Heinrich Heine
Universität
Düsseldorf

Computational Workflows to Capture Molecular Flexibility in the Design of Peptide-Based Therapeutics

Inaugural Dissertation

presented to the Faculty of Mathematics and Natural Sciences
of Heinrich Heine University Düsseldorf
for the degree of Doctor of Natural Sciences
by

Lara Scharbert
from Cologne, Germany

Cologne, July 2026

From the Institute for Biological Information Processing: Structural Biochemistry (IBI-7)
at the Forschungszentrum Jülich

Published by permission of
the Faculty of Mathematics and Natural Sciences of
Heinrich Heine University Düsseldorf

1. Reviewer: Prof. Dr. Birgit Strodel
2. Reviewer: Jun.-Prof. Dr. Jan Meisner

Date of the oral examination: September 4, 2026

Declaration

Hiermit versichere ich, Lara Scharbert, an Eides Statt, dass diese Dissertation mit dem Titel '*Computational Workflows to Capture Molecular Flexibility in the Design of Peptide-Based Therapeutics*' von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der 'Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf' erstellt worden ist.

I, Lara Scharbert, hereby declare under oath that I have produced my thesis entitled '*Computational Workflows to Capture Molecular Flexibility in the Design of Peptide-Based Therapeutics*' independently and without any undue assistance by third parties under consideration of the 'Principles for the Safeguarding of Good Scientific Practice at Heinrich Heine University Düsseldorf'.

Düsseldorf, _____ _____

Lara Scharbert

Contents

Acknowledgements IX

Abstract XI

Publications XIII

List of Abbreviations XV

List of Figures XXI

List of Tables XXV

1. Introduction 1
   1.1. Peptides as Therapeutic Agents 1
   1.2. Computer-aided Peptide Drug Design 3
      1.2.1. Pose Generation 4
      1.2.2. Pose Evaluation 6
      1.2.3. Rational Design 7
   1.3. Therapeutic Targets 9
      1.3.1. Coronavirus Disease 2019 10
      1.3.2. Prostate Cancer 16
   1.4. Aim of the Thesis 23

2. Methods 25
   2.1. Modeling Peptide–Protein Interactions 25
      2.1.1. Protein Preparation 25
      2.1.2. Molecular Docking 26
      2.1.3. Molecular Dynamics Simulations 35
      2.1.4. Binding Free Energy Calculations 41
   2.2. Computational Dynamic Saturation Mutagenesis 42
   2.3. Experimental Evaluation of Peptide Inhibitors 42
      2.3.1. Protein Production and Purification 42
      2.3.2. Determination of Peptide Inhibitory Activity 45

3. Binding Mode of SARS-CoV-2 M^{pro} Retro-Inverso D-Peptide Inhibitors 47
3.1. Methods . . . 48
3.2. Results . . . 50
3.2.1. Sampling Peptide Conformations . . . 50
3.2.2. M^{pro}–Peptide Complex Generation . . . 51
3.2.3. Simulation of M^{pro}–Peptide Complexes . . . 52
3.2.4. Extensive Analysis of 3CVL and 3CVLri Binding Modes . . . 53
3.3. Discussion . . . 59

4. Computer-Aided Peptide Inhibitor Design against SARS-CoV-2 Proteases 61
4.1. Methods . . . 62
4.2. Strategy I: All-D Retro-Inverso Versions of Substrate Peptides . . . 65
4.2.1. Main Protease . . . 66
4.2.2. Papain-like Protease . . . 72
4.3. Strategy II: Peptide–Protease Complex Library . . . 76
4.3.1. Expanding the Library with M^{pro} Peptide Inhibitors . . . 76
4.3.2. Design of Novel M^{pro} Peptide Inhibitors . . . 78
4.4. Strategy III: Computational Dynamic Saturation Mutagenesis . . . 80
4.4.1. M^{pro} Inhibitor Design. . . . 80
4.4.2. PL^{pro} Inhibitor Design. . . . 84
4.5. Experimental Validation of the Peptide Designs . . . 86
4.5.1. Inhibitory Activity of Peptide Designs against M^{pro} . . . 86
4.5.2. Correlations between *In Vitro* and *In Silico* . . . 89
4.5.3. Inhibitory Activity of Peptide Designs against PL^{pro} . . . 91
4.6. Discussion . . . 92

5. A Modeling Workflow for PSMA-targeted Radiopharmaceuticals 101
5.1. Methods . . . 102
5.2. Results . . . 105
5.2.1. Conformational Differences of Two PSMA Crystal Structures . . . 105
5.2.2. The Effect of PSMA Binding Site Hydration on Docking . . . 106
5.2.3. Pose Generation of Clinically Relevant PSMA Ligands . . . 108
5.2.4. Evaluation of Ligand Poses for Subsequent MD Simulations . . . 112
5.2.5. Dynamic Investigation of PSMA–Ligand Interactions . . . 116
5.3. Discussion . . . 126

6. Conclusion 131

A. Appendix **169**
A.1. Supplementary Figures 169
A.2. Publication I 201
A.3. Publication II 247
A.4. Publication III 257

Acknowledgements

First and foremost, I want to thank Prof. Dr. Birgit Strodel for her supervision, insightful guidance, and constant support throughout my PhD. She gave me the freedom to pursue my own ideas, and her door was always open whenever I needed advice, feedback, or a fresh perspective. I am truly grateful and honored to be part of her research group.

My sincere thanks also go to Jun.-Prof. Dr. Jan Meisner for kindly agreeing to be my second reviewer and for taking the time to evaluate this thesis.

I am especially grateful to apl. Prof. Dr. Thomas Drepper for welcoming me to his group, giving me the opportunity to carry out the experimental part of my PhD under his supervision, and to the entire team for the kind and supportive atmosphere, which made my time in the lab truly enjoyable.

I want to express my gratitude to our collaborators at the Universitätsklinikum Düsseldorf, Prof. Dr. med. Frederik L. Giesel and Dr. Jens Cardinale, whose expertise in nuclear medicine and many productive discussions greatly contributed to this work.

I would also like to thank Dr. Mônika Coronado and Dr. Raphael Eberle for the fruitful collaboration and for sharing their knowledge of peptide inhibitors and assay development.

I am grateful for the financial support of this work from the Volkswagen Foundation under the initiative 'Corona Crisis and Beyond' (Project No. 99761). I also acknowledge the computing time granted through JARA-HPC (projects 3CLDTP and rdrp) on the supercomputer JURECA at Forschungszentrum Jülich, as well as the resources provided by the Centre for Information and Media Technology at Heinrich Heine University Düsseldorf.

On a personal note, I want to thank Stephan Thies and Rebecka Molitor for the many fun lunches at the Seecasino and for their scientific and mental support through countless helpful conversations. I am also deeply grateful to Anna Jäckering, who has been not only a colleague but a close friend, always ready to help with patience and understanding, whether the challenge was scientific or personal. I would like to thank Moritz Schäffler for his consistently positive and motivating spirit and for proofreading my thesis. Jennifer Loschwitz was always happy to help, regardless of the type of question or problem, and I deeply appreciate that.

Finally, from the bottom of my heart, I thank my family and friends for their unwavering support and belief in me, which carried me through the challenging times of my PhD

journey. My sister Tanja, for offering her creative perspective and a listening ear whenever I needed one. And especially my parents, who have supported me immensely throughout my studies with boundless love, patience and encouragement, and to whom I could always turn, no matter the circumstances.

Abstract

Peptides have emerged as a highly promising class of therapeutics, offering unique advantages such as high specificity, versatility, and low toxicity. Experimental advances in modification strategies have addressed key limitations of poor proteolytic stability and rapid clearance, facilitating the clinical translation of peptide-based therapeutics. However, computational peptide design remains challenging despite rapidly developing computational methods, including artificial intelligence tools such as AlphaFold. The intrinsic flexibility of peptides complicates reliable prediction of peptide–protein structures and interactions, which is essential for structure-based drug design.

To address this challenge, we developed computational workflows combining diverse strategies, including docking and molecular dynamics (MD) simulations, to accurately model and design peptide-based therapeutics. We focused on the following target systems: the main protease (M^{pro}) and papain-like protease (PL^{pro}) of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), modeled with natural substrates and peptide inhibitors, and the prostate specific membrane antigen (PSMA) in complex with peptide-based radiopharmaceuticals.

We developed the computational dynamic saturation mutagenesis (CDSM) protocol for automated saturation mutagenesis at each peptide position using MD simulation snapshots to capture the flexibility of peptide–protein interactions. The resulting binding free energy deviations from parent residues served as design suggestions, enabling the generation of an 8-amino acid peptide inhibitor with high binding affinity for M^{pro}, validated by *in vitro* inhibition assays ($IC_{50} = 4.8 \pm 0.5\,\mu M$).

To elucidate the structure–activity relationship of PSMA radioligands, we investigated the binding modes of eight clinically relevant radioligands using a workflow that combines docking at varying hydration levels, re-scoring of docking poses, and MD simulations to generate accurate and stable ligand poses. To date, this represents the most comprehensive computational modeling study of PSMA radioligands.

The computational modeling and design strategies presented in this thesis—especially the CDSM protocol, which is broadly applicable to various peptide–protein systems, as well as the workflows for elucidating PSMA–radioligand interactions—lay the groundwork for future development of peptide-based therapeutics.

Publications

[I] Eberle, R. J., Sevenich, M., Gering, I., Scharbert, L., Strodel, B., Lakomek, N. A., Santur, K., Mohrlüder, J., Coronado, M. A., and Willbold, D. (2023) Discovery of All-D-Peptide Inhibitors of SARS-CoV-2 3C-like Protease. *ACS Chem Biol.*; doi:10.1021/acschembio.2c00735.

[II] Mamlins, E., Scharbert, L., Cardinale, J., Krotov, M., Winter, E., Rathke, H., Strodel, B., Ankrah, A. O., Sathekge, M., Haberkorn, U., Kratochwil, C., and Giesel, F. L. (2023) The Theranostic Optimization of PSMA-GCK01 Does Not Compromise the Imaging Characteristics of [^{99m}Tc]Tc-PSMA-GCK01 Compared to Dedicated Diagnostic [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA in Prostate Cancer. *Mol Imaging Biol.*; doi:10.1007/s11307-023-01881-y.

[III] Scharbert, L. and Strodel, B. (2025) Innovative strategies for modeling peptide–protein interactions and rational peptide drug design. *Curr Opin Struct Biol.*; doi:10.1016/j.sbi.2025.103083.

List of Abbreviations

K_d equilibrium dissociation constant

3CL^{pro} 3-chymotrypsin-like protease

ABS arene-binding site

ACE acetyl

AD4 AutoDock4

ADCP AutoDock CrankPep

ADFR AutoDockFR

ADMET absorption, distribution, metabolism, excretion and toxicity

ADT androgen deprivation therapy

AF2multi AlphaFold-Multimer

AF3 AlphaFold 3

AGFR AutoGridFR

AI artificial intelligence

AMBER assisted model building with energy refinement

aMD accelerated MD

Amp100 100 μg/ml ampicillin

APBS adaptive Poisson-Boltzmann solver

ATP adenosine triphosphate

BFGS Broyden-Fletcher-Goldfarb-Shanno

BUDE Bristol University Docking Engine

CADD computer-aided drug design

CAGR compound annual growth rate

CAPID computer-aided peptide inhibitor design

CARD computer-aided radiopharmaceutical design

CAS computational alanine scanning

CD circular dichroism

CDSM computational dynamic saturation mutagenesis

CHARMM chemistry at Harvard macromolecular mechanics

COVID-19 coronavirus disease 2019

CRPC castration-resistant prostate cancer

CSD Cambridge Structural Database

CT computed tomography

CV column volumes

DMSO dimethyl sulfoxide

DOTA 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid

DOTAGA 1,4,7,10-tetraazacyclododececane,1-(glutaric acid)-4,7,10-triacetic acid

Dri retro-inverso D-enantiomeric

DSSP dictionary of protein secondary structures

DTT Dithiothreitol

EAU European Association of Urology

E. coli *Escherichia coli*

EDTA ethylenediaminetetraacetic acid

EPS electrostatic potential surface

ESI-MS electrospray ionization mass spectrometry

FDA Food and Drug Administration

FEP free energy perturbation

FFT fast fourier transform

FNA 6-fluoronicotinic acid

GA genetic algorithm

GaMD Gaussian accelerated MD

ΔG_{bind} binding free energy

GCPII glutamate carboxypeptidase II

GLP-1-RA glucagon-like peptide-1 receptor agonists

GROMOS Groningen molecular simulation

HBED-CC N,N′-Bis[2-hydroxy-5-(ethylene-β-carboxy)benzyl]ethylenediamine-N,N′-diacetic acid

HBED$^-$ N,N′-Bis(2-hydroxybenzyl)ethylenediamine-N,N′-diacetic acid

HCoVs human coronaviruses

HEPES 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HREMD Hamiltonian replica exchange molecular dynamics

IC_{50} half-maximal inhibitory concentration

ILS Iterated Local Search

IMAC immobilized metal ion affinity chromatography

ipTM interface predicted Template Modeling score

ISG15 interferon-induced gene 15

LB lysogeny broth

LBDD ligand-based drug design

LJ Lennard-Jones

M^{pro} main protease

MAG_3 mercaptoacetyl triglycine

MAS_3 mercaptoacetyl tri-D-serine

MC Monte Carlo

MD molecular dynamics

MDT metastasis-directed therapies

MERS-CoV middle east respiratory syndrome-related coronavirus

MES 2-(N-morpholino)ethanesulfonic acid

ML machine learning

MM/GBSA molecular mechanics generalized Born surface area

MM/PBSA molecular mechanics Poisson–Boltzmann surface area

mpMRI multiparametric magnetic resonance imaging

mRNA messenger ribonucleic acid

MSA multiple sequence alignment

NAAG N-acetyl-aspartyl-glutamate

NAALADase I N-acetyl-L-aspartyl-L-glutamate peptidase I

$natM^{pro}$ natural M^{pro}

Ni-NTA nickel nitrilotriacetic acid

NME N-methyl amide

NMR nuclear magnetic resonance

OD optical density

PBC periodic boundary conditions

PDB Protein Data Bank

Pep-GaMD Peptide Gaussian Accelerated Molecular Dynamics

pepPIs peptide–protein interactions

PET positron emission tomography

PL^{pro} papain-like protease

pLDDT predicted Local Distance Difference Test

PME particle mesh Ewald

PPIs protein–protein interactions

PreSaVS predictive saturation variation scanning

PSA prostate-specific antigen

PSMA prostate specific membrane antigen

QM quantum mechanical

QSAR quantitative structure-activity relationship

RdRp RNA-dependent RNA polymerase

REMD replica exchange MD

RESP restrained electrostatic potential

RGS radioguided surgery

rh radiohybrid

RLT radioligand therapy

RMSD root mean square deviation

RMSF root mean square fluctuation

RNA ribonucleic acid

RP-HPLC reverse-phase high-performance liquid chromatography

RTC replication and transcription complex

SAR structure–activity relationship

SARS-CoV severe acute respiratory syndrome coronavirus

SARS-CoV-2 severe acute respiratory syndrome coronavirus 2

SASA solvent-accessible surface area

SBDD structure-based drug design

SD standard deviation

SDS-PAGE sodium dodecyl sulfate polyacrylamide gel electrophoresis

SEM standard error of the mean

SiFA silicon–fluoride–acceptor

SPECT single-photon emission computed tomography

SPR surface plasmon resonance

TCEP tris(2-carboxyethyl)phosphine

TEV Tobacco Etch Virus

TI thermodynamic integration

Triton X-100 2-[4-(2,4,4-trimethylpentan-2-yl)phenoxy]ethanol

Ubl ubiquitin-like

vdW van der Waals

Vina AutoDock Vina

WHO World Health Organization

List of Figures

1.1. Differences in protein binding site structures. 3
1.2. Workflow of a typical computer-aided peptide inhibitor design study. . . . 5
1.3. Mechanism for the conversion of L-peptides into retro-inverso D-peptides. 9
1.4. Cleavage sites of SARS-CoV-2 M^{pro} and PL^{pro} within the viral polyproteins pp1a and pp1ab. 11
1.5. Structural features of SARS-CoV-2 M^{pro}, including domain organization, dimerization, and substrate recognition. 13
1.6. Structural features of SARS-CoV-2 PL^{pro}. 15
1.7. PSMA-targeted radioligand therapy and diagnostic. 17
1.8. Schematic representation of the homodimeric PSMA transmembrane protein and its binding cavity 19
1.9. Clinically relevant Glu-ureido-based PSMA radioligands. 21

2.1. Representation of the AutoDock docking grid 30
2.2. Mechanisms of the genetic algorithm. 31
2.3. Docking algorithm of AutoDock CrankPep (ADCP). 33
2.4. Bonded and nonbonded interactions represented by the force-field potential functions. 37

3.1. Secondary structure analysis of the MD simulations with unbound inhibitory L- and Dri-peptides. 50
3.2. Docking solutions and starting structures for MD simulations in complex with M^{pro}. 52
3.3. Change of the active site volume of M^{pro} in complex with 3CVL and 3CVLri peptides 54
3.4. Binding modes of the competitive 3CVL(ri)-2 and 3CVL(ri)-7 peptide inhibitors and their interactions with M^{pro} active site residues. 55
3.5. Per-residue binding free energy decomposition of the 3CVL and 3CVLri peptides simulated in complex with M^{pro}. 56
3.6. Binding modes of the non-competitive 3CVL(ri)-4 peptide inhibitors and their interactions with M^{pro} residues of a supposed allosteric binding site. 58

4.1. Substrate peptides of PL^{pro} and M^{pro} extracted from polypeptide cleavage sequences. . . . 65
4.2. Re-scored docking solutions of L- and Dri-peptides in complex with M^{pro} . 66
4.3. Stability and flexibility of L- and Dri-peptides in complex with M^{pro} during MD simulations. . . . 68
4.4. Hydrogen bond existence for L- and Dri-peptides simulated in complex with M^{pro}. . . . 69
4.5. Comparison between co-crystallized substrate peptide nsp4/5, L-m01 and Dri-m01 forming hydrogen bonds with M^{pro} 70
4.6. Per-residue binding free energy decomposition of L- and Dri-peptides simulated in complex with M^{pro}. . . . 71
4.7. Hydrogen bond analysis and binding free energy calculations of L- and Dri-peptides simulated in complex with PL^{pro}. . . . 74
4.8. Comparison between L- and Dri-peptide p02 forming hydrogen bonds with PL^{pro} 75
4.9. Peptide design process based on binding affinities of M^{pro}–peptide library MD simulations. . . . 79
4.10. Computational Dynamic Saturation Mutagenesis protocol for peptide design involving peptide–protein MD simulations. . . . 81
4.11. Computational Dynamic Saturation Mutagenesis of the M^{pro}–m01 complex MD simulation. . . . 82
4.12. CDSM-guided rational peptide inhibitor design against M^{pro}. . . . 83
4.13. CDSM-guided rational peptide inhibitor design against PL^{pro}. . . . 85
4.14. Electrostatic interactions of the peptide designs z07 and z08 with PL^{pro}. . 86
4.15. M^{pro} inhibition after incubation with peptide designs. . . . 87
4.16. Estimation of IC_{50} for the designs a10 and c05. . . . 88
4.17. Correlation between *in silico* ΔG_{bind} estimation, *in vitro* inhibition potency and peptide design features. . . . 89
4.18. Interaction pattern of peptide designs a10 and c04 with M^{pro}. . . . 90
4.19. PL^{pro} inhibition after incubation with peptide designs. . . . 91

