

SARS-CoV-2- severe acute respiratory syndrome-related coronavirus 2

SBDD- structure-based drug design

SPR- surface plasmon resonance

+ssDNA- positive-sense single-stranded deoxyribonucleic acid

+ssRNA- positive-sense single-stranded ribonucleic acid

-ssRNA- negative-sense single-stranded ribonucleic acid

TMPRSS2- transmembrane protease serine subtype 2

WNV- West Nile virus

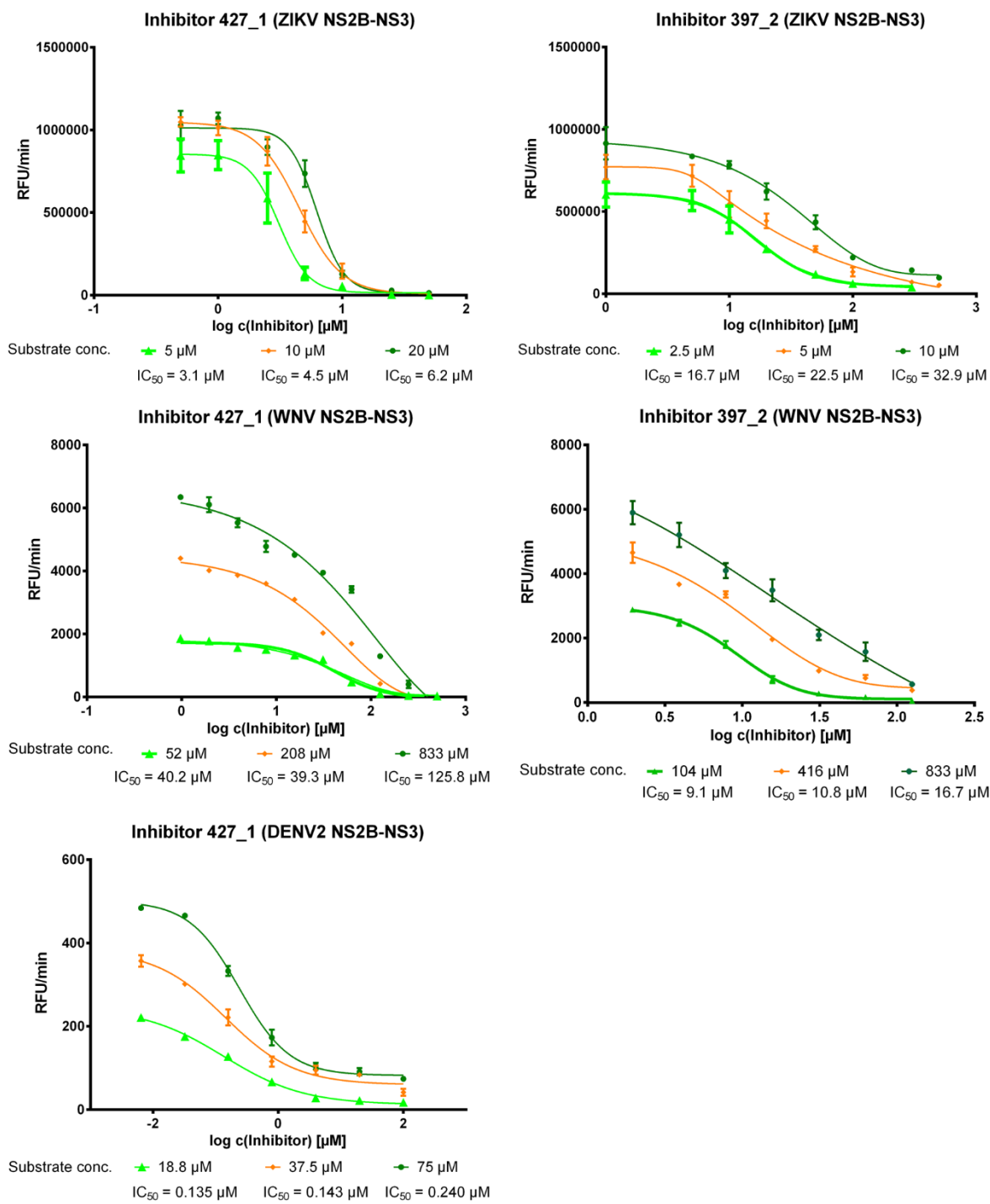YFV- yellow fever virus

ZIKV- Zika virus

## 11.2 Development of Small Molecules Targeting NS2B-NS3 Proteases
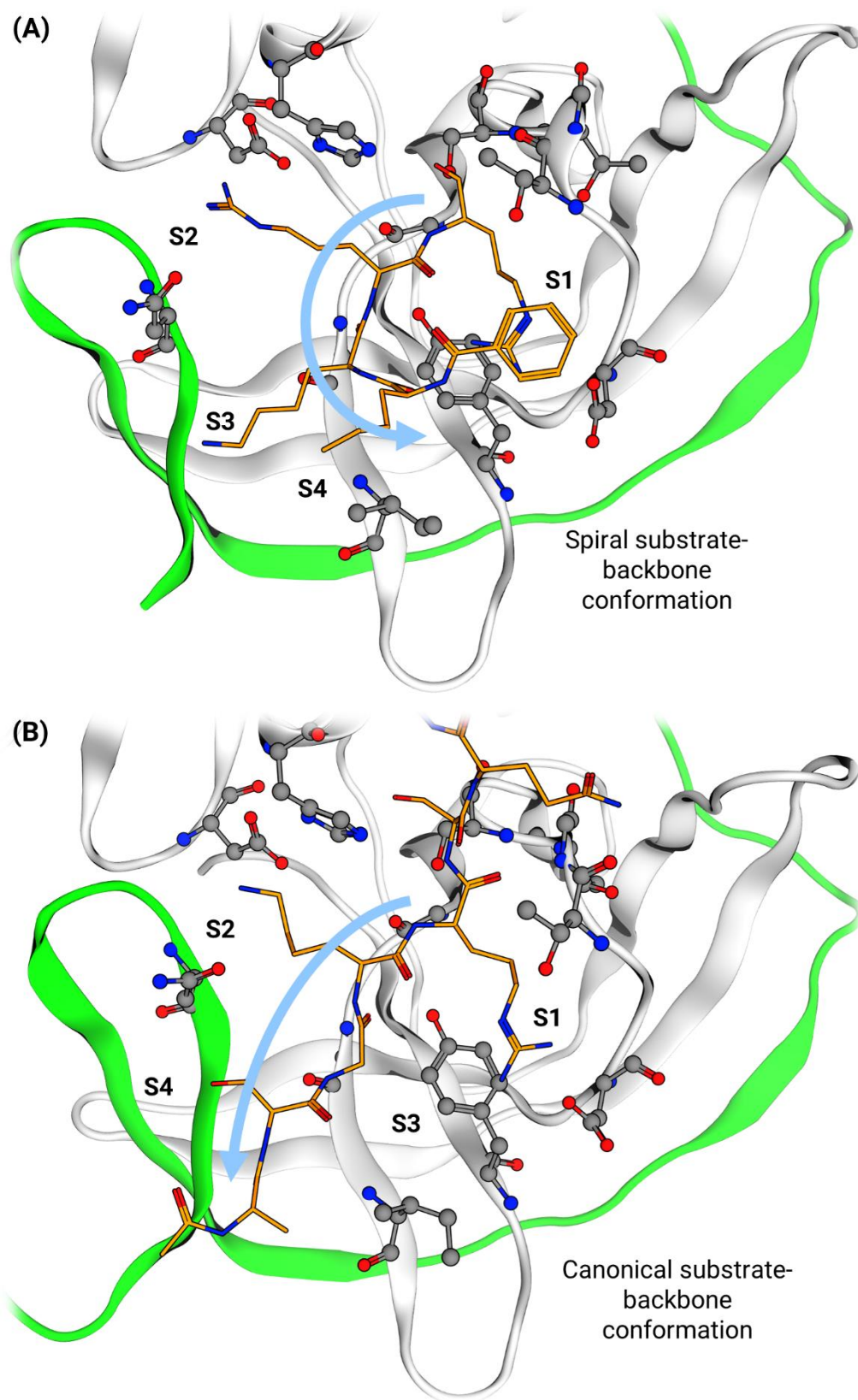
Appendix Figure 1. Hypotheses of substrate binding to the NS2B-NS3 protease reported in the literature:[159] (A) spiral-like conformation and (B) canonical conformation.