5.1. Docking conditions and re-scoring parameters for PSMA: binding site water molecules and vector-to-residue distances. . . . 103
5.2. Bonded and non-bonded modeling of the metal sites. . . . 104
5.3. Loop rearrangement in PSMA induced by PSMA-1007 opening the arene-binding site. . . . 105
5.4. Redocking of PSMA-617 and PSMA-1007 to the crystal structure of PSMA with different levels of hydration in the binding pocket. . . . 107

5.5. Chemical structures of modeled PSMA ligands. 109
5.6. Top-ranked docking scores of PSMA ligands. 111
5.7. Docking solutions of PSMA-GCK01. 113
5.8. Docking solutions of PSMA-I&S. 115
5.9. Interaction analysis of PSMA–ligand MD simulations. 118
5.10. Flexibility and interaction pattern of PSMA-GCK01. 121
5.11. Flexibility and interaction pattern of PSMA-I&S. 122
5.12. Flexibility and interaction pattern of rhPSMA-7. 124

A1. Plasmid maps for recombinant protein expression in *E. coli*. 169
A2. Evaluation of peptide purity (A2–A14) 169
A15. Volume of the active sites in the unbound M^{pro} dimer MD simulation. . . 177
A16. Docking solutions of the peptide structures (cl01–10) and selection criteria for starting structures of MD runs 1–3 in complex with M^{pro}. 177
A17. Conformational changes of 3CVL(ri) peptides in complex with M^{pro} during different MD runs. 178
A18. Binding free energy estimation of the M^{pro}-3CVL(ri) complex MD simulations 178
A19. Binding modes from different MD runs of M^{pro} in complex with competitive peptide inhibitors 179
A20. Binding modes from different MD runs of M^{pro} in complex with non-competitive peptide inhibitors 180
A21. Conformational changes of 3CVL(ri) peptides in complex with M^{pro} during extensive MD simulations. 180
A22. Total binding free energy and change in active site volume of M^{pro}-peptide complex simulations. 181
A23. Co-crystallized M^{pro} substrate peptides. 181
A24. Binding affinities of L- and Dri-peptides m01–m11 with M^{pro} estimated by Docking and MD simulations. 182
A25. Electrostatic potential surface of M^{pro} and PL^{pro}. 183
A26. Comparison of two PL^{pro} crystal structures and their suitability as receptors for peptide docking. 184
A27. Solutions for PL^{pro}–peptide complex generation with L- and Dri-peptides via a consensus docking approach. 185
A28. Stability and flexibility of L- and Dri-peptides in complex with PL^{pro} during MD simulations. 186
A29. Substrate peptides p01 and p03 forming hydrogen bonds with PL^{pro} . . . 187

A30. Stability and flexibility of peptide inhibitors p12–p16 in complex with M^{pro} during MD simulations. 187
A31. Binding free energy decomposition and structural flexibility of the M^{pro} inhibitors 3CVL(ri). 188
A32. Stability of the peptide designs a01–a21 and b01–b07 from strategy II in complex with M^{pro} during MD simulations. 189
A33. Stability and flexibility of the peptide designs c01–c07 from strategy III in complex with M^{pro} during MD simulations. 189
A34. CDSM with MD simulations of PL^{pro} in complex with p01–p03. 190
A35. Stability and flexibility of the peptide designs z04–z08 from strategy III in complex with PL^{pro} during MD simulations. 190
A36. SDS-PAGE with samples from the production steps of natural M^{pro}. . . . 191
A37. Hydrogen bond existence for all M^{pro} peptide designs tested experimentally. 191
A44. Additional S1 accessory hydrophobic pocket (R-patch pocket) addressed by PSMA inhibitor DCIBzL. 195
A45. The remote arene binding site accessed by PSMA-1007. 195
A38. Docking solutions of PSMA ligands (A38–A43) 196
A46. Flexibility and interaction patterns of PSMA ligands (A46–A51) 196
A52. Per-residue ΔG_{bind} decomposition of PSMA–ligand MD simulations. . . . 199

List of Tables

3.1. Peptide inhibitors targeting SARS-CoV-2 M^{pro}. 47

4.1. Peptide designs purchased for testing against SARS-CoV-2 PL^{pro} and M^{pro}. 64

5.1. Summary of PSMA–ligand MD simulations: Starting conditions and results of binding free energy calculations. 117

1. Introduction

The following sections are partly adapted from Publication III (Section A.4):

Innovative strategies for modeling peptide–protein interactions and rational peptide drug design
Lara Scharbert and Birgit Strodel
Current Opinion in Structural Biology, 2025, DOI:10.1016/j.sbi.2025.103083

1.1. Peptides as Therapeutic Agents

Peptides play central roles in a wide range of cellular processes, particularly within signaling pathways. They function as hormones, growth factors, neurotransmitters, or antimicrobial agents[1–3], and most of these activities are mediated through peptide–protein interactions (pepPIs). Because they must enable rapid cellular responses to stimuli, pepPIs are typically transient, allowing them to be formed and disrupted easily while maintaining high specificity[4]. Owing to their essential roles in physiological processes, combined with high specificity and low toxicity, peptides have emerged as a highly promising class of therapeutics[5,6]. Although therapeutic peptides currently represent only 7.1 % of the global pharmaceutical market, their revenues have doubled over the past decade and are projected to grow at a compound annual growth rate (CAGR) of 10.8 % from 2025 to 2030[7,8], exceeding that of small molecules (5.1 %[9]) and biologics (8.5 %[10]).
Small molecules still dominate the global pharmaceutical market, accounting for 54.7 % of revenues in 2024[8], as they offer several advantages, including low production costs, oral bioavailability, and good membrane permeability. Their small size, however, limits their ability to disrupt interactions involving large protein surfaces, such as protein–protein interactions (PPIs). While biologics can overcome this limitation, they are often associated with high production costs and potential immunogenicity[5]. Therapeutic peptides are able to bridge the gap between these two classes: they are generally less immunogenic and easier to produce than biologics, while—unlike small molecules—their larger size and intrinsic flexibility allow them to target PPIs that have long been considered ‘undruggable’[11].
Currently, therapeutic peptides are primarily used in the treatment of metabolic disorders

and cancer. The largest share of the peptide drug market is represented by insulin and its analogues[12], followed by glucagon-like peptide-1 receptor agonists (GLP-1-RA) such as dulaglutide, liraglutide, and semaglutide[5,13]. Historically, insulin also marks the beginning of peptide drug research, which dates back to the early 20th century with its discovery and first clinical use. At that time, scientists were restricted to isolating peptides from natural sources, most commonly animal tissues. This approach was limited by low yields, inconsistent purity, and the risk of adverse immune responses, including allergic reactions. With the discovery of recombinant technology in the 1980s, it became possible to produce larger quantities of peptides, including longer sequences, under controlled and clean conditions[14]. Another significant advance in peptide drug discovery was the development of target-based screening strategies. Among these, phage display enables the construction of large peptide libraries through genetic engineering of bacteriophages—ranging from millions to billions of variants—each displaying different peptides for selection against specific protein targets[15].

However, therapeutic peptides intrinsically exhibit two main limitations: their membrane impermeability, which restricts access to intracellular targets, and their short metabolic half-lives due to rapid proteolytic degradation and renal clearance[5,6]. To overcome these limitations, numerous strategies have been developed to enhance peptide stability, including chemical modifications (e.g., acetylation/amidation and the incorporation of non-natural or D-amino acids), conjugation strategies (e.g., lipidation, PEGylation, and fusion proteins), and structural stabilization (e.g., stapling or cross-linking)[16]. When peptides are chemically modified, they are generally referred to as peptidomimetics. According to Grossmann[17], they can be grouped into four main classes: (A) modified peptides with only limited side-chain or backbone alterations; (B) peptides and foldamers incorporating more extensive modifications, such as unnatural amino acids, small-molecule fragments, or major backbone changes; (C) structural mimetics in which the peptide backbone is fully replaced by a small-molecule scaffold, while substituents preserve the orientation of key native side chains; and (D) mechanistic mimetics that share little structural similarity with the parent peptide but reproduce its mode of action. Peptidomimetics investigated in this thesis include all-D-enantiomeric peptides and amino acid-based radiopharmaceuticals, which can be grouped into categories B and C, respectively.

Rational design has been greatly aided by improvements in computational algorithms and processing power, together with the growing availability of structural information, which have established computer-aided drug design (CADD) as an invaluable tool in modern drug discovery and development. These capabilities enable rapid virtual screening of vast compound libraries and provide detailed insights into protein–ligand interactions, offering an efficient alternative to the time- and resource-intensive experimental high-throughput

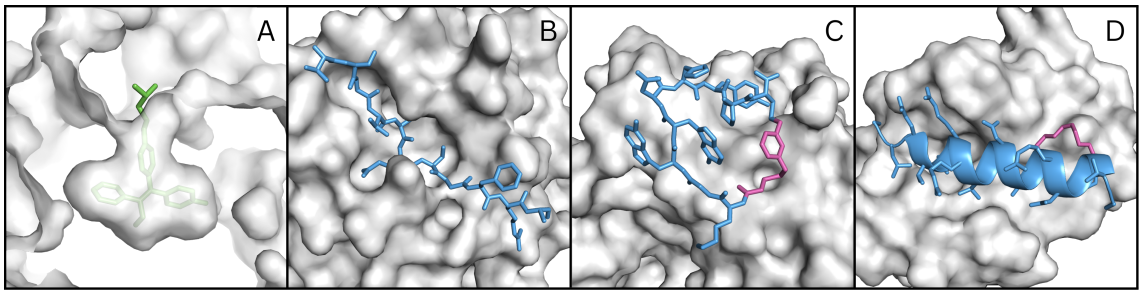


Figure 1.1.: Differences in protein binding site structures. Representative examples of different protein binding sites occupied by (A) a small molecule inhibitor (PDB: 3ERT), (B) an extended linear peptide (PDB: 7T70), (C) a cyclic peptide (PDB: 6XIB), and (D) a stapled helical peptide (PDB: 3MK8). The proteins are shown as gray surface and the ligands as sticks, with green for the small molecule, blue for the peptides, purple for peptide modifications, and a blue cartoon for the helix. Reproduced with permission from Scharbert and Strodel[18].

screening[19]. Successful CADD requires a detailed molecular understanding of the pepPIs, which depends on reliable structural information. While over 200,000 protein structures have been experimentally resolved and are available in the Protein Data Bank (PDB) through high-resolution techniques such as X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy, these methods are often labor-intensive, costly, and can produce non-native conformations[20]. In 2021, AlphaFold marked a turning point in structural biology, using artificial intelligence (AI) to predict protein structures from amino acid sequences with remarkable accuracy. Today, the AlphaFold server allows anyone to generate high-quality models within minutes[21,22].

While target protein structures are often available in databases, elucidating peptide–protein complexes remains a particular challenge. Because of their intrinsic flexibility, peptides can adopt a multitude of conformations, challenging both experimental characterization and computational modeling of pepPIs. This challenge is exacerbated by the distinct binding mode of peptides. Whereas small molecules typically occupy well-defined, deeply buried pockets, peptides often engage broad and less ordered regions on protein surfaces (Fig. 1.1A vs. B–D). To capture their conformational complexity, modeling techniques such as molecular dynamics (MD) simulations can be employed to explore the dynamic nature of pepPIs, setting an important basis for computer-aided peptide inhibitor design (CAPID) studies[19,23].

1.2. Computer-aided Peptide Drug Design

CADD integrates computational techniques to facilitate the discovery and optimization of lead compounds, by improving their affinity, specificity, and pharmacokinetic properties. It utilizes structural knowledge of either the biological target (structure-based drug design

(SBDD)) or known bioactive ligands (ligand-based drug design (LBDD)) to identify and prioritize promising candidate drugs[24]. Common ligand-based design techniques include quantitative structure-activity relationship (QSAR) and pharmacophore-based methods. QSAR studies are based on the principle that a compound's bioactivity correlates with its structural and molecular properties, enabling the generation of statistical models to predict the biological activity of novel compounds[25]. Pharmacophore models represent the three-dimensional arrangement of chemical features responsible for a small molecule's activity. Screening these models allows the identification of structurally diverse compounds that maintain the same key interactions with the target[26]. In the following, we focus on SBDD of peptide inhibitors, presenting a workflow that generates poses by molecular docking, evaluates them through MD simulations, and applies the resulting data to guide rational design of potent candidates (Fig. 1.2).

1.2.1. Pose Generation

In SBDD, access to a high-resolution structure of the therapeutic target protein is essential. Such structures can often be obtained from the PDB, typically as X-ray crystal or NMR models. If no experimental structure is available, computational prediction methods provide an alternative. For many years, homology modeling—based on the principle that evolutionarily related proteins adopt similar folds—was the standard approach. This changed with the introduction of AlphaFold[21], which employs deep learning models trained on large-scale databases of experimentally determined protein structures to produce highly accurate structural predictions[20].

To predict the structure of protein–ligand complexes, molecular docking is the standard approach. It models ligand binding by sampling conformations and orientations within the protein and ranking them with a scoring function to identify the most probable binding modes. When the binding site is known, local docking restricts sampling to that region, offering higher efficiency due to the reduced search space. In contrast, global docking explores the entire protein surface to identify potential binding sites, enabling predictions without prior knowledge of the binding pocket. Peptides, however, present a particular challenge for conventional docking programs due to their large number of rotatable bonds and high conformational flexibility, which distinguish them from both rigid small molecules and folded proteins[23].

In the work of Weng et al. (2020), they compared 14 docking programs with regard to their performance to predict the structure of pepPIs. Three protein–protein (ZDOCK[27], FRODOCK[28], and HawkDock[29]), three small molecule–protein (GOLD[30], Surflex-Dock[31], and AutoDock Vina (Vina)[32]) and eight peptide–protein docking programs (GalaxyPepDock[33], MDockPeP[34], HPEPDOCK[35], CABS-dock[36], pepATTRACT[37],

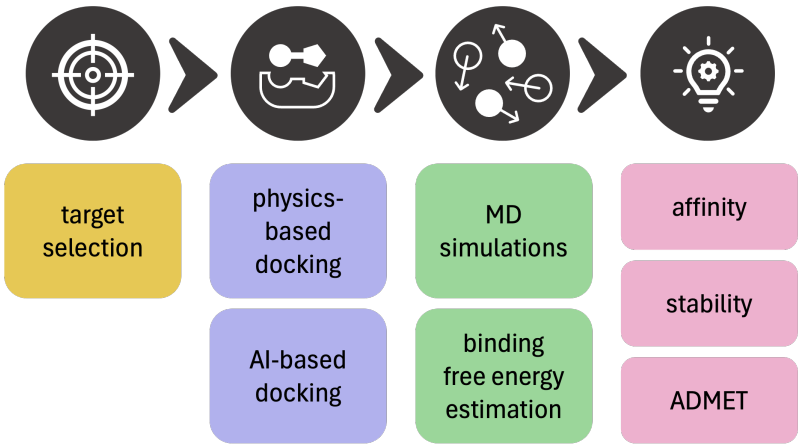


Figure 1.2.: Workflow of a typical computer-aided peptide inhibitor design study. The process begins with the selection of a drug target, including binding site identification, followed by complex generation via molecular docking. In a consensus approach, physics-based docking is combined with AI-driven pepPIs prediction. The resulting peptide poses are evaluated through MD simulations, which investigate the peptide flexibility at the binding site and enable statistical estimation of binding free energies. Finally, insights from pepPIs modeling guide the rational design of peptide candidates to improve affinity, structural stability, and pharmacological activity. Reproduced with permission from Scharbert and Strodel[18].

DINC[38], AutoDock CrankPep (ADCP)[39], and HADDOCK[40] peptide docking) were evaluated based on their performance on a set of 185 peptide–protein complexes with peptides consisting of 5-20 residues[41]. They found that peptide–protein docking algorithms, overall, performed better than protein–protein or small molecule docking algorithms. In global docking, the web server HPEPDOCK, which uses fast peptide conformation sampling and ensemble docking techniques, generated the best predictions for the entire data set.

In local docking, ADCP presented the best overall results. ADCP is a peptide docking program specifically designed to model flexible peptides binding to protein receptors. It integrates CRANKITE, which efficiently samples the conformational space of peptides and proteins[42,43], with AutoDock's grid-based representation of a rigid receptor[44], allowing simultaneous optimization of peptide conformations and their interactions with the receptor[39]. The conformational space is explored using a Metropolis Monte Carlo (MC) search protocol adapted from CRANKITE, which folds the peptide *de novo* within the receptor affinity grid while generating docking poses that capture both peptide flexibility and receptor binding[39].

In recent years, AI-based approaches have transformed protein structure prediction, with AlphaFold[21] setting the benchmark for state-of-the-art performance. The introduction of AlphaFold-Multimer (AF2multi)[45], an extension of AlphaFold 2, expanded its capabilities to predict protein–protein complexes. Although primarily developed for protein–

protein complex prediction, AF2multi has also demonstrated good performance in peptide–protein docking tasks[46,47]. Shanker and Sanner compared AF2multi with OmegaFold, AlphaFold 2 Monomer, and their peptide-docking software ADCP. AF2multi achieved the highest performance with a 53 % success rate for top-ranked solutions, while ADCP reached only 23 %. Considering all generated poses, however, ADCP's success increased to 62 %, which may point to limitations in its scoring function. The authors found that the two methods performed complementarily, each succeeding on different peptide–protein complexes. By combining them into a consensus docking strategy with an efficient selection mechanism, the authors achieved 60 % success for top-ranked and 66 % for the top five predictions[47]. With the release of the AlphaFold Server powered by AlphaFold 3 (AF3), peptide–protein complex prediction has become fast and easily accessible[22]. Recent benchmark studies indicate that AF3 achieves a performance comparable to the best-performing AF2multi in peptide–protein docking[48]. Docking approaches, however, are limited to static snapshots of peptide–protein interactions, neglecting their dynamic nature and especially the intrinsic flexibility of peptides, which can be addressed by MD simulations.