West Nile virus (WNV) NS2B

`DMWIERTADITWESDAEITGSSERVDVRLDDDGNFOLMNDPGAPWA`

Zika virus (ZIKV) NS2B

`DMYIERAGDITWEKDAEVTGNSPRLDVALDESGDFSLVE.......`

Dengue 3 virus (DENV3) NS2B

`DLTVEKAADVTWEEEAEQTGVSHNLMITVDDDGTMRIKD.......`

Dengue 2 virus (DENV2) NS2B

`DLELERAADVKWEDQAEISGSSPILSITISEDGSMSIKN.......`

Identity matrix [%]

	WNV	ZIKV	DENV3	DENV2
WNV		61.5	41.0	41.0
ZIKV	52.2		38.5	41.0
DENV3	34.8	38.5		53.8
DENV2	34.8	41.0	58.8	

WNV NS3

`.GGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFHTLWHTTKGAALM`

ZIKV NS3

`................GETTDGVYRVMTRRLLGSTQVGVGVMQEGVFHTMWHVTKGAALR`

DENV3 NS3

`SGVLWD.VPSPPETQKAELEEGVYRIKQQGIFGKTQVGVGVQKEGVFHTMWHVTRGAVLT`

DENV2 NS3

`.GVLWD.VPSPPPMGKAELEDGAYRIKQKGILGYSQIGAGVYKEGTFHTMWHVTRGAVLM`

```
SGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGHDEVQMIVVEPGKNVKNVQTKPGVFKTP
SGEGRLDPYWGDVKQDLVSYCGPWKLDAAWDGLSEVQLLAVPPGERAKNIQTLPGIFKTK
HNGKRLEPNWASVKKDLISYGGGWRLSAQWQKGEEVQVIAVEPGKNPKNFQTMPGTFQTT
HKGKRIEPSWADVKKDLISYGGGWKLEGEWKEGEEVQVLALDPGKNPRAVQTKPGLFKTN
```

```
EGEIGAVTLDYPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYISAIVQGE..
DGDIGAVALDYPAGTSGSPILDKSGRVIGLYGNGVVIKNGSYVSAITQGKRE
TGEIGAIALDFKPGTSGSPIINREGKVVGLYGNGVVTKNGGYVSGIAQTNAE
AGTIGAVSLDFSPGTSGSPIIDKKGKVVGLYGNGVVTRSGAYVSAIAQTEKS
```

Identity matrix [%]

	WNV	ZIKV	DENV3	DENV2
WNV		68.6	52.6	52.9
ZIKV	63.3		52.6	50.6
DENV3	53.3	57.7		71.2
DENV2	53.3	55.1	70.8	

Appendix Figure 2. Sequence alignment and identity matrix for the NS2B and NS3 proteins from West Nile, Zika, Dengue 3, and Dengue 2 viruses. The catalytic triad is marked in green.

Inhibitor 427_1 (ZIKV NS2B-NS3)

Substrate conc.
- 5 µM, $IC_{50} = 3.1$ µM
- 10 µM, $IC_{50} = 4.5$ µM
- 20 µM, $IC_{50} = 6.2$ µM

Inhibitor 397_2 (ZIKV NS2B-NS3)

Substrate conc.
- 2.5 µM, $IC_{50} = 16.7$ µM
- 5 µM, $IC_{50} = 22.5$ µM
- 10 µM, $IC_{50} = 32.9$ µM

Inhibitor 427_1 (WNV NS2B-NS3)

Substrate conc.
- 52 µM, $IC_{50} = 40.2$ µM
- 208 µM, $IC_{50} = 39.3$ µM
- 833 µM, $IC_{50} = 125.8$ µM

Inhibitor 397_2 (WNV NS2B-NS3)

Substrate conc.
- 104 µM, $IC_{50} = 9.1$ µM
- 416 µM, $IC_{50} = 10.8$ µM
- 833 µM, $IC_{50} = 16.7$ µM

Inhibitor 427_1 (DENV2 NS2B-NS3)

Substrate conc.
- 18.8 µM, $IC_{50} = 0.135$ µM
- 37.5 µM, $IC_{50} = 0.143$ µM
- 75 µM, $IC_{50} = 0.240$ µM

Appendix Figure 3. Inhibition curves of lead compounds **397_2** and **427_1** and the respective half-inhibitory concentration values.
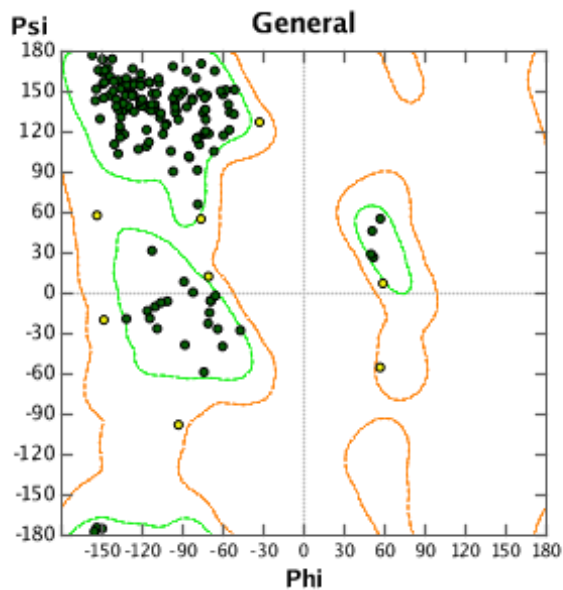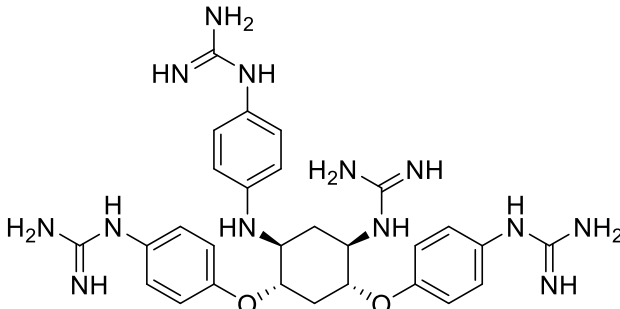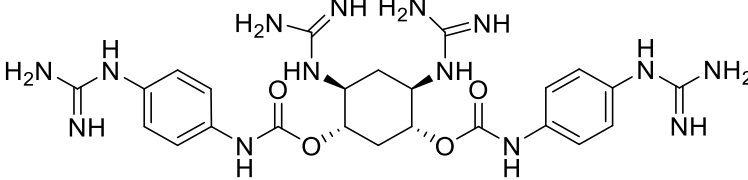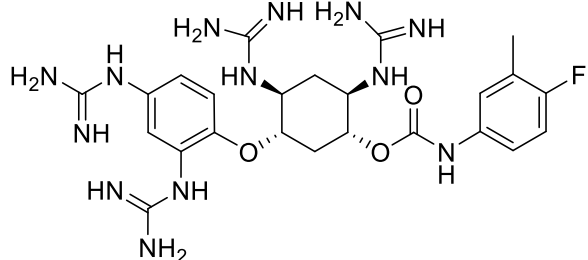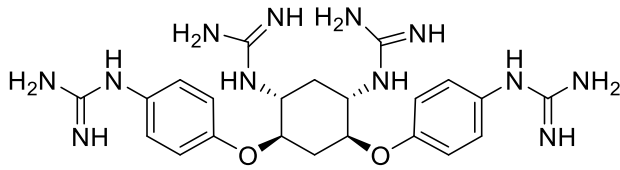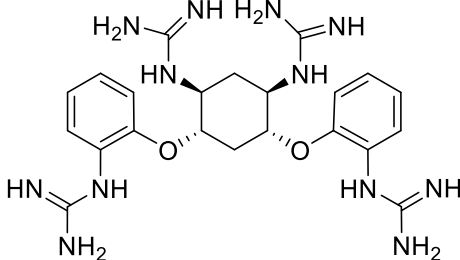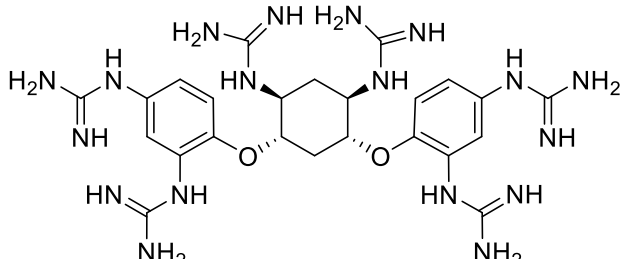
Appendix Figure 4. Ramachandran plot of the Dengue 2 virus NS2B-NS3 homology model. Symbol code: green point- residue with favorable geometry, yellow point- residue with allowed geometry.
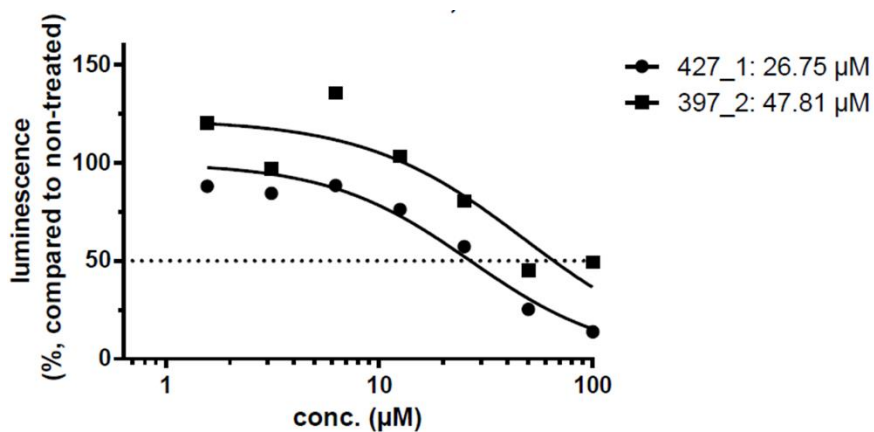
Appendix Figure 5. Cellular activity curves of lead compounds **397_2** and **427_1** and the respective half-effective concentration values. The picture was prepared by Mila Leuthold.

Appendix Table 1. Set of reportedly active NS2B-NS3 inhibitors of West Nile virus (WNV) used for the validation of the pharmacophore model.

Inhibitor structure	Inhibitory activity on WNV NS2B-NS3	Reference
	IC_{50}= 4 µM	Cregar-Hernandez et al.[169]
	IC_{50}= 9 µM	Cregar-Hernandez et al.[169]
	IC_{50}= 6 µM	Cregar-Hernandez et al.[169]
	IC_{50}= 8 µM	Cregar-Hernandez et al.[169]
	IC_{50}= 14 µM	Cregar-Hernandez et al.[169]
	IC_{50}= 1.2 µM	Cregar-Hernandez et al.[169]
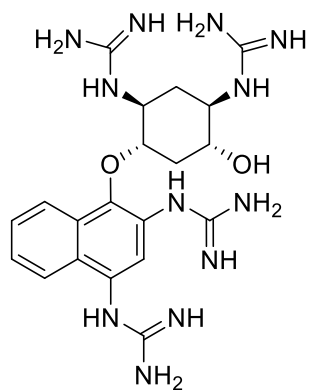
IC$_{50}$= 13 μM — Cregar-Hernandez et al.[169]
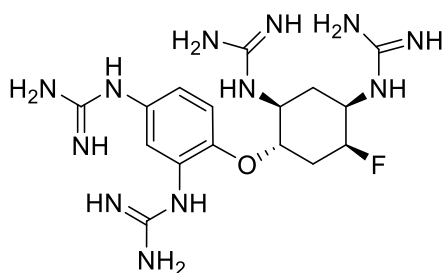
IC$_{50}$= 8 μM — Cregar-Hernandez et al.[169]
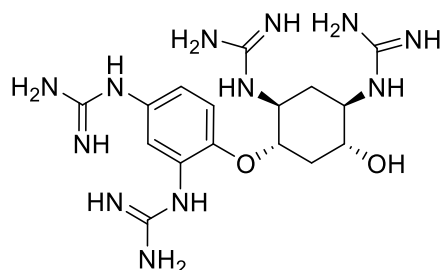
IC$_{50}$= 16 μM — Cregar-Hernandez et al.[169]
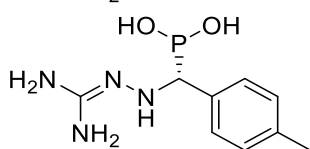
K$_i$= 13 μM — Ganesh et al.[194]
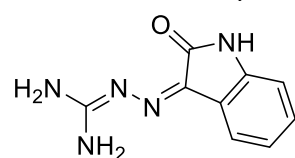
K$_i$= 16 μM — Ganesh et al.[194]
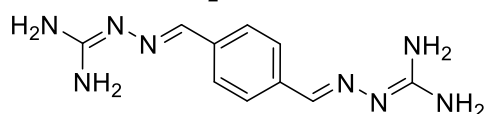
K$_i$= 35 μM — Ganesh et al.[194]
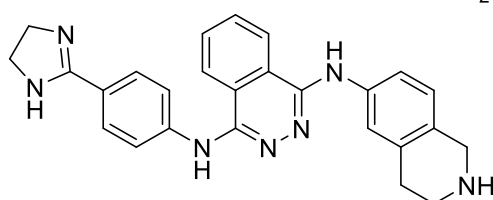
K$_D$= 7 μM — Bodenreider et al.[168]
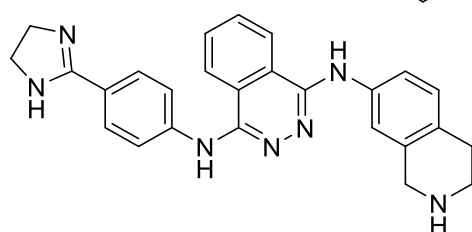
K$_D$= 4.6 μM — Bodenreider et al.[168]

144
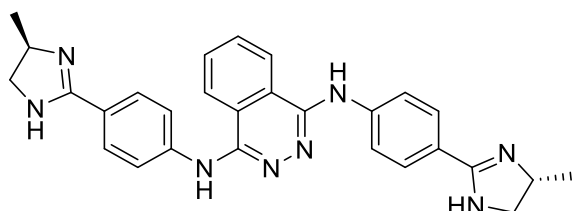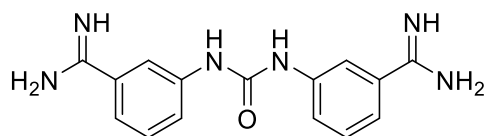
|		
	K_D= 21 μM	Bodenreider et al.[168]
	IC_{50}= 2.8 μM	Ekonomiuk et al.[192]
	K_D= 40 μM	Ekonomiuk et al.[193]
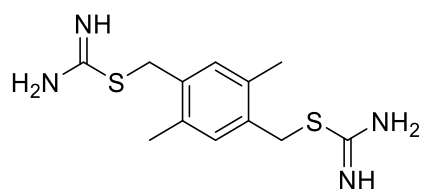
Abbreviations: WNV- West Nile virus.

11.3 Equine Herpesvirus: Targeting the MHC-I – Glycoprotein D Interface for Vaccine Development
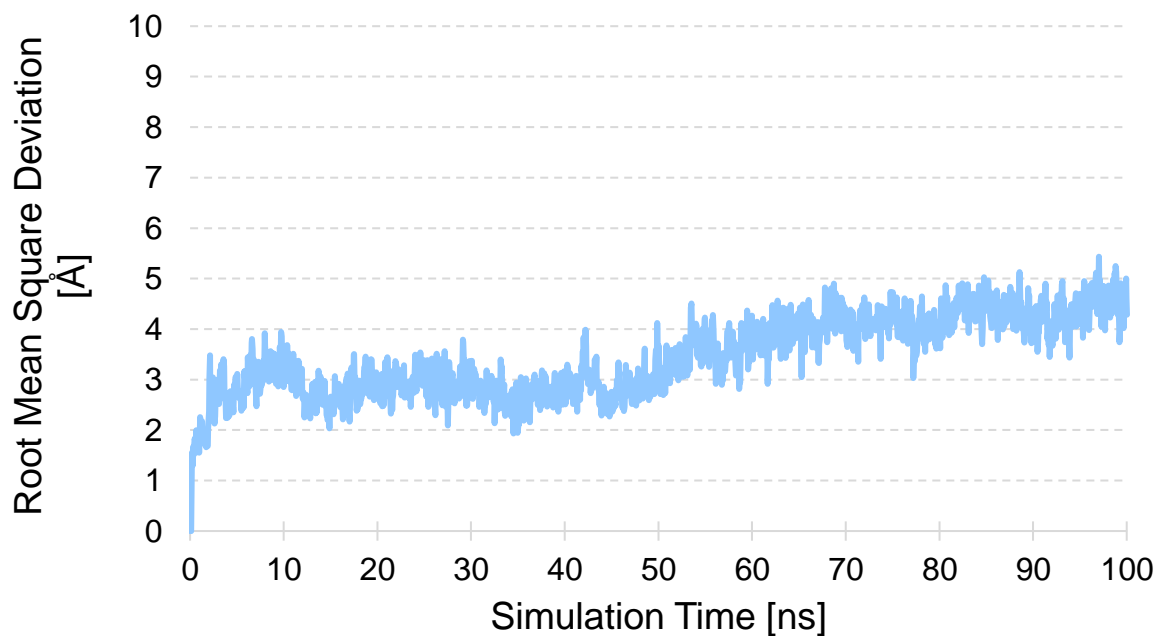
Major histocompatibility complex 1, heavy chain