1.2.2. Pose Evaluation

MD simulations are a standard tool for studying the dynamic behavior of biological macromolecules, providing atomistic insight into flexibility and interaction mechanisms, which cannot be captured by static models. They calculate atomic trajectories by numerically solving Newton's equations of motion, with forces and potential energies being described by molecular mechanics force fields[49]. Widely used platforms such as NAMD[50], GROMACS[51], and AMBER[52] enable the realization of MD simulations. MD simulations provide a means to assess docking predictions by exploring the stability and conformational dynamics of peptide–protein complexes over time. This dynamic insight is essential for characterizing interaction modes and affinities of pepPIs and serves as a basis for subsequent drug design studies[19]. Recent applications of combined docking and MD simulations in therapeutic peptide development include designing peptides targeting multiple overexpressed receptors in tumor cells[53], generating peptide inhibitors of the Omicron severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike–ACE2 interaction[54], and optimizing a potent macrocyclic peptide to disrupt the PD-1/PD-L1 protein–protein interaction through rational design[55].

While docking methods usually rely on fast empirical or force-field-based scoring functions to estimate peptide–protein affinity, MD-based free energy calculations provide a more accurate assessment of binding affinities[56]. The most commonly used free energy calculation methods include molecular mechanics Poisson–Boltzmann surface area (MM/PBSA),

molecular mechanics generalized Born surface area (MM/GBSA), free energy perturbation (FEP), and thermodynamic integration (TI). End-point methods like MM/PBSA and MM/GBSA, are computationally efficient, requiring only a single trajectory, whereas pathway-based methods such as FEP and TI offer higher accuracy at the cost of greatly increased computational expense. As a result, MM/PB(GB)SA has become a widely used approach for estimating the binding free energy (ΔG_{bind}) between proteins and ligands[57]. Among the tools available for MM/PB(GB)SA calculations, Amber provides the widely used MMPBSA.py script within the AmberTools package, allowing free energy estimations based on MD snapshots[58]. To make MM/PB(GB)SA calculations more accessible to GROMACS users, Valdés-Tresanco developed gmx_MMPBSA[59], an adaptation of MMPBSA.py for trajectories and topology files generated with GROMACS. Compared to other GROMACS implementations, such as g_mmpbsa[60] and GMXPBSA2.1[61], gmx_MMPBSA offers broad compatibility across GROMACS versions and enhanced features, including ΔG_{bind} calculations with multiple solvation models, stability assessments, computational alanine scanning, ΔG_{bind} decomposition, and entropy corrections. It also supports Amber, OPLS, and CHARMM force fields and can be easily installed via Conda or AmberTools[59]. These features have led to the growing adoption of gmx_MMPBSA for analyzing pepPIs sampled by MD simulations in peptide drug discovery studies[62–65]

1.2.3. Rational Design

After identifying promising binding peptides, their affinity and stability need to be further optimized, which requires reliable and efficient computational tools. Computational alanine scanning (CAS) is a ΔG_{bind}-based approach that systematically substitutes amino acids with alanine to identify residues critical for the affinity and stability of pepPIs. Beyond identifying these 'hot spot' residues, CAS has provided insights into the structure–activity relationship (SAR) of peptide drug candidates, revealing which residues are essential for activity and which can be modified to improve affinity, specificity, or stability[66–68]. Per-residue ΔG_{bind} decomposition is another widely applied technique in computational peptide design, providing detailed insight into the energetic contributions of each amino acid within a peptide–protein complex. Residues that contribute little to binding can be targeted for modification, and the impact of these substitutions is assessed through subsequent ΔG_{bind} calculations. This approach has successfully guided the rational design of peptide variants with enhanced binding affinity[69,70]. Several tools are available for performing residue-level ΔG_{bind} analyses of pepPIs. The BAlaS web application offers a fast and user-friendly platform for conducting CAS, requiring only a PDB structure as input[71]. For analyses based on MD trajectories with MM/PB(GB)SA calculations, options include CompASM[72]—an Amber-based VMD[73] plug-in for CAS—and gmx_MMPBSA[59],

which supports both per-residue ΔG_{bind} decomposition and CAS. Although these methods can highlight positions suitable for modification, they do not directly indicate which substitutions will enhance activity. In a peptide design study based on prior modeling of SARS-CoV-2 main protease (M^{pro}) in complex with its substrates, Chan et al. (2021)[74] employed the computational saturation mutagenesis tool BUDE_SM[75] from the Bristol University Docking Engine (BUDE)[76] to systematically substitute each peptide residue with all 19 alternative proteinogenic amino acids. Subsequent *in vitro* inhibition assays confirmed the predictions, yielding potent peptide inhibitors against M^{pro}.

The inherent flexibility of peptides can compromise their binding affinity. To maintain an optimal binding pose, stabilization strategies such as cyclization are commonly employed. Cyclization further enhances stability by protecting against enzymatic degradation, hydrolysis, and immunogenicity, and can be achieved head to tail, side chain to side chain, or head to side chain, often with nonpeptidic linkers (Fig. 1.1C)[77,78]. Peptide cyclization is frequently applied to so-called 'hot segments', an extension of the 'hot spots' concept that defines continuous binding regions essential for PPIs[11]. Secondary structures like helices can also be stabilized by stapling, which cross-links two or more side chains, most often through disulfide bonds but also via synthetic linkers or chemical groups (Fig. 1.1D). Similar to cyclic peptides, stapled peptides often offer additional benefits, including improved cell penetration and increased resistance to proteolytic degradation[79,80].

A major challenge in peptide drug development has always been their short metabolic half-lives due to proteolytic cleavage and renal clearance[13,81]. A popular approach to enhance the proteolytic stability of peptide drugs is the incorporation of non-natural amino acids, such as D-amino acids. However, converting L-peptide inhibitors into all-D-isomers often reduces biological activity due to disrupted stereospecific pepPIs. A promising approach to preserve the binding mode of L-peptides is the design of retro-inverso D-enantiomeric (Dri) peptides, which consist of D-amino acids arranged in reverse order, maintaining side-chain orientation while inverting amide bonds and terminal groups (Fig. 1.3)[82]. This technique has led to the development of several peptide inhibitors with high activity and stability[65,83,84].

Early studies recognized renal clearance as a size-dependent mechanism, which can be overcome by increasing the molecular weight of a drug using conjugation techniques[16,85]. One prominent example is the design of GLP-1-RA, where attaching a fatty acid chain to the peptide enabled reversible binding to albumin. This modification prolonged the half-lives of drugs such as liraglutide and semaglutide by reducing enzymatic degradation and slowing renal clearance[86].

The development of effective and safe drugs requires careful evaluation of their absorption, distribution, metabolism, excretion and toxicity (ADMET) properties, thus *in silico* AD-

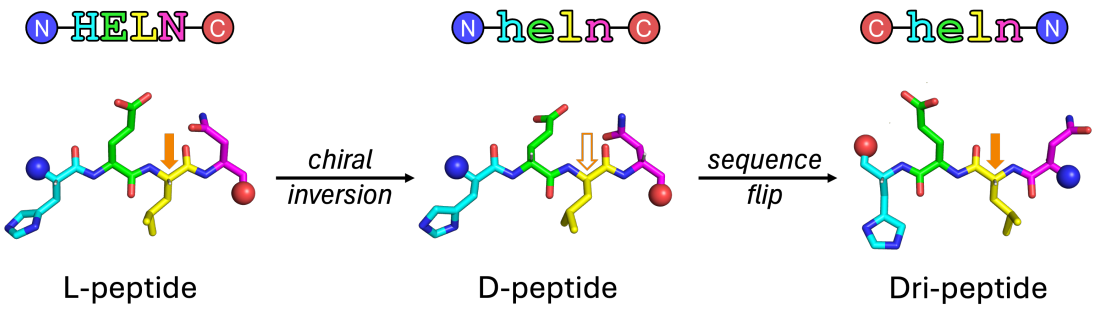


Figure 1.3.: Mechanism for the conversion of L-peptides into retro-inverso D-peptides. From left to right: A mirror image of the original L-peptide is first generated by flipping its structure along a chiral axis. The amino acid sequence of the resulting all-D-enantiomeric peptide is then reversed to obtain the Dri-peptide with inverted N/C termini (blue/red balls) and amide bonds. Orange arrows indicate the change in side chain orientation, either pointing away (full arrow) or toward the viewer (empty arrow), as exemplified by the Leu residue shown as yellow sticks. This demonstrates that Dri-peptides adopt the same side chain orientation as the original L-peptides.

MET prediction has become a critical component of pharmaceutical research[87]. Several web servers allow prediction of a wide range of ADMET properties, including solubility, bioavailability, blood-brain barrier penetration, and toxicity, and some can further assist in toxicity risk assessment and candidate selection[88,89]. However, most ADMET prediction tools are tailored to small molecules, and a dedicated platform addressing the specific ADMET challenges of peptide drugs is still lacking.

1.3. Therapeutic Targets

The selection of an appropriate biomolecular target in CADD is essential for successful drug development. According to the latest data from the ChEMBL database[90], the main classes of protein targets for bioactive compounds include enzymes, ion channels, and membrane receptors. However, the selection of a drug target is inherently context-dependent and varies based on the disease and the specific goal of the pharmaceutical, whether therapeutic or diagnostic. In general, a drug target can be selected based on its (i) druggability: the ability to bind drug-like molecules effectively, (ii) essentiality: its critical role in the survival or pathogenicity of the organism, and (iii) specificity: minimized off-target effects by being unique or sufficiently different from (other) host proteins[91]. In the following, we present the therapeutic targets used in this work: the two SARS-CoV-2 proteases, main protease (M^{pro}) and papain-like protease (PL^{pro}), and prostate specific membrane antigen (PSMA), a theranostic target in prostate cancer.

1.3.1. Coronavirus Disease 2019

Coronaviruses as part of the *Coronaviridae* are large, enveloped, single-stranded ribonucleic acid (RNA) viruses that infect humans, other mammals, and avian species, causing mainly respiratory and enteric diseases[92]. Human coronaviruses (HCoVs) such as HCoV-229E, HCoV-OC43, HCoV-NL63, and HCoV-HKU1 generally cause mild, seasonal respiratory infections. In contrast, severe acute respiratory syndrome coronavirus (SARS-CoV), middle east respiratory syndrome-related coronavirus (MERS-CoV), and SARS-CoV-2, which have emerged over the past two decades, can induce severe, life-threatening respiratory disease by infecting bronchial epithelial cells, pneumocytes, and upper respiratory tract cells in humans[93,94].

The outbreak of SARS-CoV-2 in December 2019 led to a global pandemic—known as coronavirus disease 2019 (COVID-19)—which had caused over 6.9 million deaths by the time the World Health Organization (WHO) declared the end of the emergency phase in May 2023[95,96]. Beyond the severe health crisis, the pandemic also had a profound impact on the global economy, including income loss, rising unemployment, and disruption across key industrial sectors[97]. This emergency situation led scientists from various disciplines to shift their focus to COVID-19 research and work together to achieve rapid advances in diagnostics, treatments, and vaccine development[98].

In this context, significant attention has been directed toward the trimeric spike protein, which decorates the surface of the virion and facilitates viral entry into human cells through the interaction between its receptor-binding domain (RBD) and the human angiotensin-converting enzyme 2 (ACE2)[99]. Neutralizing monoclonal antibodies—primarily targeting epitopes located at this critical interface—were among the first therapeutics used against SARS-CoV-2, aiming to prevent disease progression and reduce hospitalization[100]. The spike protein also played a major role in the development of most common anti-SARS-CoV-2 vaccines. The Pfizer–BioNTech messenger ribonucleic acid (mRNA) BNT162b2 (Comirnaty) and Moderna mRNA 1273 vaccines both use lipid nanoparticles to deliver mRNA encoding the full-length SARS-CoV-2 spike protein[101]. Compared to vector vaccines that use a modified, replication-deficient adenovirus for displaying the spike protein (e.g. Astra-Zeneca ChAdOx1-S or Janssen by Johnson & Johnson), the mRNA vaccines showed the highest efficacies in preventing COVID-19, with reported rates of 95 %/87.5 % (Pfizer–BioNTech) and 94.5 %/100 % (Moderna), for symptomatic and severe disease, respectively[102].

However, mutations in critical regions of the spike protein have been associated with altered transmissibility, infectivity, and reduced therapeutic efficacy of vaccines and antibody-based immunotherapies among emerging SARS-CoV-2 variants[103]. Thus, there remains the need for pharmaceuticals that effectively target also new variants with minimal side

effects[104]. To develop broadly effective therapeutics against current and emerging SARS-CoV-2 variants—and potentially other coronaviruses—the inhibitors need to target essential viral protein binding sites with conserved sequences and structures. Notable SARS-CoV-2 proteins that exhibit these features and present strong drug-target potential include the M^{pro}, also known as 3-chymotrypsin-like protease ($3CL^{pro}$), the PL^{pro} and the RNA-dependent RNA polymerase (RdRp)[105–107]. Following entry of the virion into the host cell and release of the viral genome, the open reading frames ORF1a and ORF1b, which encode 16 non-structural proteins (nsp1–nsp16), are translated into the shorter polyprotein pp1a and, via a ribosomal frameshift, into the longer polyprotein pp1ab. PL^{pro} and M^{pro} proteolytically process pp1a and pp1ab at three and eleven cleavage sites, respectively, yielding nsp1–11 from pp1a, as well as nsp1–10 and nsp12–16 from pp1ab (Fig. 1.4A). Since PL^{pro} and M^{pro} are encoded within nsp3 and nsp5, respectively, they undergo autoproteolytic cleavage to liberate themselves. Nsp2–16 assemble into the viral replication and transcription complex (RTC), which is directed to defined subcellular sites and exploits host cell membranes and resources for viral RNA synthesis. While nsp12 (contains RdRp domain), nsp13, nsp14 and nsp16 comprise the core enzymatic machinery for replication, proofreading, and RNA modification, nsp2–11 provide supporting functions such as membrane remodeling, immune evasion, and cofactor activity[94,108].

With the COVID-19 outbreak, numerous drug-repurposing studies were launched in

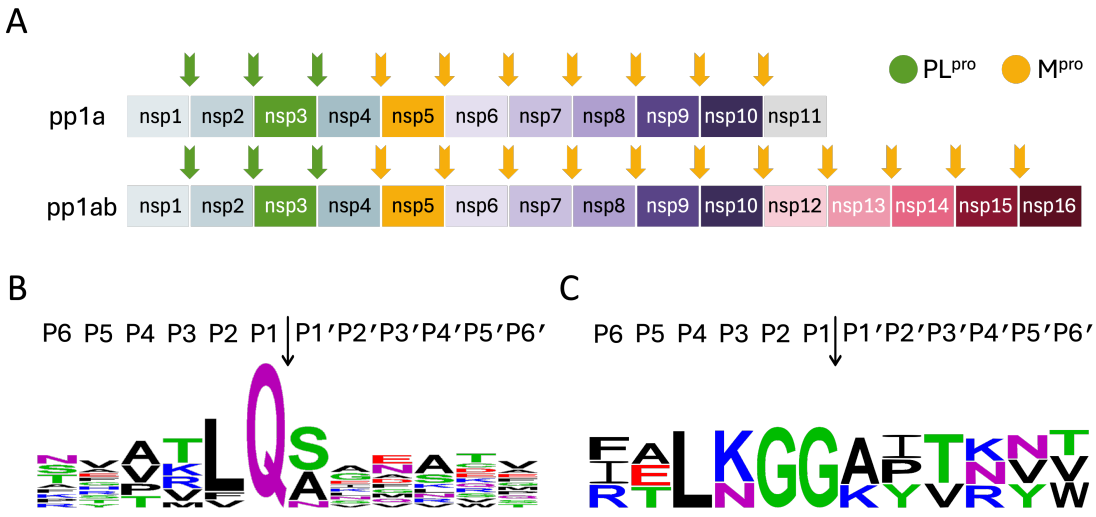


Figure 1.4.: Cleavage sites of SARS-CoV-2 M^{pro} and PL^{pro} within the viral polypeptides pp1a and pp1ab. The polypeptides pp1a and pp1ab are cleaved by PL^{pro} and M^{pro} at three and eleven sites, respectively. (B) SARS-CoV-2 M^{pro} and (C) SARS-CoV-2 PL^{pro} cleavage site sequence logo plotted with WebLogo Version 3.7.9[109]. The data was obtained from the Uniprot database (accession code: P0DTD1) by extracting the eleven and three cleavage sites, respectively, from the polyprotein 1ab ranging from P6–P6′ position. The height of symbols indicates the frequency of each amino acid at a given substrate position.

hopes of rapidly identifying effective COVID-19 therapies among already approved drugs. Remdesivir, which had previously demonstrated antiviral activity against RNA viruses such as Ebola, was among the first treatments for COVID-19 to receive emergency use authorization from the Food and Drug Administration (FDA) in May 2020. As an adenosine triphosphate (ATP) analogue, it competes with natural ATP substrates for incorporation into nascent viral RNA chains, resulting in delayed chain termination[110]. In December 2021, two further oral antiviral therapies, Paxlovid and Molnupiravir, have been authorized by the FDA to reduce the risk of hospitalization or death in patients with mild-to-moderate COVID-19. Paxlovid combines Nirmatrelvir, a potent inhibitor of SARS-CoV-2 M^{pro} developed by Pfizer[111], and Ritonavir, which is used to enhances the effectiveness and bioavailability of Nirmatrelvir by slowing hepatic metabolism[112]. Similar to Remdesivir, Molnupiravir targets the viral RdRp; however, instead of causing chain termination, it functions as a nucleoside analogue that introduces mutagenic base substitutions, resulting in an 'error catastrophe'[113,114].
Although currently available COVID-19 antivirals have markedly reduced the risk of hospitalization and death[115,116], there remains a critical need for inexpensive and easily administered therapies to address emerging drug resistance and to ensure preparedness against future pathogenic coronavirus outbreaks.

1.3.1.1. The Main Protease

Due to its essential role in viral replication and its unique substrate recognition motif, characterized by a single conserved glutamine, M^{pro} is considered one of the most promising drug targets against SARS-CoV-2. M^{pro} catalyzes proteolytic processing of the viral polyproteins pp1a and pp1ab at a total of eleven cleavage sites, which can be collectively represented as a sequence logo summarizing the cleavage site consensus, as illustrated in Fig. 1.4B. According to the enzyme-substrate interaction nomenclature introduced by Schechter and Berger[117], substrate recognition by M^{pro} is mainly driven by the amino acid residues at positions P4–P1′. In addition to the conserved P1 residue, the adjacent positions P2 and P1′ contribute mostly to substrate specificity. P2 usually contains hydrophobic residues (Leu/Phe/Val), whereas P1′ prefers small residues (Ser/Ala/Asn) [106,118].
M^{pro} is highly conserved across coronaviruses. In the case of HCoVs, sequence identities range between approximately 40 % and 96 %—with the highest identity observed between SARS-CoV-2 and SARS-CoV M^{pro}—while the structural fold of the protease is highly conserved[119,120], especially in the active site region (RMSD $\leq$ 1.02 Å))[121].
M^{pro} can be structurally divided into three distinct domains (Fig. 1.5A). The N-terminal domains I (residues 8–101) and II (residues 102–184) exhibit a β-sheet-rich chymotrypsin-like fold, with the substrate-binding site located at the interface between the two domains.