Genotype

```
                25
1.18.7-6    GSHSMRYFYTAVSRPGRGEPRFIAVGYVDDTQFVRFDSDAASPRMEPRAPWVEQEGPEYW
     3.1    GSHSMRYYKTAVSRPGRGEPRFISVGYVDDTQFVRFDSDAASPRMEPRARWVEQEGPQYW

            ERETRNMKEATQNFRVGLNTLHGYYNQSEAGSHTLQRMYGCDVGPDGRLLRGYRQDAYDG
            EEETRTAKGHAQTFRGNLRIALGYYNQSEAGSHTFQWMSGCDVGPDGRLLRGYSQFAYDG

            ADYIALNEDLRSWTAADAAAQITRRKREEAGEAEQCRNYLEGTCVEWLLRYLENGNETLQ
            ADYIALNEDLRSWTAADTAAQITRRKWEAAGEAEQHRNYLEGRCVEWLLRYLENGKETLQ

            RADAPKTHVTHHPISDHEVTLRCWALGFYPEEISLSWQRDGEDVTQDTEFVETRPAGDRT
            RVDPPKTHVTHHPSSDHEVTLRCWALGFYPAEITLTWQRDGEDLTQDTEFVATRPAGDGT

            FQKWAAVVVPSGEEQRYTCHVQHEGLAEPVTLRW
            FQKWAAVVVPSGEEQRYTCHVQHEGLPEPVTRRW
```

		1.18.7-6	3.1
Identity matrix [%]	1.18.7-6		85.4
	3.1	85.4	

Major histocompatibility complex 1, β$_2$ microglobulin

Species

```
        1
Mouse   IQKTPQIQVYSRHPPENGKPNILNCYVTQFHPPHIEIQMLKNGKKIPKVEMSDMSFSKDW
Horse   VPRVPKVQVYSRHPAENGKPNFLNCYVSGFHPPEIEIDLLKNGEKM.KVDRSDLSFSKDW

        SFYILAHTEFTPTETDTYACRVKHDSMAEPKTVYWDRDM
        SFYLLVHTDFTPNGVDEYSCRVQHSTLKDPLIVKWDRDL
```

|		Mouse	Horse
Identity matrix [%]	Mouse		63.3
	Horse	62.6	

Appendix Figure 6. Sequence alignment and identity matrix for the major histocompatibility complex-I genotype 1.18.7-6 (template) and 3.1 (target). Residue 173 is highlighted in green.
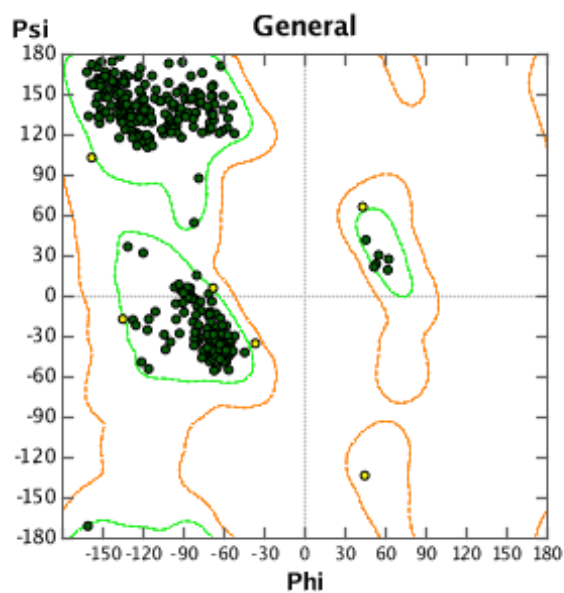
Appendix Figure 7. Ramachandran plot of the major histocompatibility complex-I genotype 3.1 homology model. Symbol code: green point- residue with favorable geometry, yellow point- residue with allowed geometry.
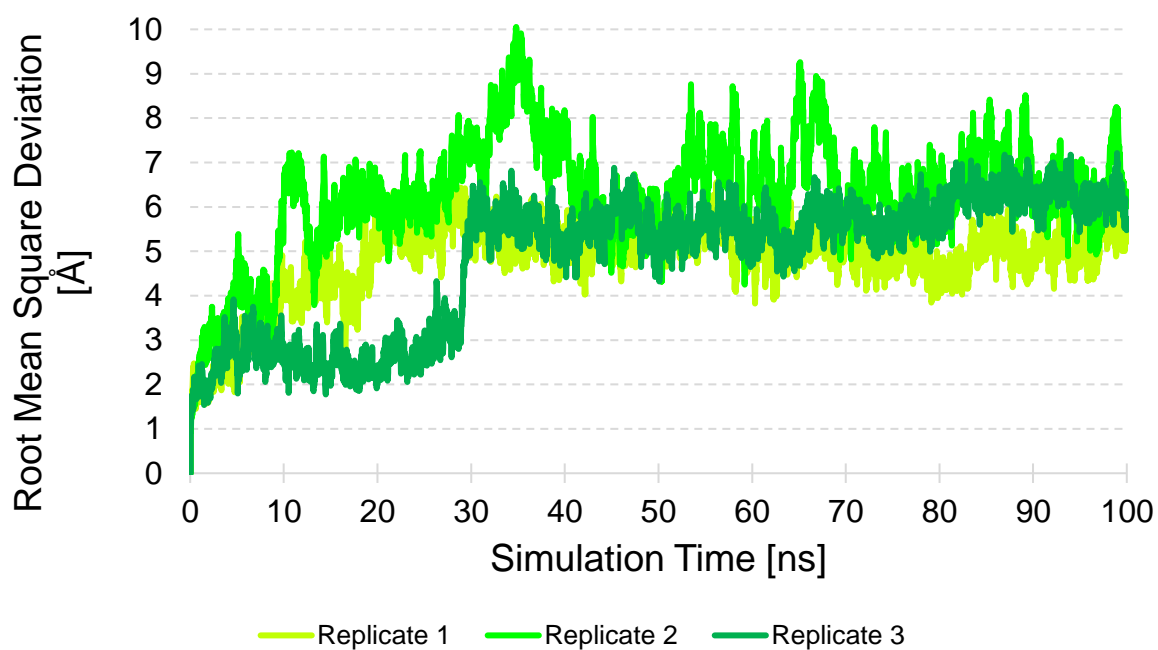