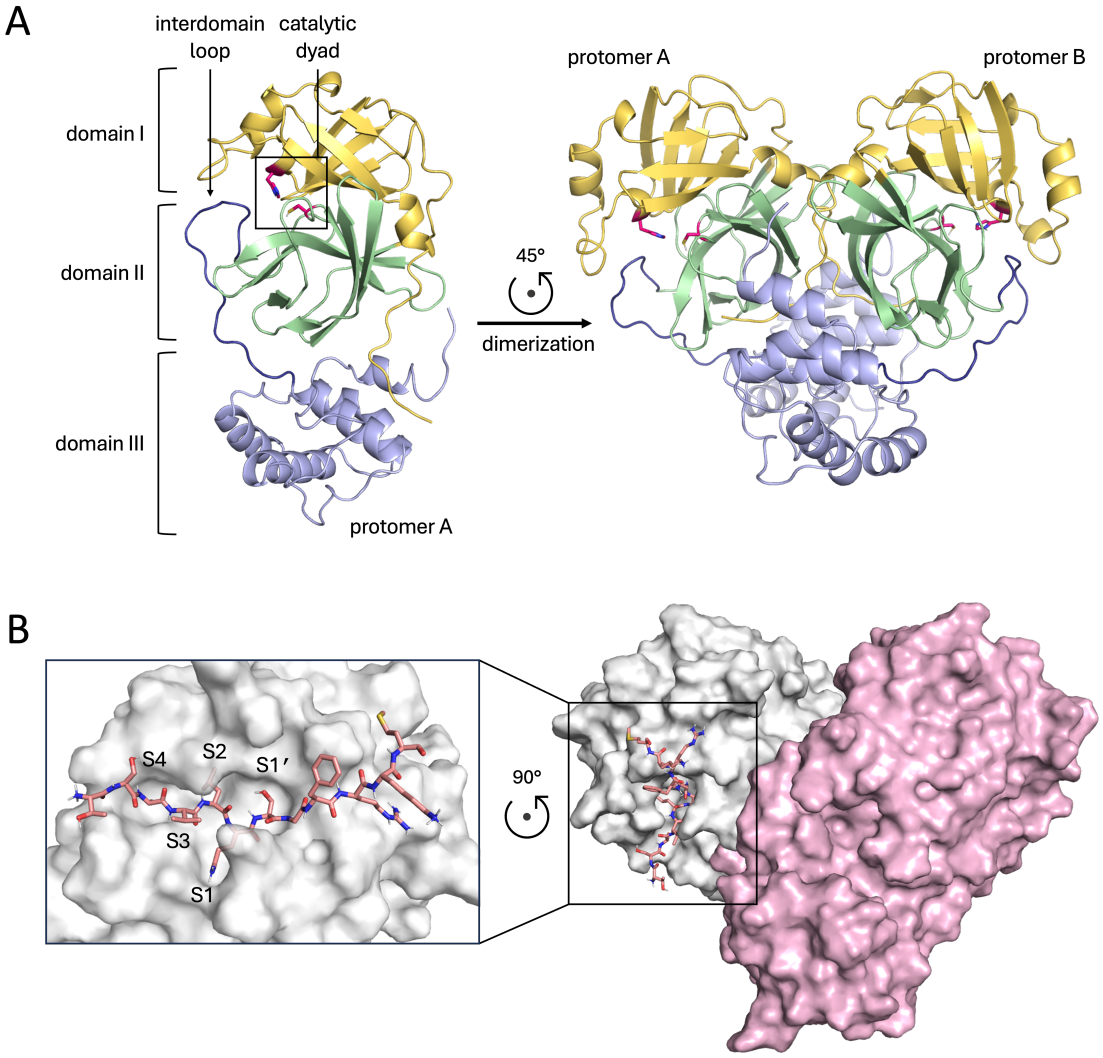


Figure 1.5.: Structural features of SARS-CoV-2 M^{pro}, including domain organization, dimerization, and substrate recognition. (A) Cartoon representation of M^{pro} (PDB ID: 7T70), showing the three domains, the interdomain loop connecting domains II and III, and the catalytic dyad (His41 and C145) between domains I and II. M^{pro} is active only as a homodimer, with the dimerization guided through domain III interactions between two protomers. (B) Surface representation of M^{pro} with substrate nsp4/5 bound at the active site (PDB ID: 7T70). Subsites S4–S1′, which are primarily responsible for substrate recognition, are indicated.

The C-terminal domain III (residues 201–303), which is connected to domain II via a long interdomain loop (residues 185–200), is primarily composed of α-helices and plays a crucial role in the dimerization of M^{pro} into its catalytically active homodimeric form[120,122,123].

Proteolytic cleavage of polypeptide substrates by M^{pro} is mediated through the catalytic dyad, comprising His41 in domain I and Cys145 in domain II. Analogous to the substrate positions P4–P1′, the most structurally well-defined subsites within the substrate-binding

region correspond to S4–S1′. These subsites are adapted to the residue-specific preferences of the substrate, with the shallow S1′ subsite that can only accommodate residues with short side chains at P1′, and the highly hydrophobic S2 favoring hydrophobic residues at P2 (Fig. 1.5B)[124,125].

Numerous drug repurposing virtual screening studies targeting M^{pro}, sometimes using libraries of millions to billions of compounds, have been reported[126–133]. However, many of these studies employed simplified docking-only approaches without experimental validation and showed limited consistency across results[134]. Nonetheless, owing to the extensive global efforts, the COVID Moonshot project—an open-science drug discovery campaign to identify, synthesize, and test inhibitors against SARS-CoV-2 M^{pro}—has identified over 2,000 compounds, mainly small molecules, with confirmed biological activity[135,136].

To date, the only FDA-approved M^{pro} inhibitor is nirmatrelvir (Paxlovid), a covalent peptidomimetic inhibitor developed by Pfizer through rational drug design[111]. While nirmatrelvir shows sufficient efficacy, its poor pharmacokinetic properties require coadministration with ritonavir. Moreover, the clinical benefit of Paxlovid strongly depends on treatment timing: delayed administration reduces its impact on peak viral shedding, whereas very early treatment increases the likelihood of viral rebound without lowering infectiousness[137]. Thus, the development of inhibitors with enhanced efficacy and pharmacological properties remains an urgent challenge in antiviral drug design, particularly in preparing for future COVID-19 variants.

1.3.1.2. The Papain-like Protease

Another promising target for SARS-CoV-2 drug development is PL^{pro}, which represents one of eight domains within the membrane-anchored, multi-domain protein nsp3. It cleaves the polyprotein pp1a at three sites, between nsp1/nsp2, nsp2/nsp3 and nsp3/nsp4, releasing proteins that are essential for viral replication[107,138]. The three cleavage sites are summarized as a sequence logo in Fig. 1.4C.

In addition to its role in viral polyprotein processing, PL^{pro} functions as a deubiquitylating and de-ISGylating enzyme, preferentially hydrolyzing interferon-induced gene 15 (ISG15) and K48-linked polyubiquitin chains from proteins. By reversing these post-translational modifications, PL^{pro} suppresses specific antiviral signaling pathways, thereby promoting immune evasion and enhancing viral replication[139–141]. Substrate recognition by PL^{pro} for proteolytic cleavage, de-ubiquitination and de-ISGylation is governed by the LXGG motif, corresponding to the P4–P1 positions of the substrate[107,140,142].

Similar to M^{pro}, PL^{pro} is conserved across coronaviruses. While SARS-CoV-2 PL^{pro} shares 83 % sequence identity with that of SARS-CoV—with the highest conservation

observed within the active site—and only 29 % identity with MERS-CoV, their overall structures are highly conserved (RMSD $\leq$ 1.11 Å)[107,143]. The secondary structure of SARS-CoV-2 PL^{pro} can be divided into four subdomains: the N-terminal Ubl domain (residues 1–61), the α-helical thumb domain (residues 62–178), the β-stranded finger domain (residues 179–240), and the palm domain (residues 241–315), comprising six β-strands (Fig. 1.6A). The catalytic triad, comprising Cys111, His272, and Asp286, is located at the interface between the thumb and palm domains. Within the finger domain,

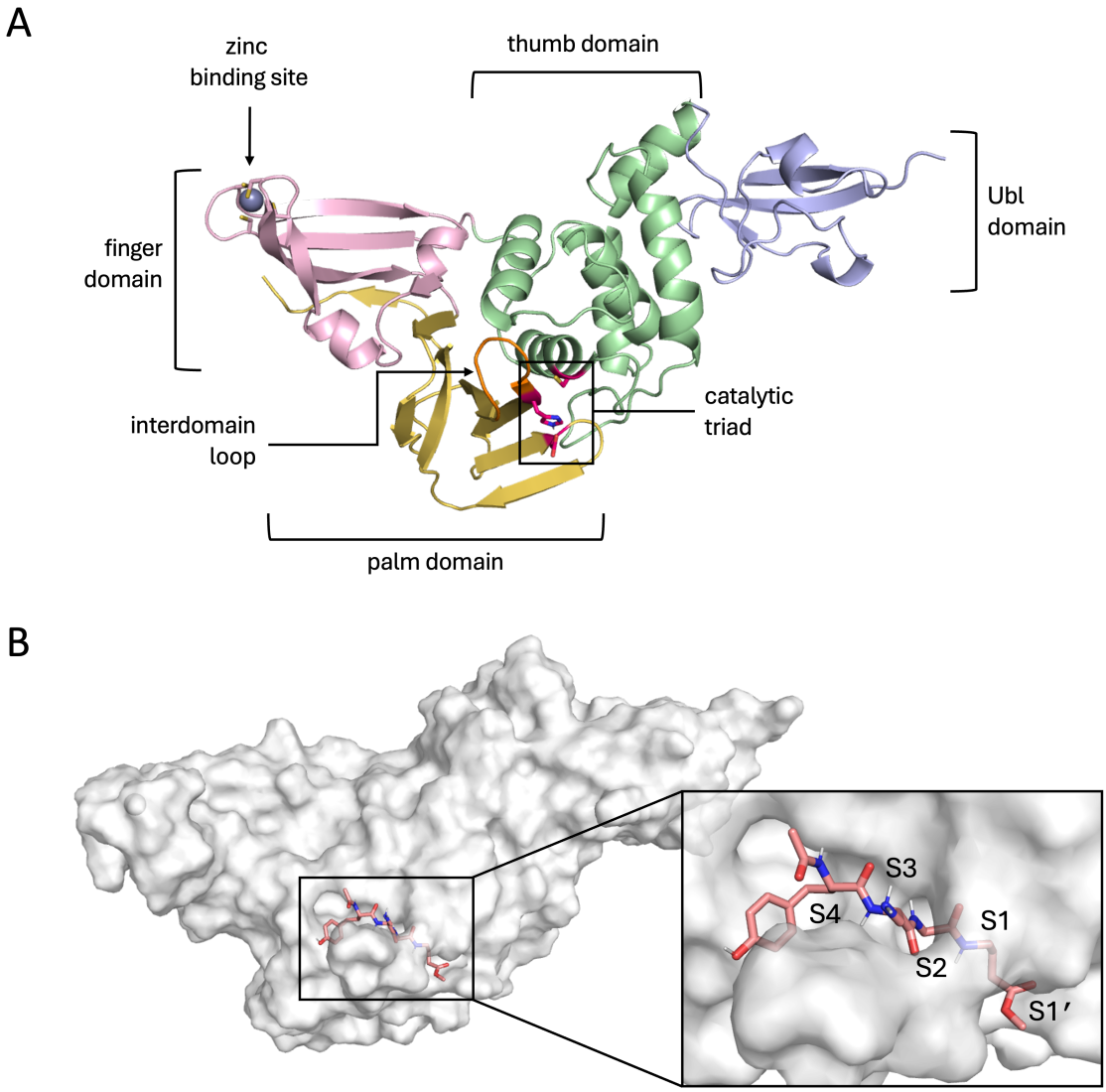


Figure 1.6.: Structural features of SARS-CoV-2 PL^{pro}. (A) Cartoon representation of the four domains—ubiquitin-like (Ubl), thumb, palm, and fingers—of PL^{pro} (PDB ID: 6WX4). The catalytic triad (Cys111, His272, Asp286), the Zn-binding site (coordinated by Cys189, Cys192, Cys224, and Cys226), and the BL2 loop are indicated. (B) Surface representation of PL^{pro} in complex with the covalent peptide inhibitor VIR251 (PDB ID: 6WX4). The subsites S4–S1′ are indicated, among these, S4–S1 mediate recognition of the LXGG motif.

four conserved cysteine residues (Cys189, Cys192, Cys224, and Cys226) coordinate a Zn ion[107,143], which is essential for catalytic activity[144]. A flexible β-turn, the BL2 loop (residues Gly266–Gly271), is located near the active site and plays a key role in substrate binding. By closing over the active site, it stabilizes the protease–substrate complex[138,143]. Recognition of the LXGG motif is mediated by subsites S4–S1, as illustrated in Fig. 1.6B with the peptidomimetic inhibitor VIR251 bound to PL^{pro}.

Beyond M^{pro}, PL^{pro} has also drawn considerable attention in high-throughput virtual screening campaigns[133,145–148]. Among the inhibitors identified, GRL0617—a noncovalent naphthalene compound—and its analogs represent some of the earliest and most thoroughly characterized PL^{pro} inhibitors[149,150]. However, to date, there are no FDA-approved inhibitors targeting PL^{pro}, nor are there any PL^{pro} inhibitors currently in clinical trials.

1.3.2. Prostate Cancer

Prostate cancer is the second most common cancer and the fifth leading cause of cancer deaths in men globally, with over 1.46 million new cases and 396,000 deaths in 2022[151]. While the highest incidence rates are observed in predominantly high-income regions, including Australia, the United States, and Northern Europe, the highest mortality rates occur in sub-Saharan Africa and the Caribbean, mainly due to limited access to early detection methods and effective treatment. As the global population continues to grow and age, the number of prostate cancer cases and related deaths is expected to rise significantly in the coming decades[152,153].

In patients with localized prostate cancer, i.e. the cancer is confined to the prostate gland, therapeutic approaches including active surveillance, radical prostatectomy, and radiotherapy are associated with excellent long-term prostate cancer-specific survival outcomes.[154,155]. However, once prostate cancer has metastasized—most commonly to bones and lymph nodes—the treatment becomes more complex. In advanced and metastatic prostate cancer, systemic therapies such as androgen deprivation therapy (ADT) or chemotherapy are typically employed. As the current mainstay of treatment, ADT controls disease progression by suppressing androgen receptor signaling, thereby depriving prostate cancer cells of critical growth and survival stimuli. Although most patients initially respond to ADT, therapeutic resistance usually develops within 2–3 years, resulting in disease progression despite continuous hormonal suppression, a stage referred to as castration-resistant prostate cancer (CRPC)[156,157]. At this point, chemotherapy becomes the main treatment; however, improvements in survival are modest relative to the risk of significant toxicity, and resistance to chemotherapy inevitably develops[157,158]. Overall, patients with metastatic CRPC have a median survival of approximately 16 months[159].

Thus, accurate and sensitive detection of metastatic lesions is essential for optimizing targeted treatment strategies[160].

Over the past decades, the diagnostic pathway for prostate cancer has evolved into a multimodal approach. Initial assessment typically involves a combination of prostate-specific antigen (PSA) testing and digital rectal examination. This is commonly followed by multiparametric magnetic resonance imaging (mpMRI), which, in accordance with European Association of Urology (EAU) guidelines, is used to improve lesion localization and guide targeted prostate tissue biopsy. To determine disease stage and, accordingly, select the optimal treatment strategy, distinguishing localized from advanced prostate cancer is critical. While conventional imaging modalities such as computed tomography (CT) and bone scintigraphy remain widely used, their sensitivity for detecting small or early metastatic lesions is limited[161–163]. In contrast, PSMA-targeted positron emission tomography (PET) offers a more accurate alternative. PSMA, a transmembrane protein commonly overexpressed in prostate cancer cells, provides an excellent molecular target for imaging with radiolabeled ligands (Fig. 1.7)[164]. PSMA–PET is most often performed in combination with CT in a hybrid PET/CT system, combining the high molecular specificity of PSMA-targeted detection with the precise anatomical imaging by CT[165,166]. This enables accurate detection of lymph node and bone metastases, with clinical trials confirming superior sensitivity and specificity over conventional imaging[167–169]. In ad-

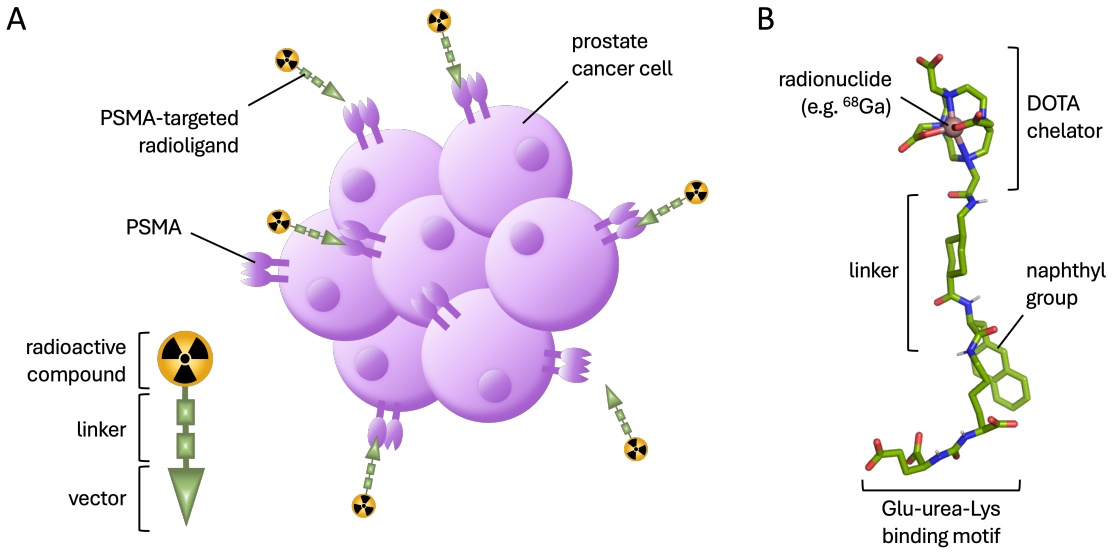


Figure 1.7.: PSMA-targeted radioligand therapy and diagnostic. (A) Radioligands targeting and binding specifically to PSMA on the surface of prostate cancer cells, enabling either diagnostic imaging or cancer therapy through radioactivity. (B) The structure of PSMA-targeted radioligands is exemplified by the theranostic ligand PSMA-617: the DOTA chelator, which allows the incorporation of a radiometal (e.g., ^{68}Ga for imaging or ^{177}Lu for therapy), is connected via a naphthyl- and cyclohexyl-containing linker to the PSMA-targeting vector, a Glu-urea-Lys binding motif.

dition to PET, single-photon emission computed tomography (SPECT) can be employed for PSMA-targeted imaging, using radiopharmaceuticals labeled with photon-emitting radionuclides such as ^{99m}Tc. By contrast, PET utilizes positron-emitting radionuclides, for example ^{68}Ga or ^{18}F, conjugated to PSMA ligands. While PET provides superior spatial resolution and higher sensitivity, SPECT tracers are more widely available and cost-effective, making SPECT a practical alternative when PET access is limited[170–172]. By improving the detection and localization of small metastatic lesions, PSMA-targeted imaging has facilitated the implementation of metastasis-directed therapies (MDT) approaches such as stereotactic body radiotherapy, a high-precision technique in which large doses of radiation are carefully delivered to metastases over only a few sessions, and which has been shown to be both effective and well-tolerated. MDT can postpone the initiation of systemic therapy, thereby reducing exposure to the side effects of hormonal therapy, and prolong progression-free survival in patients with oligometastatic prostate cancer[173–175]. MDT is primarily employed in cases of oligometastatic or oligorecurrent disease, where the number of metastatic lesions is limited and approachable for local treatment. However, in patients with metastatic CRPC involving multiple tumor sites throughout the body, PSMA-targeted radioligand therapy (RLT) has emerged as an effective therapeutic option. RLT employs targeted radionuclides, such as the β-emitter ^{177}Lu or the α-emitter ^{223}Ra, to deliver cytotoxic radiation directly to cancer cells and their microenvironment, while largely sparing healthy tissue in contrast to external beam radiotherapy[176,177].

1.3.2.1. The Prostate Specific Membrane Antigen

PSMA is a type II transmembrane glycoprotein encoded by the *FOLH1* gene, exhibiting dual enzymatic activities as a folate hydrolase and a carboxypeptidase. It is expressed in multiple tissues, including the prostate, kidney, small intestine, and central nervous system. Depending on its tissue localization and function, it is also referred to as glutamate carboxypeptidase II (GCPII)—due to its ability to hydrolyze poly-glutamated folates and thereby release glutamate—or as N-acetyl-L-aspartyl-L-glutamate peptidase I (NAALADase I), referring to its role in metabolizing the neuropeptide N-acetyl-aspartyl-glutamate (NAAG) in the brain.

From a clinical perspective, PSMA has gained attention due to its increased expression in prostate cancer cells compared to benign prostate tissue, which increases with disease progression, making it an attractive target for both diagnostic and therapeutic applications[178,179]. The role of PSMA in the metabolism of folates has been associated with tumor growth and progression, however its exact function in prostate cancer remains unclear[180].

The native conformation of PSMA is a homodimer (Fig. 1.8A), and homodimerization

is essential for its enzymatic activity[181]. Each protomer is organized into three distinct structural regions: a short N-terminal intracellular segment (residues 1–19), a single-pass transmembrane helix (residues 20–43), and an extensive C-terminal extracellular domain (residues 44–750). The extracellular domain can be further subdivided into a protease domain (residues 56–116 and 352–591), an apical domain (residues 117–351), and a helical C-terminal domain (residues 592–750)[178,182].

The substrate-binding site is located at the interface of these three domains and consists of a large cavity containing a deeply buried catalytic site defined by two Zn^{2+} ions. Each Zn ion is coordinated by three residues: a histidine (His553 or His377), an aspartate (Asp453) or glutamate (Glu425), and a bridging aspartate (Asp387). Both Zn ions are additionally coordinated by a bridging water molecule, which serves as a nucleophile in the cleavage of the peptide bond in a bound substrate. Within the substrate-binding cavity, the arginine patch at the S1 site is formed by three arginines (Arg463, Arg534, and Arg536), which interact with the residue immediately N-terminal to the scissile bond, such as the aspartate in NAAG. A fourth arginine (Arg210) is positioned on the opposite side of the cavity at the S1′ glutamate pocket, which accommodates the negatively charged C-terminal glutamate residue of the substrate[183,184].

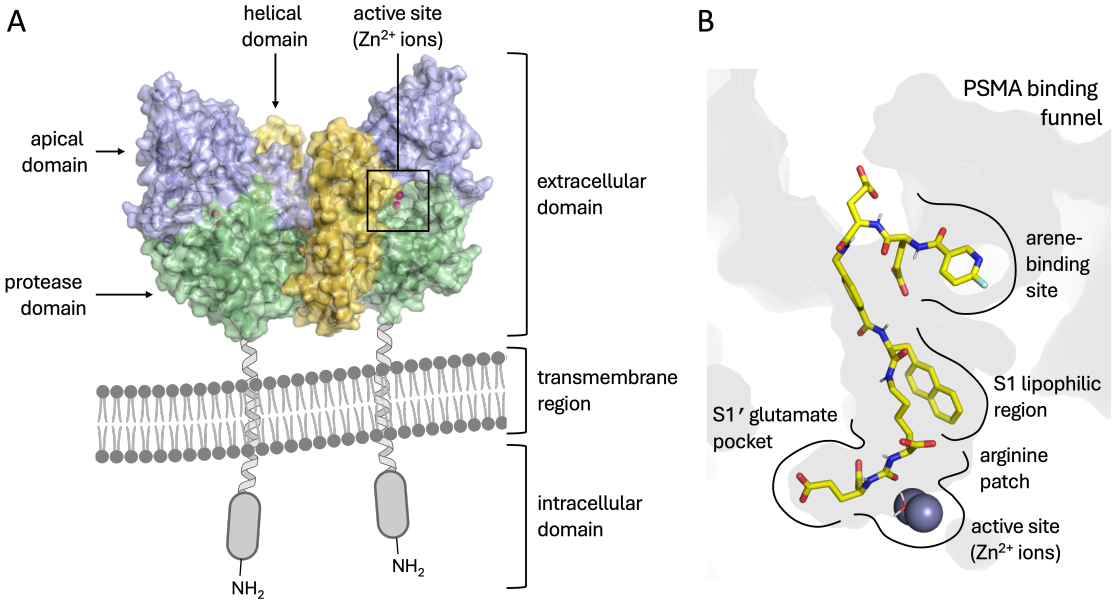


Figure 1.8.: Schematic representation of the homodimeric PSMA transmembrane protein and its binding cavity. (A) The extracellular domain of the PSMA homodimer, including its domains and active site, is shown based on the biological assembly of the PDB structure 5o5t. The transmembrane and intracellular regions are added schematically to illustrate the complete transmembrane protein. (B) Binding funnel of PSMA co-crystallized with the inhibitor PSMA-1007, highlighting key protein–inhibitor interaction subsites (PDB ID: 5o5t).

Additional subsites of the PSMA binding funnel, including the S1 accessory lipophilic region[185] and a remote arene-binding site[186], have been characterized. Both subsites are engaged by the inhibitor PSMA-1007: its ligand's naphthyl group occupies the S1 lipophilic region, while its fluoronicotinic acid moiety is accommodated within the arene-binding site (Fig. 1.8B)[187–189].

1.3.2.2. PSMA Radioligands

In general, PSMA radiotracers consist of three components: a vector (or binding motif), a radionuclide—coordinated by a chelator or prosthetic group—and a linker inbetween (Fig. 1.7B). PSMA targeting vectors typically contain a Glu-ureido–based moiety that exhibits high affinity towards the binding site[190,191]. The chelator determines which radionuclides can be stably incorporated, enabling either diagnostic imaging (e.g. ^{111}In and ^{99m}Tc for SPECT or ^{68}Ga and ^{18}F for PET) or targeted RLT (e.g. ^{177}Lu and ^{90}Y for β-therapy or ^{225}Ac for α-therapy). The linker connects the chelator and vector, and can also include modifications that enhance binding affinity by engaging specific PSMA binding pockets[187,192].

[^{18}F]FDG and capromab pendetide (ProstaScint) were among the first molecular imaging agents approved for prostate cancer imaging. [^{18}F]FDG enables the detection of many cancers through uptake by tumor cells with high metabolic activity. However, its utility in prostate cancer is limited due to the disease's low glycolytic rate[193]. Capromab pendetide is an ^{111}In-labeled antibody targeting PSMA that binds the intracellular domain. As it cannot be internalized by viable cells, detection is largely limited to necrotic tumor regions, with markedly reduced sensitivity for bone and lymph node metastases[194].

The first small-molecule PSMA inhibitors for prostate cancer imaging, [^{123}I]MIP-1072 and [^{123}]MIP-1095, entered clinical trials in 2008. They are based on a Glu-urea-Lys binding motif and incorporate aromatic moieties that enable stable labeling with ^{123}I, a single-photon emitter suitable for both imaging and RLT[195–197].

A selection of clinically relevant Glu-ureido-based PSMA radioligands for PET/SPECT diagnostics and RLT are summarized in Fig. 1.9. The first widely used PET tracer, introduced in 2012, was [^{68}Ga]Ga-PSMA-11[198,199], a Glu-ureido–based peptidomimetic PSMA inhibitor labeled with ^{68}Ga via the chelator N,N′-Bis[2-hydroxy-5-(ethylene-β-carboxy)benzyl]ethylenediamine-N,N′-diacetic acid (HBED-CC). However, HBED-CC cannot complex clinically relevant therapeutic radiometals such as ^{177}Lu, ^{90}Y or ^{225}Ac.

To address this limitation, PSMA-617 was developed and introduced in 2015[200,201]. It retains the Glu-urea-Lys binding motif while incorporating a 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid (DOTA) chelator, which enables labeling with both diagnostic and therapeutic radionuclides, and a naphthyl linker group to optimize binding. Several

clinical studies of [^{177}Lu]Lu-PSMA-617 in metastatic CRPC indicated improved survival outcomes and higher response rates compared to standard treatments, with fewer severe adverse events[202–205]. Based on the phase 3 VISION trial (NCT03511664), demonstrating the efficacy and safety of [^{177}Lu]Lu-PSMA-617[206], the FDA granted approval of Pluvicto (lutetium Lu 177 vipivotide tetraxetan) for the treatment of adult patients

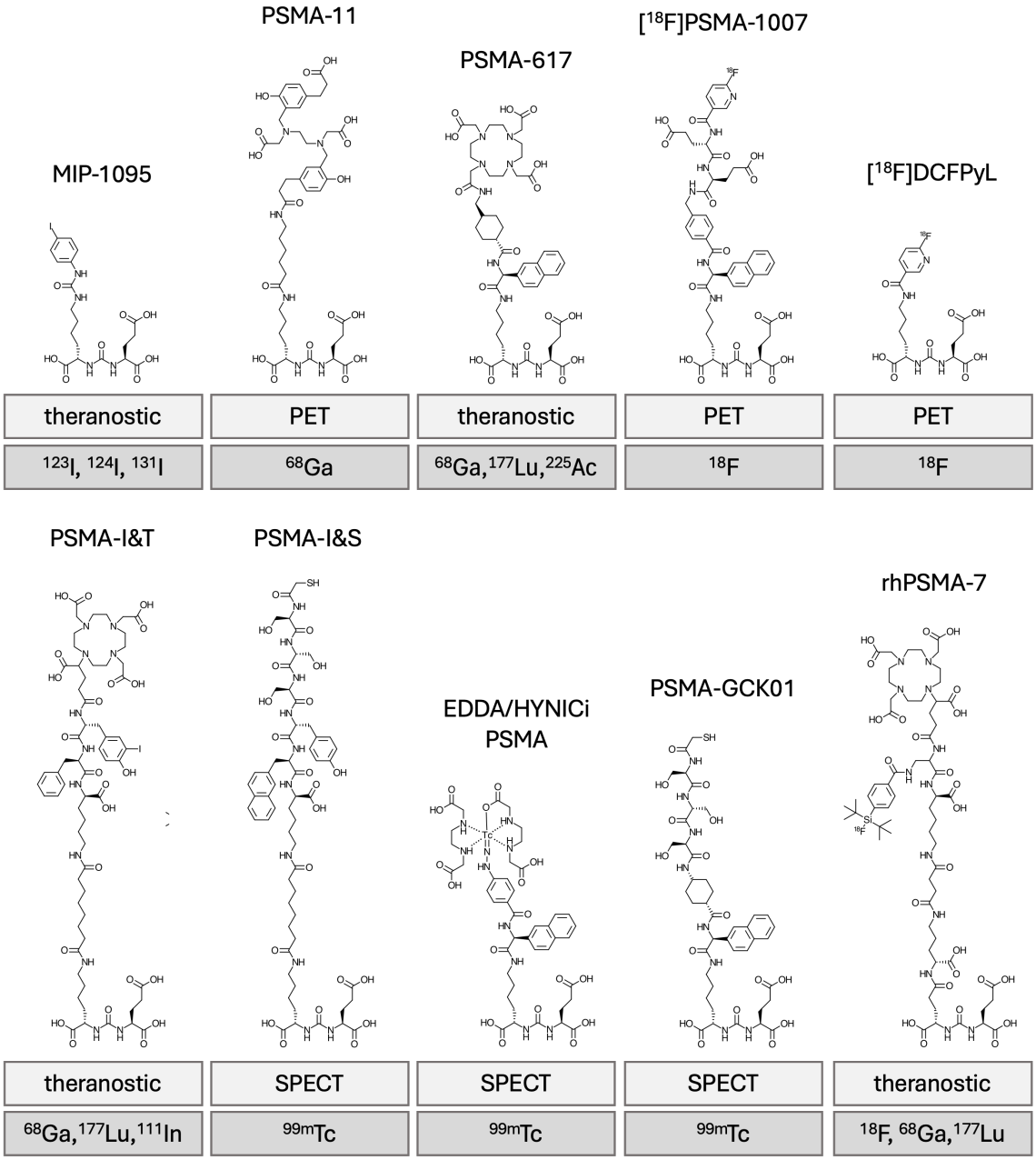


Figure 1.9.: Clinically relevant Glu-ureido-based PSMA radioligands. Chemical structures of clinically relevant PSMA-targeted radioligands. Below each structure, their classification—diagnostic (PET or SPECT), therapeutic, or theranostic—and radionuclides used for labeling are indicated.

with PSMA-positive metastatic CRPC[207]. PSMA-617 has been shown to exhibit significantly faster renal clearance than PSMA-11, which is important for minimizing potential radiation-induced kidney toxicity and reducing the risk of long-term side effects[200].
While [^{68}Ga] Ga-PSMA-11 and ^{177}Lu or ^{225}Ac-labeled PSMA-617 were able to cover the diagnostic and therapeutic aspects of clinical prostate cancer care, the limited availability and short half-life of ^{68}Ga restrict large-scale production and routine use. This created a strong demand for ^{18}F-labeled PSMA ligands, leading to the development of [^{18}F]PSMA-1007. PSMA-1007 was derived from PSMA-617, sharing the Glu-ureido-based binding motif and a similar linker structure featuring a naphthyl group. The key difference lies in their radiolabel: PSMA-617 is labeled via its DOTA chelator (e.g., with ^{68}Ga), whereas PSMA-1007 is labeled with a covalently bound ^{18}F at the 6-fluoronicotinic acid (FNA) group, which engages the arene-binding site of PSMA. To counterbalance the loss of hydrophilicity caused by replacing DOTA with FNA, two glutamic acid residues were added to the linker of PSMA-1007[187,189,208].
Clinical studies have demonstrated that [^{18}F]PSMA-1007 performs comparably to ^{68}Ga-PSMA-11 and other ^{18}F-labeled PET tracers, such as [^{18}F]DCFPyL, which was introduced in 2011, and is based on a ^{18}F-fluoropyridyl-substituted Glu-urea-Lys motif[209,210]. However, [^{18}F]PSMA-1007 offers distinct advantages, including its longer half-life, which enables large-scale production, reducing the need for multiple daily syntheses, and its low urinary excretion together with high tumor-to-background ratios allowing highly sensitive detection of even small lesions in the body[211,212].
Another theranostic PSMA radioligand, PSMA-I&T (imaging and therapy), incorporates a 1,4,7,10-tetraazacyclododecane,1-(glutaric acid)-4,7,10-triacetic acid (DOTAGA) chelator, a Glu-urea-Lys binding motif, and a D-amino acid-substituted peptidyl spacer[213,214]. In early clinical studies, [^{177}Lu]Lu-PSMA-I&T achieved promising outcomes in heavily pretreated patients with metastatic CRPC, with 68 % demonstrating stable disease or complete remission[215]. In addition to PSMA-targeted therapy, the ^{111}In-labeled PSMA-I&T has been used for radioguided surgery (RGS), enabling intraoperative detection and resection of even small or atypically PSMA-positive lymph node metastases[216]. However, its suboptimal nuclear properties, high cost, and limited availability restrict clinical use. To develop a more practical alternative, [^{111}In]In-PSMA-I&T was adapted to ^{99m}Tc-labeling by replacing the DOTAGA chelator with mercaptoacetyl tri-D-serine (MAS_3), yielding PSMA-I&S (imaging and surgery). This SPECT radioligand has shown promising initial clinical results, with early data indicating comparable or even superior performance for PSMA-targeted RGS[217].
[^{99m}Tc]Tc-EDDA/HYNIC-iPSMA, a SPECT radioligand developed primarily for diagnostics, proved effective in clinical studies comparing it with the performance of

[^{68}Ga]Ga-PSMA-11[218] or [^{68}Ga]Ga-PSMA-617[219]. However, since HYNIC is unsuitable for coordination with ^{188}Re—a therapeutic nuclide that can serve as a theranostic tandem with ^{99m}Tc—it was replaced by the ^{188}Re-compatible MAS_3 chelator, yielding PSMA-GCK01[220]. The theranostic tracer [^{99m}Tc]Tc/[^{188}Re]Re-PSMA-GCK01 showed similar imaging characteristics compared to [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA in a clinical study of patients with metastatic CRPC[221].
A novel class of PSMA-targeted radiotracers are radiohybrid (rh) ligands, which integrate a silicon–fluoride–acceptor (SiFA) moiety and a chelator within a single molecule. These ligands can be radiolabeled either with ^{18}F via isotopic exchange at the SiFA site or with radiometals such as ^{68}Ga or ^{177}Lu at the chelator[222]. An example is [^{18}F,^{nat}Ga]rhPSMA-7, which has shown high detection rates in PET imaging of primary and metastatic CRPC[223].
Over the past two decades, a diverse array of clinically relevant PSMA-targeting radioligands has been developed. However, important challenges such as high renal uptake and unresolved pharmacological questions remain to be addressed.