Appendix Figure 8. Root mean square deviation plot obtained from a molecular dynamics simulation of the major histocompatibility complex-I genotype 3.1 homology model.
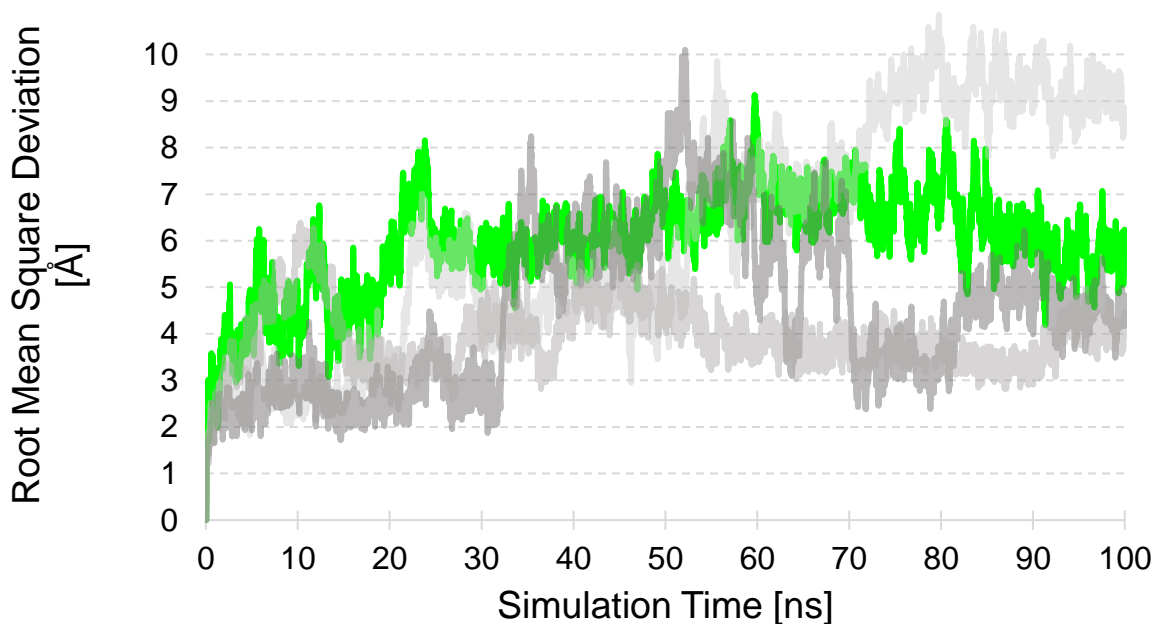
Appendix Figure 9. Root mean square deviation plot obtained from a molecular dynamics simulation of the major histocompatibility complex-I genotype 3.1 homology model in the peptide-free state in complex with the glycoprotein D from the equine herpesvirus 1. Color code: green- selected docking pose, grey- other docking poses.
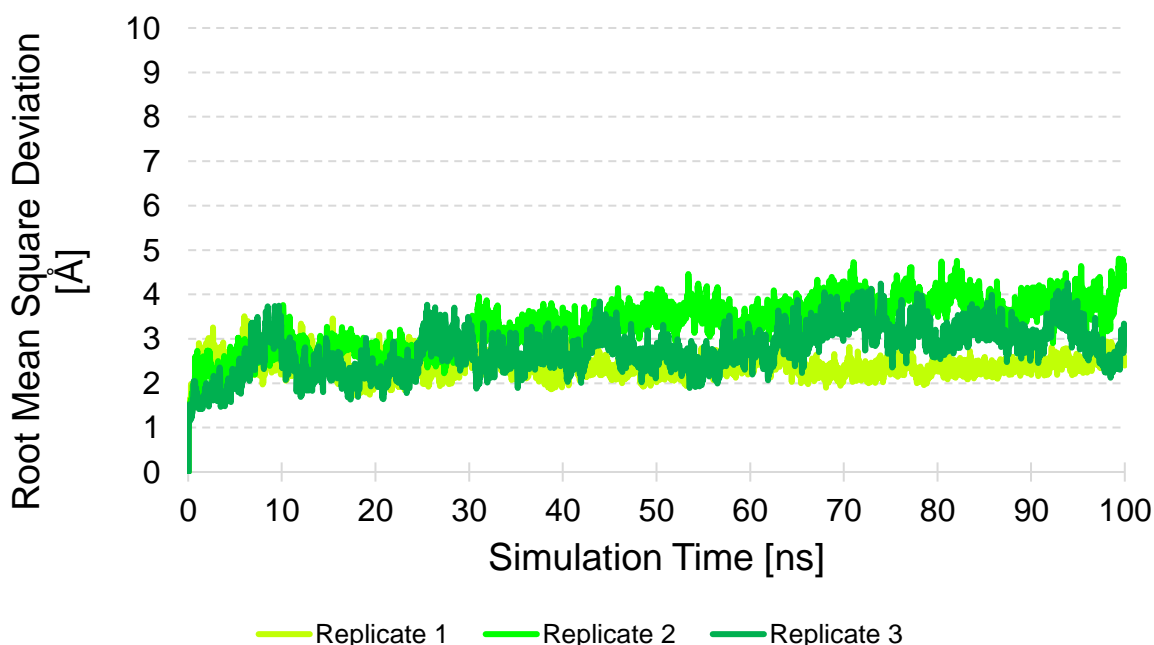
Appendix Figure 10. Root mean square deviation plot obtained from a molecular dynamics simulation of the major histocompatibility complex-I genotype 3.1 homology model in the peptide-bound state in complex with the glycoprotein D from the equine herpesvirus 1. Different shades of green indicate simulation replicates 1-3.

148

Appendix Figure 11. Root mean square deviation plot obtained from a molecular dynamics simulation of the major histocompatibility complex-I genotype 3.1 homology model in the peptide-bound state in complex with the glycoprotein D from the equine herpesvirus 4. Different shades of green indicate simulation replicates 1-3.

Appendix Table 2. Genotypes of the equine major histocompatibility complex-I reported in the literature and their glycoprotein D binding.

UniProt Code	Genotype	Residue 173	EHV-4 gD binding reported (Wazab et al.[34])
M4PTG5	-	E	-
Q38RB9	9.1	A	Yes
Q95480	2.16 (1-29)	A	Yes
Q0R0D3	9.16	A	Yes
Q38RB8	3.2	A	No
M4PPR7	-	E	-
Q38RB4	3.6	V	No
Q0R0B5	2.7	A	-
Q38RB3	3.7	A	No
Q860N6	1.18.7−6	E	-
D9MNN1	-	E	-
Q95479	2.X (8-9)	A	Yes
Q30483	3.1	A	Yes
Q30485	-	E	-
Q30486	3.6	V	No
Q30482	-	E	-
Q30484	-	E	-
Q30488	3.5	E	No
C9K796	5.X (A68)	A	Yes
Q38RB7	3.3	V	No
Q860N5	5.2	A	Yes
Q6UAL6	1.18.141	E	-
Q38RB6	3.4	T	No
Q860N9	-	E	-
Q860N8	10.1	A	Yes
Q30481	-	E	-
Q860N7	-	D	-

Abbreviations: gD- glycoprotein D, EHV-4- equine herpesvirus 4.