1.4. Aim of the Thesis

Therapeutic peptides are emerging as a promising class of drugs, distinguished by their versatility, target specificity, and adjustable pharmacokinetics. Despite rapid advancements in computational techniques, including AI models like AlphaFold, the intrinsic conformational flexibility of peptides still poses a challenge for the prediction of peptide–protein complex structures. As reliable structural information on pepPIs is essential for structure-based peptide design, this thesis aims to develop optimized modeling workflows that serve as basis for subsequent computer-aided design of peptide-based therapeutics. By exploring different strategies adapted to the intrinsic flexibility of peptides, we further aim to establish computational techniques for the accurate design of peptide inhibitors.
To achieve this, we model natural substrates and Dri-peptides in complex with the critical SARS-CoV-2 proteases M^{pro} and PL^{pro}. We employ docking and MD simulations to study peptide–protein interactions, forming the basis for various design strategies, which are validated through *in vitro* inhibition assays of the predicted peptide designs.
To advance PSMA-targeted radiopharmaceutical design for prostate cancer diagnostics and therapy, we aim to develop an optimized computational workflow for the accurate prediction of PSMA–radioligand interactions. By performing docking under varying conditions and employing a re-scoring process, we generate promising starting structures for subsequent MD simulations. These simulations aim to elucidate the structure–activity relationships governing PSMA–radioligand binding, with the resulting insights establishing a foundation for future computer-aided radiopharmaceutical design (CARD) studies.

2. Methods

In the following, the methods and algorithms employed in this work are outlined. The workflow begins with molecular docking followed by a rescoring approach to reevaluate the docking solutions. Starting from the docked poses, molecular dynamics (MD) simulations are performed, with binding affinities assessed using free energy calculations. Next, the computational dynamic saturation mutagenesis (CDSM) protocol developed in this study is presented. Finally, experimental methods for evaluating the inhibitory activity of the peptide designs predicted through the computer-aided peptide inhibitor design (CAPID) workflow are described.

2.1. Modeling Peptide–Protein Interactions

To predict and analyze peptide–protein interactions (pepPIs), we applied a combination of molecular docking and MD simulations. Docking was used to generate initial peptide binding poses within the receptor binding site, while MD simulations were performed to explore the conformational space of the peptide–protein complex and to refine the predicted binding modes. In the following, we describe the theoretical background and specific methodologies used for modeling pepPIs through docking and MD simulations.

2.1.1. Protein Preparation

All protein structures obtained from the Protein Data Bank (PDB), including the X-ray crystal structures of the main protease (M^{pro}) (PDB: 6M2N, 7T70), papain-like protease (PL^{pro}) (PDB: 6WX4) and PSMA (PDB: 8BOW, 5O5T), were processed using the protein preparation wizard in Maestro (Schrödinger)[224]. Default settings (pH 7.4 and OPLS_2005 as force field) were applied for adding hydrogens to protein, ligand and water molecules, reconstructing missing side chains and loops, and optimizing hydrogen bonds and water orientations.

2.1.2. Molecular Docking

Docking is widely used to predict the predominant binding modes of a ligand in protein–ligand complex structures. In this study, we employed four different docking tools: AutoDock Vina (Vina)[32], AutoDockFR (ADFR)[225], AutoDock CrankPep (ADCP)[39], and AlphaFold3[22]. Vina is a fast docking engine—suitable especially to screen large databases—which combines knowledge-based and empirical scoring functions, whereas ADFR and ADCP, which are both part of the ADFR software suite, use the AutoDock4 (AD4)[226] scoring function that is based on a semi-empirical free energy force field. ADFR extends the docking capabilities of AD4 and Vina by introducing selective receptor flexibility, support for covalent docking, and a custom Genetic Algorithm optimized for handling ligands with a large number of rotatable bonds. ADCP is specifically designed for docking flexible peptides to protein receptors. It combines CRANKITE[43]—a software that can efficiently sample the conformational space of peptides and proteins—with AutoDock's grid-based representation of a rigid receptor, enabling the simultaneous optimization of peptide conformations and their interactions with the receptor. AlphaFold 3 (AF3), developed by DeepMind, enables structure prediction of various complexes, including proteins, nucleic acids, small molecules, ions and modified residues, through a machine learning (ML)-based framework.

2.1.2.1. Scoring Functions

Scoring functions in molecular docking estimate the binding affinity between a ligand and its receptor, allowing the predicted binding poses to be ranked accordingly. They can be classified into four major classes: (i) physics-based (or force field-based), (ii) empirical (or regression-based), (iii) knowledge-based (or potential of mean force-based), and (iv) machine learning-based (or descriptor-based)[227]. However, the boundaries between these categories are often indistinct, as many approaches integrate elements from multiple strategies[228].

Force Field-based. The AD4 scoring function—shared by ADFR and ADCP—is based on a semi-empirical force field that estimates the binding free energy (ΔG_{bind}) of a protein–ligand interaction in two steps: (i) the intramolecular energy required for both the ligand (L–L) and the protein receptor (P–P) to adopt their bound conformations, and (ii) the change in intermolecular energy upon complex formation (P–L), which also accounts for the conformational entropy loss (ΔS_{conf}):

$$\Delta G_{\text{bind}} = \left(U^{\text{L–L}}_{\text{bound}} - U^{\text{L–L}}_{\text{unbound}}\right) + \left(U^{\text{P–P}}_{\text{bound}} - U^{\text{P–P}}_{\text{unbound}}\right) + \left(U^{\text{P–L}}_{\text{bound}} - U^{\text{P–L}}_{\text{unbound}} + \Delta S_{\text{conf}}\right) \quad (2.1)$$

Each of the energetic terms (U) comprises estimates of dispersion/repulsion, hydrogen bonding, electrostatics, and desolvation:

$$U = W_{\text{vdw}} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + W_{\text{hbonds}} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + W_{\text{elec}} \sum_{i,j} \frac{q_i q_j}{e(r_{ij}) r_{ij}} + W_{\text{sol}} \sum_{i,j} \left(S_i V_j + S_j V_i \right) e^{\left(-r_{ij}^2 / 2\sigma^2 \right)} \tag{2.2}$$

The first term represents van der Waals (vdW) interactions by a 6/12 potential for dispersion/repulsion based on the Amber force field. Hydrogen bonds are described by a 10/12 potential and the $E(t)$ function, which provides directionality using the angle t from an ideal H-bonding geometry. The Coulomb potential represents electrostatic interactions using partial point charges q of atoms i and j, and the distance r_{ij} between them.
The last term is a desolvation potential that involves a volume of solvent shielding atoms (V), a solvation parameter (S), and an exponential distance factor (σ = 3.5 Å)[226]. The weighting constants W were optimized using experimentally determined binding constants to calibrate the empirical free energy function.
ADCP builds upon the AD4 scoring function by integrating additional conformational energy terms specifically tailored for peptides[39]. Its scoring function consists of two components: an *internal score*, which evaluates the conformational quality of the peptide, and an *interaction score*, which assesses the peptide–protein affinity based on the AD4 scoring function. Given that ADCP employs a coarse-grained peptide model, in which each side chain is represented by a single bead, the internal score is based on CRANKITE's Go-type potentials to approximate side-chain interactions within the peptide[42,43]. In addition, the internal scoring function incorporates backbone torsion biases derived from Ramachandran propensities for the ϕ and ψ dihedral angles, which are converted into energetic penalties using the Boltzmann distribution.

Knowledge/Empirical-based. Vina uses a scoring function that combines knowledge-based potentials with empirical terms to estimate the binding affinity of protein–ligand complexes[32]. Unlike force field-based methods, Vina does not explicitly compute electrostatics or solvation energies, but instead relies on statistically derived potentials and parameterization through regression against empirical data.
The binding affinity is estimated as the sum over atom-pair interactions, each defined as a function of the surface distance $d = r_{ij} - R_i - R_j$ (with the interatomic distance r_{ij} and the vdW radius R of the atoms i and j):

$$E = \sum e_{\text{pair}}(d) \tag{2.3}$$

Each pairwise term includes steric and, depending on the atom type, hydrophobic and non-directional H-bonding interactions. More specifically, the binding score is a weighted sum of two Gaussian functions $G(d)$, a repulsion term $f^{\text{rep}}(d)$, a hydrophobic term $f^{\text{h}}(d)$, and a directional hydrogen bond potential $f^{\text{hbond}}(d)$:

$$e_{\text{pair}}(d) = w_1 \cdot G_1(d) + w_2 \cdot G_2(d) + w_3 \cdot f^{\text{rep}}(d) + w_4 \cdot f^{\text{h}}(d) + w_5 \cdot f^{\text{hbond}}(d) \qquad (2.4)$$

where w_i are the respective weights. The combination of the Gaussian terms and the repulsive parabolic function approximates the shape of a Lennard-Jones (LJ) potential, capturing both attraction, moderate distances and strong repulsion at short ranges. These steric interaction terms are defined as:

$$G_1(d) = e^{-(d/0.5)^2}$$
$$G_2(d) = e^{-((d-3)/2)^2}$$
$$f^{\text{rep}}(d) = \begin{cases} d^2, & \text{if} \quad d < 0 \\ 0, & \text{if} \quad d \geq 0 \end{cases}$$

The hydrophobic and hydrogen bonding terms are modeled by:

$$f^{\text{h}}(d) = \begin{cases} 1, & \text{if} \quad d < 0.5 \\ 1.5 - d, & \text{if} \quad 0.5 \leq d \leq 1.5 \\ 0, & \text{if} \quad d \geq 1.5 \end{cases}$$
$$f^{\text{hbond}}(d) = \begin{cases} 1, & \text{if} \quad d \leq -0.7 \\ -d/0.7, & \text{if} \quad -0.7 \leq d \leq 0 \\ 0, & \text{if} \quad d \geq 0 \end{cases}$$

All interaction functions are cut off at an interatomic distance of r = 8 Å.

Machine Learning-based. ML-based, or descriptor-based, scoring functions have evolved from early ligand-based quantitative structure-activity relationship (QSAR) models to more advanced ML algorithms to learn non-linear relationships between structural descriptors and binding affinities[48,229]. Instead of predicting binding affinities from predefined descriptors, AF3 employs a deep neural network trained end-to-end to directly predict the three-dimensional structure of molecular complexes from sequence inputs[22]. To assess the reliability of its structural predictions, AF3 uses a dedicated confidence head, which produces internal scoring metrics:

- **The predicted Local Distance Difference Test (pLDDT)**—trained to predict a version of LDDT[230]—provides an estimate of the reliability of each atom's position in the predicted structure
- **The Predicted Aligned Error (PAE)** estimates the positional uncertainty of any token (group of atoms, e.g. amino acids) when aligned according to the frame of another token, and is primarily used to assess the confidence of interfaces.
- **The Predicted Distance Error (PDE)** evaluates the overall structure accuracy by determing the deviation between predicted and true absolute interatomic distances for all atom pairs.
- **The experimentally resolved prediction** indicates whether an atom would be experimentally resolved (e.g. by X-ray crystallography), thereby distinguishing between confidently predicted and potentially disordered or flexible regions.

For the ranking of predicted complex structures, AF3 employs a specialized confidence metric that builds on the scoring strategy introduced in AlphaFold-Multimer[45], which combines two predictors of the Template Modeling score (TM-score)[231]. The predicted TM (pTM) is an alignment-based measure of global structural accuracy that includes all tokens in the complex, whereas the interface pTM (ipTM) is computed exclusively from residue pairs belonging to different chains, capturing the confidence of inter-chain interfaces. In addition, AF3 incorporated penalty terms to account for steric clashes and structural disorder, resulting in an overall model confidence score calculated as:

$$\text{model confidence} = 0.8 \cdot \text{ipTM} + 0.2 \cdot \text{pTM} + 0.5 \cdot \text{disorder} - 100 \cdot \text{has_clash}$$

A structure is marked as having a clash (has_clash), if the total number of inter-chain clashes—atom pairs with a distance less than 1.1 Å—exceeds 100, or if the number of clashes divided by the atom count of the smaller chain exceeds 0.5. The disorder term is based on the relative solvent accessible surface area (rSASA), where atoms with an rSASA value above 0.581 are considered disordered.

2.1.2.2. Docking Box Preparation

In local docking approaches, prior knowledge about the binding site is used to restrict the conformational search to a specific region of the receptor—the docking box—to improve both computational efficiency and prediction accuracy. In this work, local docking was performed using the AutoDock-based programs ADFR, ADCP, and Vina, all of which require receptor and ligand structures in the PDBQT format. This AutoDock-specific extension of the standard PDB format contains additional information such as

atomic partial charges, atom types, and torsional degrees of freedom for rotatable bonds. Conversion from PDB to PDBQT format was carried out using the AutoDock scripts prepare_receptor4.py and prepare_ligand4.py. To define the docking box, the center coordinates (usually the geometric center of the binding site) and box dimensions along each axis are specified. These parameters are sufficient for docking with Vina and were set using AutoDockTools as part of the MGLTools suite[32]. In contrast, ADFR and ADCP require pre-computed grid maps containing affinity potentials for each atom type present in the ligand, generated using AutoGridFR (AGFR)[232], which is built on AutoGrid4[233]. Here, the receptor is embedded within a three-dimensional grid (default spacing: 0.375 Å), and at each grid point, the interaction energy between a probe atom and each receptor atom is calculated (Fig. 2.1). These pre-calculated grid maps allow for rapid estimation of the ligand–receptor interaction energies during docking by trilinear interpolation of the stored grid values[44,226]. In addition, AGFR integrates AutoSite[234] to identify potential binding pockets by clustering high-affinity grid points from carbon, oxygen, and hydrogen affinity maps using predefined energy thresholds. Spatially contiguous clusters of favorable points, referred to as fills, are generated to obtain well-defined binding pocket descriptions. Version 1.1 of AGFR further improves this process by varying energy cutoffs, allowing smaller fills to merge into larger clusters, which is especially useful for larger ligands such as peptides.

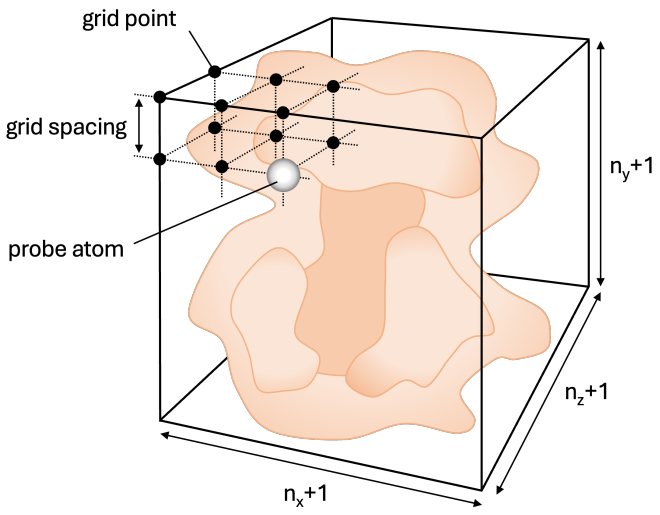


Figure 2.1.: Representation of the AutoDock docking grid. The receptor is embedded within a three-dimensional grid, where grid points are typically spaced at 0.375 Å. At each grid point, a probe atom is placed, and its interaction energy with each receptor atom is computed using AutoGrid. Adapted from Morris et al. (2010)[233].

2.1.2.3. Search Methods

A critical component of molecular docking is the search algorithm, which explores the conformational space to determine the optimal binding pose of a ligand within its receptor. In the following, we highlight four representative approaches: genetic algorithm (GA), an evolutionary algorithm employed in ADFR; iterated local search, a metaheuristic strategy integrated in Vina; a Monte Carlo (MC)-based search algorithm, used in ADCP; and diffusion-based structure prediction, as implemented in AF3.

Genetic Algorithm. The GA is a stochastic optimization technique inspired by the principles of evolution and natural selection. This method represents the initial ligand conformation as a vector of structural parameters—referred to as a chromosome (Fig. 2.2A). A population of such chromosomes is then randomly generated to broadly explore the conformational and energetic landscape. Each member of the population is evaluated by the scoring function, and the best-performing individuals with the lowest energies are selected to produce the next generation. New offspring generations are generated through genetic operators such as crossover by recombining parent chromosomes and mutation by introducing random changes (Fig. 2.2B). This process progressively lowers the average energy of successive generations by preserving favorable structural features, thereby narrowing the conformational space that needs to be explored. Through repeated cycles of conformational sampling and evaluation, the genetic algorithm ideally converges to a conformation corresponding to the global energy minimum[30,235]. ADFR extends the traditional GA framework with several additions, including clustering of top-ranking individuals based on structural similarity (RMSD < 2Å) to maintain diversity by removing duplicate solutions, support adaptive elitism through automatic retention of the best struc-

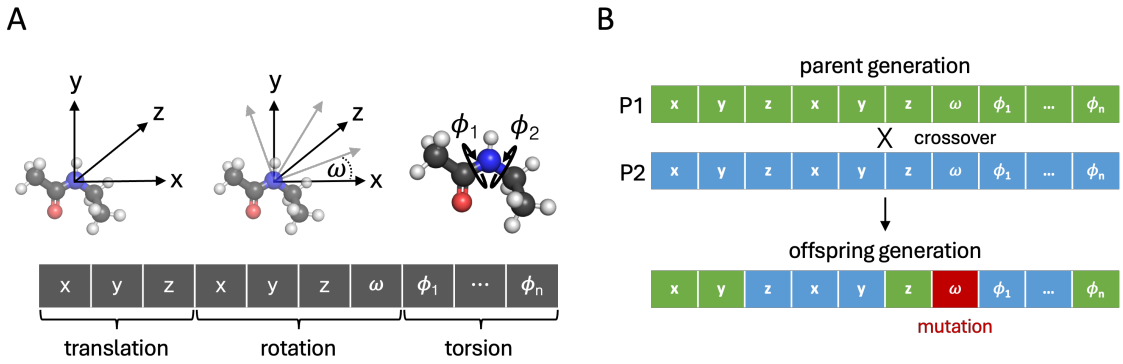


Figure 2.2.: Mechanisms of the genetic algorithm. (A) Representation of a chromosome. The ligand conformation is encoded as a vector of translational, rotational, and torsional parameters, defining its position, orientation, and internal flexibility in the three-dimensional space. (B) Offspring generations are generated by crossover of parent generations and mutation.

tures from each cluster, and monitor convergence to enable a robust termination criterion. The best individual from each cluster is automatically transferred to the next generation (adaptive elitism), while the GA selects parents for crossover, mutation, and subsequent minimization to generate offspring, which then compete with their parents for inclusion in the next generation population. If the cluster solutions remain unchanged for three consecutive generations, the entire population undergoes an additional minimization step. If no changes are observed for five consecutive generations, the docking run is considered converged and the search is terminated.