11.4 Predicting the Susceptibility of Animal Species to COVID-19

Human	19 STIEEQAKTFLDKFNHEAEDLFYQSSLASWNYNTNITEENVQNMNNAGDKWSAFLKEQST
Cat	19 STTEELAKTFLEKFNHEAEELSYQSSLASWNYNTNITDENVQKMNEAGAKWSAFYEEQSK
Dog	18 QSTEDLVKTFLEKFNYEAEELSYQSSLASWNYNINITDENVQKMNNAGAKWSAFYEEQSK
Ferret	19 STTEDLAKTFLEKFNYEAEELSYQNSLASWNYNTNITDENIQKMNIAGAKWSAFYEEESQ
Hamster	19 SIIEEQAKTFLDKFNQEAEDLSYQSALASWNYNTNITEENAQKMNEAAAKWSAFYEEQSK
Mouse	19 SLTEENAKTFLNNFNQEAEDLSYQSSLASWNYNTNITEENAQKMSEAAAKWSAFYEEQSK
Rat	19 SLIEEKAESFLNKFNQEAEDLSYQSSLASWNYNTNITEENAQKMNEAAAKWSAFYEEQSK
Chinese hamster	19 SIIEEQAKTFLDKFNQEAEDLSYQSALASWNYNTNITEENAQKMNEAAAKWSAFYEEQSK
C. dwarf hamster	19 SIIEEQAKTFLDKFNQEAEDLSYQSSLASWNYNTNITEENAQKMNEAAAKWSAFYEEQSK
Red squirrel	19 STIEESAKTFLDKFNQEAEDLSYQSSLASWDYNTNISEKNAQKMNEAGAKWSAFYEEQSK

```
LAQMYPLQEIQNLTVKLQLQALQQNGSSVLSEDKSKRLNTILNTMSTIYSTGKVCNPDNP
LAKTYPLAEIHNTTVKRQLQALQQSGSSVLSADKSQRLNTILNAMSTIYSTGKACNPNNP
LAKTYPLEEIQDSTVKRQLRALQHSGSSVLSADKNQRLNTILNSMSTVYSTGKACNPSNP
HAKTYPLEEIQDPIIKRQLRALQQSGSSVLSADKRERLNTILNAMSTIYSTGKACNPNNP
LAKNYSLQEVQNLTIKRQLQALQQSGSSALSADKNKQLNTILNTMSTIYSTGKVCNPKNP
TAQSFSLQEIQTPIIKRQLQALQQSGSSALSADKNKQLNTILNTMSTIYSTGKVCNPKNP
IAQNFSLQEIQNATIKRQLKALQQSGSSALSPDKNKQLNTILNTMSTIYSTGKVCNSMNP
LAKNYSLQEVQNLIIKRQLQALQQSGSSALSADKNKQLNTILNTMSTIYSTGKVCNPKNP
LAKNYPLQDVQNLTIKRQLQALQQSGSSALSADKNKQLNTILNTMSTIYSTGKVCNPKNP
LAKTYPLQQIQNLTVKRQLQALQQSGSSVLSTDKQKQLNTILNMMSTIYSTGKVCNPNKP
```

```
QECLLLEPGLNEIMANSLDYNERLWAWESWRSEVGKQLRPLYEEYVVLKNEMARAN..HY
QECLLLEPGLDDIMENSKDYNERLWAWEGWRAEVGKQLRPLYEEYVALKNEMAKSKQVNY
QECLLLEPGLDDIMENSKDYNERLWAWEGWRSEVGKQLRPLYEEYVALKNEMARAN..NY
QECLLLEPGLDDIMENSKDYNERLWAWEGWRSEVGKQLRPLYEEYVALKNEMARAN..NY
QECLLLEPGLDDIMATSTDYNERLWAWEGWRAEVGKQLRPLYEEYVVLKNEMARAN..NY
QECLLLEPGLDEIMATSTDYNSRLWAWEGWRAEVGKQLRPLYEEYVVLKNEMARAN..NY
QECFLLEPGLDEIMATSTDYNRRLWAWEGWRAEVGKQLRPLYEEYVVLKNEMARAN..NY
QECLLLEPGLDDIMATSTDYNERLWAWEGWRAEVGKQLRPLYEEYVVLKNEMARAN..NY
QECLLLEPGLDDIMATSTDYNERLWAWEGWRAEVGKQLRPLYEEYVVLKNEMARAN..NY
QECLLLEPGLDDIMANSTDYNERLWVWEGWRSEVGKQLRPLYEEYVVLKNEMARAN..DY
```

```
EDYGDYWRGDYEVNGVDGYDYSRGQLIEDVEHTFEEIKPLYEHLHAYVRAKLMNAYPSYI
EDYGDYWRGDYEEEWTDGYNYSRSQLIKDVEHTFTQIKPLYQHLHAYVRAKLMDTYPSRI
EDYGDYWRGDYEEEWENGYNYSRNQLIDDVELTFTQIMPLYQHLHAYVRTKLMDTYPSYI
EDYGDYWRGDYEEEWADGYSYSRNQLIEDVEHTFTQIKPLYEHLHAYVRAKLMDAYPSRI
EDYGDYWRGDYEAEGADGYNYNGNQLIEDVERTFKEIKPLYEQLHAYVRTKLMNTYPSYI
NDYGDYWRGDYEAEGADGYNYNRNQLIEDVERTFAEIKPLYEHLHAYVRRKLMDTYPSYI
EDYGDYWRGDYEAEGVEGYNYNRNQLIEDVENTFKEIKPLYEQLHAYVRTKLMEVYPSYI
KDYGDYWRGDYEAEGADGYNYNGNQLIEDVERTFKEIKPLYEQLHAYVRTKLMDTYPSYI
KDYGDYWRGDYEAEGENGYNYNGNQLIEDVERTFKEIKPLYEQLHAYVRTKLVNTYPSYI
EDYGDYWRGDYEAEGADGYGYNRNQLIEDVERTFAEIKPLYEHLHAYVRAKLMDTYPSYI
```

```
SPIGCLPAHLLGDMWGRFWTNLYSLTVPFGQKPNIDVTDAMVDQAWDAQRIFKEAEKFFV
SPTGCLPAHLLGDMWGRFWTNLYPLTVPFGQKPNIDVTDAMVNQSWDARRIFKEAEKFFV
SPTGCLPAHLLGDMWGRFWTNLYPLTVPFGQKPNIDVTNAMVNQSWDARKIFKEAEKFFV
SPTGCLPAHLLGDMWGRFWTNLYPLMVPFRQKPNIDVTDAMVNQSWDARRIFEEAETFFV
SPTGCLPAHLLGDMWGRFWTNLYPLTVPFGQKPNIDVTDAMVNQGWNAERIFKEAEKFFV
SPTGCLPAHLLGDMWGRFWTNLYPLTVPFAQKPNIDVTDAMMNQGWDAERIFQEAEKFFV
SPTGCLPAHLLGDMWGRFWTNLYPLTTPFLQKPNIDVTDAMVNQSWDAERIFKEAEKFFV
SPTGCLPAHLLGDMWGRFWTNLYPLTVPFGQKPNIDVTDAMVNQGWDAERIFKEAEKFFV