Iterated Local Search. Vina is based on the Iterated Local Search (ILS) algorithm, which performs iterations of mutation followed by local optimization[32]. After each step, the resulting pose is accepted or rejected according to the Metropolis criterion[236] with propability p_{acc}:

$$p_{\text{acc}} = \min[\,1, \exp(-\beta \Delta E)\,] \tag{2.5}$$

where $\Delta E = E_{\text{new}} - E_{\text{old}}$ is the difference in the total energy between the proposed (E_{new}) and previous (E_{old}) pose, and β is a scaling factor that can be used to adjust the probability of acceptance. The local optimization is performed using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method, a quasi-Newton algorithm commonly applied to solve unconstrained nonlinear optimization problems[237,238]. The BFGS minimization relies on the first- and second-order derivatives of the scoring function (Section 2.1.2.1) with respect to all optimization variables, including translation, rotation and torsion values of the ligand. The second-order derivatives, which are also known as the Hessian, are approximated in BFGS due to their high computational cost. Thus, Vina applies BFGS by evaluating the gradients of the scoring function to determine the search direction and updating the Hessian approximation at each iteration until the method converges to a local optimum, indicated by gradients approaching zero in all directions. Multiple independent runs are performed from random starting conformations, with the number of steps adjusted adaptively based on the problem complexity.

Monte Carlo Search. ADCP uses a Metropolis MC search protocol, adapted from the polypeptide backbone conformation sampler CRANKITE[42,43], that folds the peptide *de novo* within the receptor affinity grid while simultaneously optimizing peptide-receptor interactions[39]. At each step, a new peptide conformation is proposed via an MC move, and the updated structure is evaluated using the force field-based scoring function (Section 2.1.2.1). The move is then accepted or rejected according to a modified MC criterion. ADCP incorporates two backbone perturbation strategies from CRANKITE: (i) crankshaft motions between two randomly selected C_α atoms and (ii) rotations near the peptide ter-

mini (Fig. 2.3A). These were extended with two additional moves: (iii) a local translation to perturb the peptide's position, and (iv) a translational jump that repositions the peptide's central atom to a new translational point—a high-affinity position within the docking box identified by AGFR. As the peptide is modeled in a coarse-grained representation, side chains are reconstructed after each move using a rotamer library[239], allowing for an accurate interaction energy evaluation between peptide and receptor. The probability (p_{acc}) of accepting a proposed move is determined by the Metropolis-like MC criterion

$$p_{\text{acc}} = \begin{cases} 1, & \text{if } \exp\Big(-k_T \Delta E \cdot f^{\text{HC}}(\Delta E)\Big) > r \\ 0 & \text{otherwise} \end{cases}$$

with $\Delta E = E_{\text{new}} - E_{\text{old}}$ being the energy difference between the proposed (E_{new}) and previous (E_{old}) pose, k_T is the temperature scaling factor, and $r \in [0,1]$ is a random number

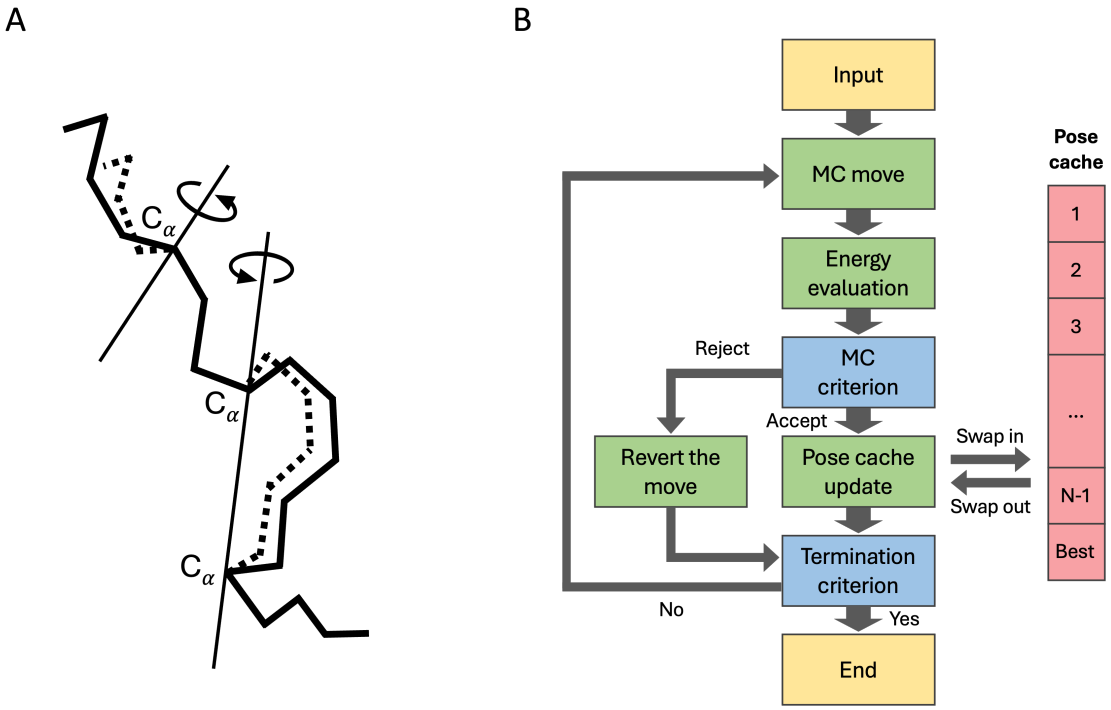


Figure 2.3.: Docking algorithm of AutoDock CrankPep (ADCP). (A) Two types of MC moves are adopted by CRANKITE: crankshaft motion along two selected C_α atoms or rotation near the peptides termini. Adapted from Podtelezhnikov and Wild[43].(B) The workflow of a ADCP docking simulation starts with an MC move, followed by an energy evaluation of the proposed configuration. Depending on the MC criterion, the move is either rejected or accepted. If the MC move is accepted, the new pose is checked against the pose cache. A new cycle begins, if the termination criterion is not met. Adapted from Zhang and Sanner[39].

drawn from a uniform distribution. The HillClimb term $f^{\mathrm{HC}}(\Delta E)$ is defined as:

$$f^{\mathrm{HC}}(\Delta E) = \begin{cases} 0.05, & \text{if } \Delta E > 5.9\,\text{kcal/mol} \\ 1 & \text{otherwise} \end{cases}$$

This approach introduces a controlled bias toward low-energy solutions while still permitting uphill moves, which improves conformational exploration (e.g., moves with a 5.9 kcal/mol score increase have a 60.6 % chance of acceptance, compared to a 36.8 % chance at 11.8 kcal/mol score increase). If the MC move is accepted, the resulting pose is compared to a pose cache and may either be added or replace a similar entry with higher energy. If the search stagnates, it is restarted from a cached pose. The docking run terminates when the maximum number of steps is reached or if no improvement in the best score is observed for 10 million steps. A workflow of a ADCP docking simulation is shown in Fig. 2.3B.

Diffusion-Based Structure Prediction. Instead of explicitly sampling conformational space using traditional search algorithms (e.g., MC or GA), AF3 employs a deep learning approach based on a diffusion model architecture. The AF3 model can be divided into three main sections:

1. **Input preparation**: In the first stage, the model processes the input data—e.g. the sequences of a peptide–protein complex—which are converted into two separate multi-dimensional arrays, referred to as tensors, storing atom-level single (atom identifiers) or pair (atom pair distances) representations. The model further uses Hidden Markov Model (HMM) based methods to scan protein databases for similar sequences, which are aligned and collected into an multiple sequence alignment (MSA). Based on the constructed MSA, the PDB is screened for resembling structures, of which up to four with the highest quality are included in the model as structure templates. The atom-level embeddings are then converted into token-level tensors, while integrating evolutionary information from the MSA.

2. **Representation learning**: The goal of this section is to learn improved representations of the token-level single (TLS) and pair (TLP) tensors. It contains (i) a template module to update TLP based on the structure templates, (ii) an MSA module, which first updates the MSA and then includes it to the TLP and (iii) the Pairformer, in which TLS and TLP are updated in a reciprocal manner using geometry-based (triangle) operations. Each block is repeated multiple times, and the whole section is "recycled" by feeding its output back as input for further refinement.

3. **Structure prediction**: In the final stage, a conditional diffusion model predicts the complex structure guided by the post-Pairformer TLS and TLP embeddings and the original input embeddings. The diffusion model was trained to predict noise that has been iteratively and randomly added to real data during the training phase. At inference time, starting with a matrix of random coordinates, the model removes the predicted noise step-by-step, ultimately generating a fully de-noised 3D atomic structure of the complex.

To promote the generation of multiple conformationally diverse structures, a small amount of noise is added within the diffusion process. The resulting models of the complex structure are then ranked using a global confidence metric that incorporates the pTM and ipTM scores as described in Section 2.1.2.1.

2.1.2.4. Re-Scoring of the Docking Solutions

To filter docking results for ligand poses adopting the desired binding mode, we developed a Python-based re-scoring script. The script evaluates the top-10 ranked solutions from triplicate docking runs based on predefined distance criteria between receptor and ligand atoms. These criteria were selected to capture key interactions observed in reference structures (e.g. X-ray crystal structures of a protein–ligand complex). For each pose, the script loads the corresponding PDB file using the PyMOL (Schrödinger)[240] module and measures specific intermolecular distances between certain ligand atoms and residues within the receptor binding site. Poses exceeding a predefined distance cutoff are excluded from further analysis, while the remaining solutions are re-ranked according to specific criteria, such as their original docking scores.

2.1.3. Molecular Dynamics Simulations

MD simulations enable the study of molecular systems by capturing their dynamical behavior at atomic resolution. This approach is based on the numerical integration of Newton's second law, the fundamental equation of motion in classical mechanics. The motion of each atom i is described by its mass m_i, its acceleration a_i, and the force f_i acting upon it. The force is obtained from the gradient of the potential energy function $U(R)$, where $R = (r_1, r_2, \ldots, r_N)$ represents the complete set of $3N$ atomic coordinates for a system comprising N atoms[241]:

$$f_i = m_i a_i \quad , \quad f_i = -\frac{\partial U(R)}{\partial r_i} \tag{2.6}$$

The potential energy function $U(R)$ is determined by the selected force field, which defines the mathematical form and parameterization of interatomic interactions.

2.1.3.1. The Force Field

In MD simulations, the motion of atoms is determined by calculating the forces acting on each atom. These forces are derived from a molecular mechanics force field, a mathematical model that defines the system's potential energy $U(R)$, which is typically divided into bonded and nonbonded contributions (Fig. 2.4)[242]:

$$U(R) = U_{\text{bonded}}(R) + U_{\text{nonbonded}}(R) \tag{2.7}$$

Bonded interactions include harmonic potentials for bond stretching and angle bending, and periodic functions for dihedral (torsional) rotations:

$$U_{\text{bonded}}(R) = \frac{1}{2}\sum_{\text{bonds}} k_b(b-b_0)^2 + \frac{1}{2}\sum_{\text{angles}} k_\theta(\theta-\theta_0)^2 + \sum_{\text{dihedrals}} k_\phi\,[1+cos(n\phi-\delta)] \tag{2.8}$$

Bond stretching and angle bending introduce energetic penalties for deviations from the equilibrium values b_0 and θ_0, scaled by the respective force constants k_b and k_θ. Torsional rotations around bonds are described by a periodic cosine function with force constant k_ϕ, periodicity n, and phase offset δ. The dihedral angle ϕ is defined by the relative positions of four atoms connected by three covalent bonds[241,243]. Nonbonded interactions include van der Waals forces modeled by the LJ potential and electrostatic forces modeled by Coulomb's law:

$$U_{\text{nonbonded}}(R) = \sum_{\text{LJ}} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{\text{Coulomb}} \frac{q_i q_j}{r_{ij}} \tag{2.9}$$

The LJ potential captures short-range repulsion (r^{-12}) and long-range dispersion (r^{-6}) effects, with ε_{ij} controlling the well depth and σ_{ij} the interatomic distance at which the potential reaches its minimum. Electrostatic interactions are described by Coulomb's law, where q_i and q_j are the atomic charges and r_{ij} is the distance between atoms i and j. The strength of electrostatic interactions increases with the magnitude of the charges, with opposite charges attracting, and atoms with the same type of charge repelling each other[241,244].

Periodic Boundary Conditions. In MD simulations, the finite size of the simulation box can introduce artificial boundary effects that may bias the physical behavior of the system. To avoid these effects, periodic boundary conditions (PBC) are applied, surrounding

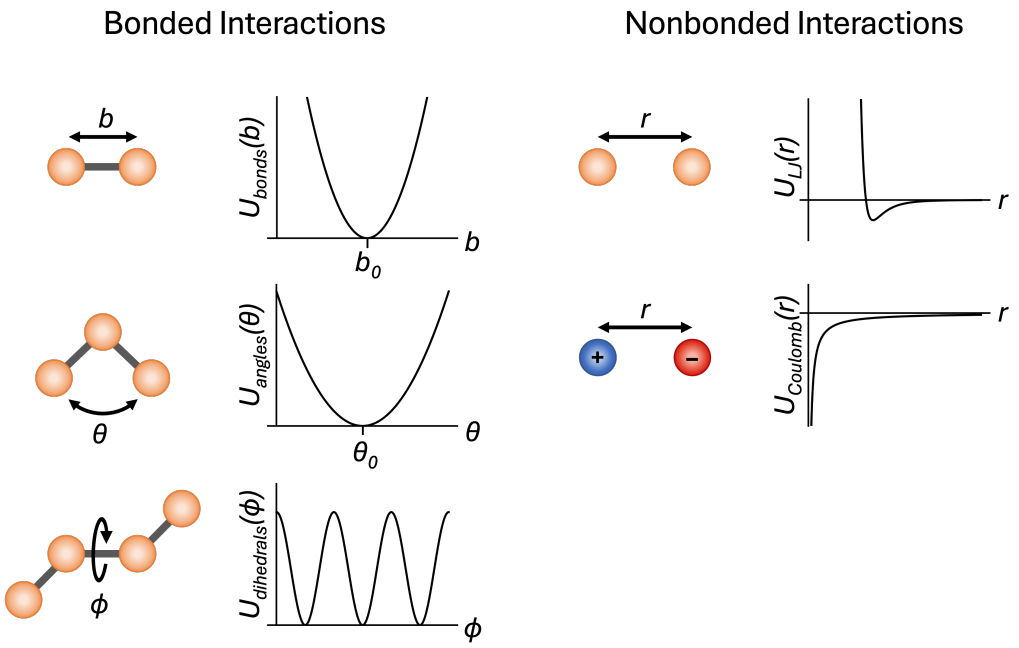


Figure 2.4.: Bonded and nonbonded interactions represented by the force-field potential functions.

the simulation box by identical copies of itself in all directions to create a continuously repeating structure. When a particle moves across the boundary of the primary cell, it re-enters from the opposite side, ensuring conservation of particle number and continuity of the system. As a result, however, each particle can interact with an infinite number of periodic images.

Cutoff Methods. To reduce the computational cost of computing non-bonded interactions (Eq. 2.9), cutoff distances are typically applied. Lennard-Jones interactions, which decay rapidly with distance ($1/r^6$), can be truncated safely using a simple cutoff. In contrast, electrostatic interactions decay slowly ($1/r$) and require a more accurate treatment to avoid artifacts. The particle mesh Ewald (PME) method is widely used to compute long-range electrostatics efficiently in PBC[245,246]. Instead of calculating all pairwise Coulomb interactions explicitly—which would scale as $\mathcal{O}(N^2)$—PME splits the interaction into real space and reciprocal space components:

$$U_{\text{Coulomb}} = U_{\text{real}} + U_{\text{rec}} - U_{\text{self}} \tag{2.10}$$

While the potential U_{real} is calculated in real space by explicitly summing all short-range interactions within a defined cutoff distance, long-range interactions (U_{rec}) are computed in reciprocal space using a Fourier transformation. The term U_{self} corrects for self-interactions.

For the calculation of the long-range interactions, particle charges are interpolated onto a three-dimensional mesh. This charge grid is then transformed into reciprocal space using the fast fourier transform (FFT) algorithm. The electrostatic potential is computed efficiently in Fourier space and subsequently transformed back to real space. Finally, forces on individual atoms are recovered by interpolating the potential values from the mesh back to the particle positions. This approach reduces the computational cost to $\mathcal{O}(N \log N)$, making PME significantly more efficient than traditional Ewald summation, which scales as approximately $\mathcal{O}(N^{3/2})$ [242,247].

The AMBER Force Field. The selection of a force field in an MD simulation depends on the specific application and molecular system. Among the most widely used are the assisted model building with energy refinement (AMBER)[52], chemistry at Harvard macromolecular mechanics (CHARMM)[248,249] and Groningen molecular simulation (GROMOS)[250,251] force field families. In this study, we used the AMBER ff14SB force field[252]. The force field family of the molecular modeling program AMBER, originally developed by Kollman and colleagues in the early 1980s[52], has undergone continuous refinement over the years. Updates such as ff94[253] improved parameter accuracy and laid the foundation for subsequent versions. Later, ff99SB[254] enhanced backbone torsion parameters to better capture protein secondary structure balance. The ff14SB force field further refined both backbone and side-chain torsions based on high-level quantum mechanical calculations and extensive benchmarking, making it particularly well-suited for protein simulations[252].