SPTGCLPAHLLGDMWGRFWTNLYPLTVPFGQKPNIDVTDAMVKQGWGAERIFKEAEKFFV
SPTGCLPAHLLGDMWGRFWTNLYSLTVPFQEKPNIDVTDAMMNQNWDAMRIFKEAEKFFV

SVGLPNMTQGFWENSMLTDPGNVQKAVCHPTAWDLGKGDFRILMCTKVTMDDFLTAHHEM
SVGLPNMTQGFWENSMLTEPGDSRKVVCHPTAWDLGKGDFRIKMCTKVTMDDFLTAHHEM
SVGLPNMTQEFWGNSMLTEPSDSRKVVCHPTAWDLGKGDFRIKMCTKVTMDDFLTAHHEM
SVGLPNMTEGFWQNSMLTEPGDNRKVVCHPTAWDLGKRDFRIKMCTKVTMDDFLTAHHEM
SVGLPYMTQGFWENSMLTDPGDDRKVVCHPTAWDLGKGDFRIKMCTKVTMDNFLTAHHEM
SVGLPHMTQGFWANSMLTEPADGRKVVCHPTAWDLGHGDFRIKMCTKVTMDNFLTAHHEM
SVGLPQMTPGFWTNSMLTEPGDDRKVVCHPTAWDLGHGDFRIKMCTKVTMDNFLTAHHEM
SVGLPHMTQGFWGNSMLTDPGDDRKVVCHPTAWDLGKGDFRIKMCTKVTMDNFLTAHHEM
SVGLPHMTKGFWQNSMLTDPGDDRKVVCHPTAWDLGKEDFRIKMCTKVTMDNFLTAHHEM
SVGLPNMTQGFWENSMLTEPADGRKVVCHPTAWDLGKGDFRIKMCTKVTMDNFLTAHHEM

GHIQYDMAYAAQPFLLRNGANEGFHEAVGEIMSLSAATPKHLKSIGLLSPDFQEDNETEI
GHIQYDMAYAVQPFLLRNGANEGFHEAVGEIMSLSAATPNHLKTIGLLSPGFSEDSETEI
GHIQYDMAYAAQPFLLRNGANEGFHEAVGEIMSLSAATPNHLKNIGLLPPSFFEDSETEI
GHIQYDMAYAEQPFLLRNGANEGFHEAVGEIMSLSAATPNHLKNIGLLPPDFSEDSETDI
GHIQYDMAYATQPFLLRNGANEGFHEAVGEIMSLSAATPEHLKSIGLLPSDFQEDNETEI
GHIQYDMAYARQPFLLRNGANEGFHEAVGEIMSLSAATPKHLKSIGLLPSDFQEDSETEI
GHIQYDMAYAKQPFLLRNGANEGFHEAVGEIMSLSAATPKHLKSIGLLPSNFQEDNETEI
GHIQYDMAYATQPFLLRNGANEGFHEAVGEIMSLSAATPKHLKSIGLLPSNFHEDNETEI
GHIQYDMAYATQPFLLRNGANEGFHEAVGEIMSLSAATPEHLKSIGLLPSNFQEDSETEI
GHIQYDMAYAIQPYLLRNGANEGFHEAVGEIMSLSAATPKHLKSIGLLPPDFREDNETEI

NFLLKQALTIVGTLPFTYMLEKWRWMVFKGEIPKDQWMKKWWEMKREIVGVVEPVPHDET
NFLLKQALTIVGTLPFTYMLEKWRWMVFKGEIPKEQWMQKWWEMKREIVGVVEPVPHDET
NFLLKQALTIVGTLPFTYMLEKWRWMVFKGEIPKDQWMKTWWEMKRNIVGVVEPVPHDET
NFLLKQALTIVGTLPFTYMLEKWRWMVFKGEIPKEQWMQKWWEMKRDIVGVVEPLPHDET
NFLLKQALTIVGTLPFTYMLEKWRWMVFKGDIPKEQWMEKWWEMKREIVGVVEPLPHDET
NFLLKQALTIVGTLPFTYMLEKWRWMVFRGEIPKEQWMKKWWEMKREIVGVVEPLPHDET
NFLLKQALTIVGTLPFTYMLEKWRWMVFQDKIPREQWTKKWWEMKREIVGVVEPLPHDET
NFLLKQALTIVGTLPFTYMLEKWRWMVFKGDIPKEKWMEKWWEMKREIVGVVEPLPHDET
NFLLKQALTIIGTLPFTYMLEKWRWMVFKGDIPKEQWMEKWWEMKREIVGVVEPLPHDET
NFLLKQALTIVGTLPFTYMLEKWRWMVFKGEIPKEQWMKKWWEMKREIVGVVEPVPHDET
```

```
YCDPASLFHVSNDYSFIRYYTRTLYQFQFQEALCQAAKHEGPLHKCDISNSTEAGQKLFN
YCDPASLFHVANDYSFIRYYTRTIYQFQFQEALCRIAKHEGPLHKCDISNSSEAGKKLLQ
YCDPASLFHVANDYSFIRYYTRTIYQFQFQEALCQIAKHEGPLHKCDISNSSEAGQKLLE
YCDPAALFHVANDYSFIRYYTRTIYQFQFQEALCQIAKHEGPLYKCDISNSSEAGQKLHE
YCDPAALFHVSNDYSFIRYYTRTIYQFQFQEALCQAAKHDGPLHKCDISNSTEAGQKLLN
YCDPASLFHVSNDYSFIRYYTRTIYQFQFQEALCQAAKYNGSLHKCDISNSTEAGQKLLK
YCDPASLFHVSNDYSFIRYYTRTIYQFQFQEALCQAAKHDGPLHKCDISNSTEAGQKLLN
YCDPAALFHVSNDYSFIRYYTRTIYQFQFQEALCQAAKHDGPLHKCDISNSTEAGQKLLN
YCDPAALFHVSNDFSFIRYYTRTIYQFQFQEALCKAAKHDGPLHKCDISNSTEAGQKLVN
YCDPATLFHVSNDYSFIRYYTRTIYQFQFQEALCQAAKHEGPLYKCDISNSLEAGQKLLN

MLRLGKSEPWTLALENVVGAKNMNVRPLLNYFEPLFTWLKDQNKNSFVGWSTDWSPYAD
MLTLGKSKPWTLALEHVVGEKKMNVTPLLKYFEPLFTWLKEQNRNSFVGWNTDWRPYAD
MLKLGKSKPWTYALEIVVGAKNMDVRPLLNYFEPLFTWLKEQNRNSFVGWNTDWSPYAD
MLSLGRSKPWTFALERVVGAKTMDVRPLLNYFEPLFTWLKEQNRNSFVGWNTDWSPYAD
MLRLGKSEPWTLALENVVGARNMDVRPLLNYFEPLSVWLKEQNKNSFVGWNTDWSPYAD
MLSLGNSEPWTKALENVVGARNMDVKPLLNYFQPLFDWLKEQNRNSFVGWNTEWSPYAD
MLSLGNSGPWTLALENVVGSRNMDVKPLLNYFQPLFVWLKEQNRNSTVGWSTDWSPYAD
MLRLGKSEPWTLALENVVGARNMDVRPLLNYFEPLSVWLKEQNKNSFVGWNTDWSPYAD
MLRLGKSGPWTLALEKVVGARNMDVRPLLNYFEPLSVWLKEQNKNSFVGWNTDWSPYAD
MLRLGKSEPWTLALENVVGAKNMDVRPLLNYFEPLFVWLKDQNRNSFVGWNTDWSPYAD
```

[%]	Human	Cat	Dog	Ferret	Ham-ster	Mouse	Rat	Chinese ham-ster	C. dwarf ham-ster	Red squirrel
Human		85.3	84.3	83.6	87.6	84.3	84.3	86.8	85.6	88.4
Cat	85.5		90.8	90.3	85.8	84.1	82.9	85.4	84.4	86.8
Dog	84.1	90.5		90.3	85.1	83.9	82.6	85.1	84.3	85.9
Ferret	83.4	90.0	90.3		84.9	84.6	82.4	84.9	83.9	86.1
Hamster	87.5	85.5	85.1	84.9		91.1	91.1	98.3	96.5	90.8
Mouse	84.1	83.8	83.9	84.6	91.1		90.8	91.5	89.1	88.8
Rat	84.1	82.6	82.6	82.4	91.1	90.8		91.0	89.6	86.8
Chinese hamster	86.6	85.1	85.1	84.9	98.3	91.5	91.0		96.1	90.5
C. dwarf hamster	85.5	84.1	84.3	83.9	96.5	89.1	89.6	96.1		88.4
Red squirrel	88.3	86.5	85.9	86.1	90.8	88.8	86.8	90.5	88.4	

Appendix Figure 12. Sequence alignment and identity matrix for the angiotensin-converting enzyme 2 from human and animal orthologs. (The subfigures are presented on the previous page)
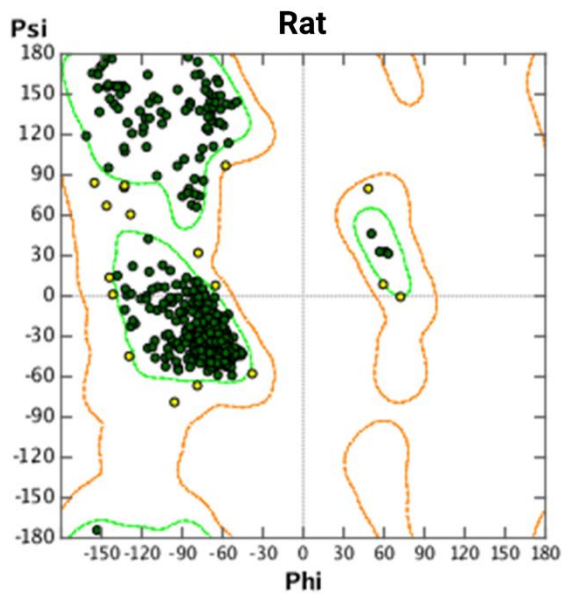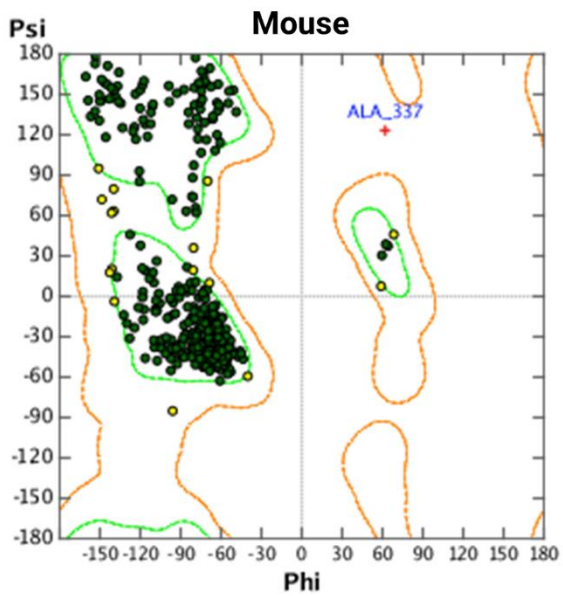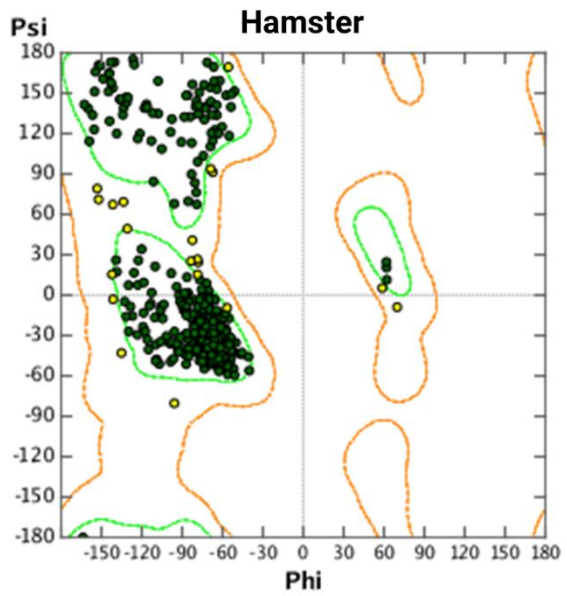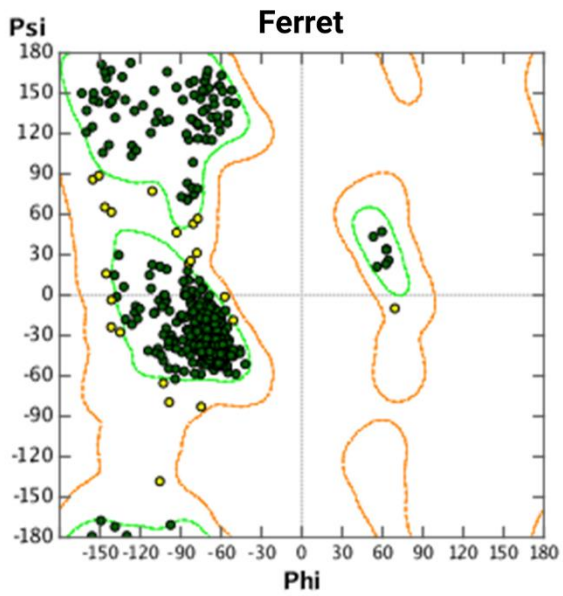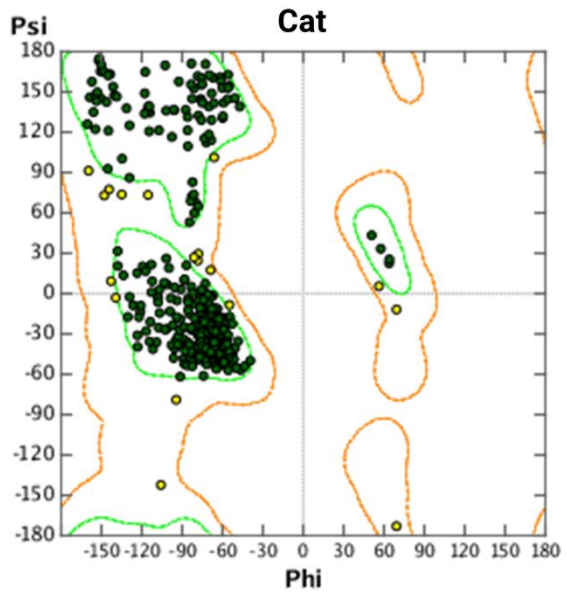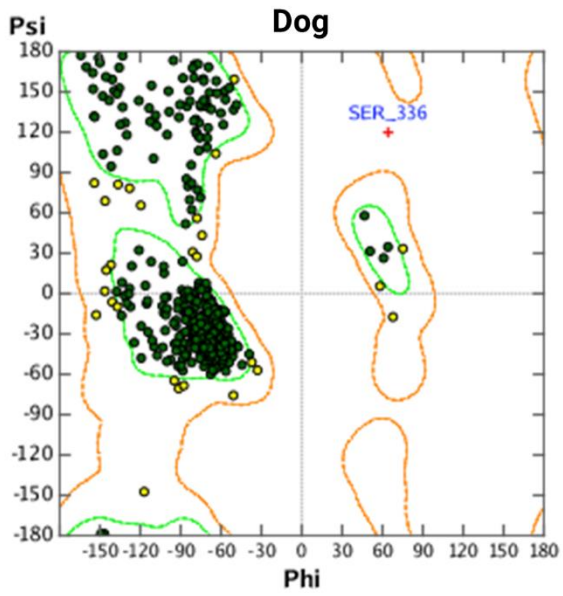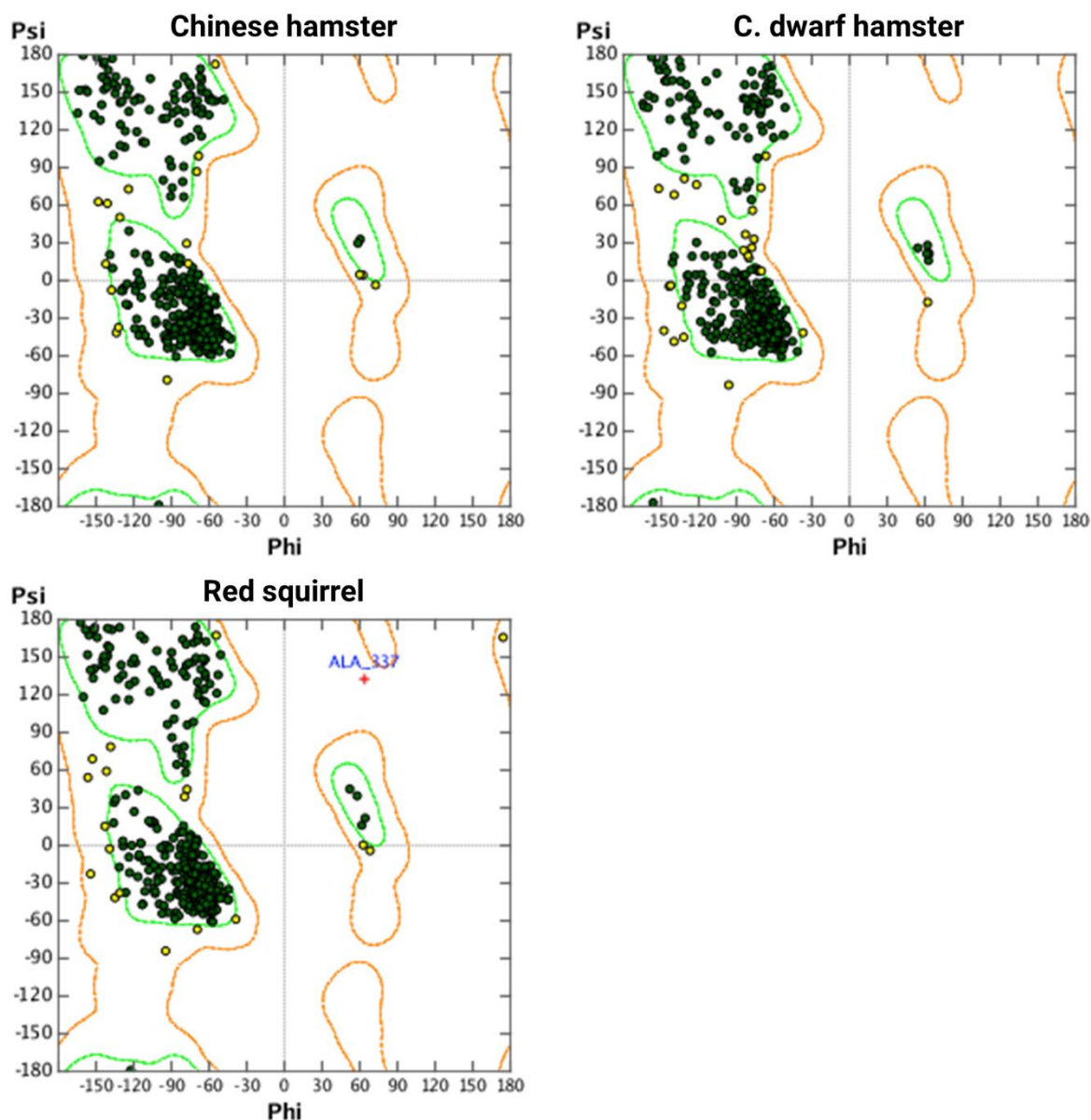
Appendix Figure 13. Ramachandran plot of the orthologs of animal angiotensin-converting enzyme 2 homology models. Symbol code: green point- residue with favorable geometry, yellow point- residue with allowed geometry, red cross- outlier. (The subfigures are presented on the previous page)