2.1.3.2. Numerical Integration of the Equations of Motion

The forces computed from the force field serve as the basis for determining atomic trajectories, which are obtained by numerically integrating Newton's equations of motion (Eq. 2.6). Here, the general idea is to estimate the positions, velocities and other factors at the time $t + \Delta t$ based on the previous step at time t. This happens in a step-by-step manner, using a discrete time step Δt, which must be small enough to accurately resolve atomic motions[255,256]. There are several techniques to solve the equation of motion. The Verlet algorithm, introduced by Verlet in 1967[257], and its variants are the most commonly used methods in MD programs[255]. The original Verlet algorithm is based on a Taylor expansion with the positions $r(t)$, the velocities $v(t)$ and accelerations $a(t)$:

$$r(t + \Delta t) = r(t) + \Delta t v(t) + \frac{1}{2}\Delta t^2 a(t) + ... \tag{2.11}$$

$$r(t - \Delta t) = r(t) - \Delta t v(t) + \frac{1}{2}\Delta t^2 a(t) + ... \tag{2.12}$$

By combining both equations we obtain the following function:

$$r(t+\Delta t) = 2r(t) + \Delta t^2 a(t) - r(t-\Delta t) \tag{2.13}$$

The Verlet algorithm is time-reversible because it calculates new positions using both the current and previous positions, $r(t)$ and $r(t-\Delta t)$[255]. Although velocities are essential for calculating the system's kinetic and total energy, they are not explicitly required to propagate the trajectory. A related method is the Leapfrog algorithm[258], which evaluates velocities at half-integer time steps, such as $v(t+\frac{1}{2}\Delta t)$, effectively causing them to "leap over" the positions during integration:

$$v(t+\tfrac{1}{2}\Delta t) = v(t+\tfrac{1}{2}\Delta t) + \Delta t a(t) \tag{2.14}$$

$$r(t+\Delta t) = r(t) + \Delta t v(t+\tfrac{1}{2}\Delta t) \tag{2.15}$$

The Leapfrog algorithm enhances stability and accuracy by updating positions and velocities in a temporally offset manner. However, this decoupling complicates the direct calculation of total energy, since kinetic and potential energies are determined at separate times[256]. To overcome this limitation, the Velocity Verlet algorithm[259] was introduced, which updates both positions and velocities at each full time step:

$$r(t+\Delta t) = r(t) + \Delta t v(t) + \tfrac{1}{2}\Delta t^2 a(t) \tag{2.16}$$

$$v(t+\Delta t) = v(t) + \tfrac{1}{2}\Delta t \left[a(t) + a(t+\Delta t)\right] \tag{2.17}$$

In this scheme, the new velocities can only be computed after the updated positions $r(t+\Delta t)$ and the resulting accelerations $a(t+\Delta t)$, with $a = f/m$, have been calculated[255,256]. The Velocity Verlet algorithm maintains the numerical stability and simplicity of the original Verlet method, while enabling direct computation of total energy and improving its practicality for MD simulations.

The choice of the time step has a significant impact on the computational efficiency. Although larger time steps reduce the total number of integration cycles, their size is limited by the motions with the highest frequency in the system like bond vibrations involving hydrogen atoms. These fast vibrations typically occur on the femtosecond timescale and can therefore restrict the time step to approximately 1 fs. However, when such vibrations are not critical to the simulated process, they can be effectively constrained using specialized algorithms. SHAKE[260] and LINCS[261] are widely used methods that impose rigid constraints on bond lengths. SHAKE applies an iterative correction scheme, while LINCS employs a non-iterative, matrix-based projection approach that

is especially efficient for large and complex biomolecular systems. By applying these constraints, simulations can safely use time steps of up to 2 fs without impairing physical accuracy[242,255].

2.1.3.3. MD Simulation Protocol

The protocol of a typical MD simulation performed in this study using GROMACS (version 2021 or 2023)[51,262] is outlined below.

System Setup. The AMBER ff14SB force field[252] was employed for all simulations. Protonation states of ionizable residues were assigned under standard conditions at neutral pH (Lys and Arg protonated; Asp and Glu deprotonated; His neutral). Each structure was placed in a dodecahedral simulation box with a minimum distance of 1 nm from the solute to the boundaries and solvated using the TIP3P water model[263]. The system was then neutralized, and a physiological salt concentration of 150 mM NaCl was established by randomly replacing water molecules with sodium and chloride ions.

Energy Minimization. To remove initial atom clashes, the system was energy-minimized using the steepest descent algorithm[264] until the maximum force on any atom was below 1 000 kJ mol^{-1} nm^{-1}.

Equilibration. Prior to the production run, the system was equilibrated in two sequential stages. First, a 100 ps MD simulation in the canonical ensemble (NVT) was used to equilibrate the solvent around the position-restrained protein and to establish the target temperature of 310 K (37 °C), using the velocity-rescaling thermostat[265]. Second, a 1 ns simulation in the isothermal-isobaric ensemble (NPT) was conducted to stabilize the pressure at 1 bar, applying the Berendsen barostat[266]. Bond constraints were handled using the LINCS algorithm[261], and long-range electrostatics were computed using the PME method[245,246] under PBC.

Production Simulations. Production simulations were performed at 310 K and 1 bar using the Nosé–Hoover thermostat[267,268] and Parrinello–Rahman barostat[269–271], respectively. A cutoff of 12 Å was applied for van der Waals and short-range Coulombic interactions. Electrostatics were treated with PME, and all bonds were constrained using LINCS.

2.1.4. Binding Free Energy Calculations

The ΔG_{bind} of a protein–ligand complex provides a quantitative measure of the affinity between the ligand and its target protein. A widely used approach to estimate ΔG_{bind} is the molecular mechanics Poisson–Boltzmann surface area (MM/PBSA) or molecular mechanics generalized Born surface area (MM/GBSA) method[272,273]. Here, ΔG_{bind} of the ligand (LIG) and the protein receptor (REC) forming a complex (COM),

$$\Delta G_{\text{bind}} = G_{\text{COM}} - G_{\text{REC}} - G_{\text{LIG}} \tag{2.18}$$

can also be described as

$$\Delta G_{\text{bind}} = \Delta H - T\Delta S = \Delta E_{\text{MM}} + \Delta G_{\text{solv}} - T\Delta S \tag{2.19}$$

where the enthalpy ΔH is decomposed into the molecular mechanics energy ΔE_{MM}, typically consisting of internal ΔE_{int} (bond, angle and dihedral), electrostatic ΔE_{ele}, and van der Waals ΔE_{vdW} interactions,

$$\Delta E_{\text{MM}} = \Delta E_{\text{int}} + \Delta E_{\text{ele}} + \Delta E_{\text{vdW}} \tag{2.20}$$

and the solvation free energy ΔG_{solv} which includes polar terms, generally calculated with MM/PBSA or MM/GBSA, and non-polar contributions, which are usually estimated using the molecules' total solvent-accessible surface area (SASA):

$$\Delta G_{\text{solv}} = \Delta G_{\text{polar}} + \Delta G_{\text{non-polar}} \tag{2.21}$$

The entropic contribution $T\Delta S$ is often neglected, as its evaluation is computationally demanding and highly sensitive. The remaining effective free energy is generally sufficient for comparing the relative binding affinities of similar ligands.
In this study, binding free energy calculations were performed using the gmx_MMPBSA tool[59], which automates MM/PB(GB)SA calculations from GROMACS simulation trajectories. For each protein–ligand system, snapshots from the production MD trajectories were extracted and evaluated using the single-trajectory approach.

2.2. Computational Dynamic Saturation Mutagenesis

We developed and implemented an automated workflow, referred to as CDSM, to provide peptide design suggestions based on differences in ΔG_{bind} after systematically substituting each peptide residue using structural information derived from a peptide–protein complex MD simulation. The protocol begins by extracting a user-defined number of snapshots from the MD trajectory of the peptide–protein complex. For each snapshot, every residue of the peptide is systematically mutated to all other 19 standard proteinogenic amino acids using PyMOLs[240] mutagenesis wizard. Following each mutation, the system undergoes a three-step relaxation procedure using GROMACS: (i) energy minimization, (ii) a 10 ps equilibration under NPT conditions, and (iii) a 100 ps production MD simulation, all under the conditions described in Section 2.1.3.3. After relaxation, the final structures from each snapshot simulation are combined and used for per-residue ΔG_{bind} decomposition calculations with the gmx_MMPBSA tool[59]. The change in binding free energy ($\Delta\Delta G$) upon mutation is calculated for residue i as the difference in ΔG_{bind} between the mutant and the parent (pt):

$$\Delta\Delta G_i = \Delta G_{\mathrm{mut},i} - \Delta G_{\mathrm{pt},i} \tag{2.22}$$

Finally, the $\Delta\Delta G$ values are visualized as a color-coded heat map, providing an overview of favorable and unfavorable mutations at each peptide position.

2.3. Experimental Evaluation of Peptide Inhibitors

The peptide designs obtained from the computer-aided drug design (CADD) study are evaluated *in vitro* for their inhibitory activity against the M^{pro} and PL^{pro}. The experimental work comprises protein expression and purification of both proteases, followed by protease inhibition assays. All chemicals used in this study were purchased from Carl Roth GmbH & Co. KG. (Germany, Karlsruhe), ThermoFisher Scientific (USA, MA, Waltham), Merck KGaA (Germany, Darmstadt) and Th. Geyer GmbH & Co. KG (Germany, Renningen).

2.3.1. Protein Production and Purification

2.3.1.1. Plasmid Generation

Codon-optimized genes encoding M^{pro} (Genbank entry: MT039874, virus strain: SARS-CoV-2/human/CHN/IME-HZ01/2020, ORF1ab 3264-3568 aa) and PL^{pro} (Genbank entry: MN908947.3, virus strain: Wuhan-Hu-1, ORF1ab 1564-1876 aa) were synthesized and

cloned into Novagen pET22b(+) expression vectors by GenScript (Netherlands, Rijswijk). The *pET22b–PL^{pro}–His$_6$* construct includes a C-terminal hexahistidine tag (His-tag) to facilitate purification (Fig. A1a). In contrast, the *pET22b–His$_6$–TEV-site–M^{pro}* plasmid encodes an N-terminal His-tag followed by a Tobacco Etch Virus (TEV) protease cleavage site (TEV-site), enabling production of M^{pro} with an authentic N-terminus after cleavage (Fig. A1b). The sequence ENLYSQS was selected as TEV-site based on findings by Zhao et al. (2024)[274], who identified it as the most efficient for processing in similar systems. To ensure an authentic C-terminus, a termination codon was introduced immediately after the M^{pro} sequence.

2.3.1.2. Recombinant Protein Production

The M^{pro} and PL^{pro} genes were heterologously expressed in *Escherichia coli* (*E. coli*) BL21(DE3) (Genotype: *E. coli* B F$^-$ *dcm ompT hsdS*(rB$^-$ mB$^-$) *gal* λ(DE3)). It carries the gene for the T7 RNA polymerase under control of the *lacUV5* promoter, which can be induced by lactose or lactose analogs such as IPTG, thereby driving expression of target genes cloned downstream of a T7 promoter.

Transformation. Chemically competent *E. coli* BL21(DE3) cells were transformed using the standard heat shock method[275]. Typically, concentrations of 50–100 ng of either the *pET22b–PL^{pro}–His$_6$* or *pET22b–His$_6$–TEV-site–M^{pro}* plasmid was added to an aliquot of competent cells. After incubation on ice for 30 min, a 30 sec heat shock was performed at 42 °C. Subsequently, 700 µL of lysogeny broth (LB) medium supplemented with 0.5 % glucose was added, and the cells were incubated at 37 °C for 10 min to allow recovery and expression of the ampicillin resistance gene. In order to prepare pre-cultures, the transformed cells were then cultured overnight at 37 °C in LB medium containing 0.5 % glucose and 100 μg/ml ampicillin (Amp100).

Gene Expression. Protein production was carried out in LB medium supplemented with the auto-induction solution (0.6 % (v/v) glycerol, 0.2 % (w/v) lactose, 0.05 % (w/v) glucose)[276].
Initially, pre-cultures were prepared in Erlenmeyer flasks filled to one-tenth of their volume with LB medium supplemented with 0.5 % glucose and Amp100, and inoculated with transformed *E. coli* cells. For large-scale production, 500 mL of auto-induction LB medium containing Amp100 was inoculated with the pre-culture to an initial optical density (OD) of 0.05 (λ = 580 nm). Cultures were grown in 5 L Erlenmeyer flasks, filled to one-tenth of their volume for adequate aeration, and shaken at 160 rpm. The incubation was conducted

at 30 °C for 4 h, followed by further incubation at reduced temperatures depending on the target protein: 18 h at 18 °C for PL^{pro} and 16 h at 16 °C for M^{pro}.

Harvest. Cells were harvested by centrifugation at 6 000 × g for 30 min at 4 °C The resulting pellets were either processed immediately or stored at −20 °C.

2.3.1.3. Protein Purification

The proteins were purified by immobilized metal ion affinity chromatography (IMAC) using nickel nitrilotriacetic acid (Ni-NTA) agarose. Cell pellets were resuspended in lysis buffer (20 mM Na_2HPO_4, 500 mM NaCl, 10 mM imidazole, pH 7.4), resulting in a 10–15 % (w/v) suspension. The cells were lysed using a Sonopuls UW 2070 ultrasonic sonotrode (BANDELIN, Germany, Berlin), applying ultrasonic waves twice for 5 min at 30 % power and 50 % cycle. Cell debris and insoluble material were removed by centrifugation (30 min, 4 °C, 36 000 × g).
The Ni-NTA agarose (ThermoFisher Scientific, USA, MA, Waltham) was washed with water and equilibrated with lysis buffer before use. For binding, 6 mL of Ni-NTA matrix per liter of culture (i.e. 3 mL Ni-NTA matrix for 500 mL culture) were added to the supernatant comprising the soluble proteins and incubated for 30 min at 4 °C. The mixture was transferred to a gravity flow column and washed with at least 10 column volumes (CV) of lysis buffer. Bound proteins were eluted with 3 CV of elution buffer (20 mM Na_2HPO_4, 500 mM NaCl, 500 mM imidazole, pH 7.4).
Eluted proteins were concentrated to approximately 2.5 mL by centrifugal ultrafiltration at 3 600 rpm and 4 °Cusing Vivaspin 20 tubes (10,000 MWCO; Sartorius AG, Germany, Göttingen). Buffer exchange was performed using PD-10 columns (GE Healthcare GmbH, Germany, Düsseldorf) following the manufacturer's protocol. Proteins were rebuffered into either M^{pro} buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA), 1 mM tris(2-carboxyethyl)phosphine (TCEP), pH 7.2) or PL^{pro} buffer (50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), 5 mM Dithiothreitol (DTT), 0.01 % 2-[4-(2,4,4-trimethylpentan-2-yl)phenoxy]ethanol (Triton X-100), pH 7.5), and stored at −20 °C.

2.3.1.4. Production of Natural M^{pro}

In order to produce natural M^{pro} ($natM^{pro}$), 1 mg of purified His-tagged M^{pro} with an N-terminal TEV cleavage site was incubated with 1000 IU of TEV protease (GenScript, Rijswijk, Netherlands) for 48 h at 15 °C in M^{pro} buffer. Subsequently, non-cleaved M^{pro} and the His-tagged TEV protease were separated from the $natM^{pro}$ by IMAC. The entire

1 mL digestion mixture was incubated with 1 mL of Ni-NTA agarose (ThermoFisher Scientific, USA, MA, Waltham) for 30 min at 4 °C.
After transferring the matrix to a gravity flow column, the through-flow containing the unbound natM^{pro} was collected. The column was then washed with 2.5 CV of M^{pro} buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, 1 mM TCEP, pH 7.2), and the bound fraction was finally eluted with 3 CV of elution buffer (20 mM Na_2HPO_4, 500 mM NaCl, 500 mM imidazole, pH 7.4).

2.3.1.5. Measurement of Protein Concentration

Protein concentrations were determined using the NanoDrop Spectrophotometer (ThermoFisher Scientific, USA, MA, Waltham). The ProtParam web service[277] was used to compute the molecular weight (natM^{pro}: 33 797 Da, PL^{pro}–His_6: 36 455 Da) and the extinction coefficient (natM^{pro}: 32 890 $M^{-1}\,cm^{-1}$, PL^{pro}–His_6: 45 270 $M^{-1}\,cm^{-1}$) which are required to determine the protein concentration with the Protein A_{280} method.

2.3.1.6. Evaluation of Protein Purification

Protein purity was assessed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) following the protocol of Laemmli[278]. Samples from various purification steps were mixed with SDS sample buffer (50 mM Tris-HCl, 100 mM DTT, 2 % (w/v) SDS, 0.1 % (w/v) bromophenol blue, 8 % (v/v) glycerol) and denatured by boiling for 10 min at 98 °C. The PageRuler Prestained Protein Ladder (ThermoFisher Scientific, USA, MA, Waltham) and prepared samples were loaded onto NuPAGE Bis-Tris 4–12 % gradient gels (ThermoFisher Scientific, USA, MA, Waltham).
Electrophoresis was performed using the XCell SureLock Mini-Cell system (ThermoFisher Scientific, USA, MA, Waltham) with 2-(N-morpholino)ethanesulfonic acid (MES) running buffer (50 mM MES, 50 mM Tris-HCl, 0.1 % SDS, 1 mM EDTA, pH 7.3) at 100 V for 15 min, followed by 200 V for 40 min. Finally, gels were stained with a Coomassie Brilliant Blue R-250 solution (0.1 % Coomassie Brilliant Blue R-250, 10 % (w/v) ammonium sulfate, 1.2 % (v/v) phosphoric acid (85 %)) to visualize protein bands.

2.3.2. Determination of Peptide Inhibitory Activity

The proteolytic activities of natM^{pro} and PL^{pro} were determined using *in vitro* fluorogenic assays. The respective reaction was performed in 96-well microplates with flat bottom in black (Greiner Bio-One GmbH, Österreich, Kremsmünster) and fluorescence intensities (excitation wavelength: 360 nm, emission wavelength: 460 nm) were measured at 30 s intervals over 20 min with the infinite M1000 Pro microplate reader (Tecan Group Ltd.,

Switzerland, Männedorf) at 30 °C.
The reaction mixtures were prepared with M^{pro} buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, 1 mM TCEP, pH 7.2) or PL^{pro} buffer (50 mM HEPES, 5 mM DTT, 0.01 % Triton X-100, pH 7.5) containing 0.5 µM of $natM^{pro}$ or PL^{pro}-His_6, respectively. The fluorogenic substrate Dabcyl-KTSAVLQ↓SGFRKME-Edans (GenScript, Netherlands, Rijswijk) or Dabcyl-FTLKGG↓APTKVTE-Edans was added in a final concentration of 10 µM to initiate the proteolytic reaction by $natM^{pro}$ or PL^{pro}, respectively.
Inhibitory activity was assessed by adding the peptide designs to the reaction mixture, incubating for 15 min at 30 °C, and initiating the reaction with the substrate as described above.

3. Binding Mode of SARS-CoV-2 M^{pro} Retro-Inverso D-Peptide Inhibitors

The emergence of the coronavirus disease 2019 (COVID-19) pandemic created an urgent need for therapeutics targeting severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), with the main protease (M^{pro}) identified as a critical drug target[106].
Eberle and colleagues used phage display to identify L-enantiomeric peptides binding to M^{pro}, and inhibitory assays highlighted three promising candidates: 3CVL-2, 3CVL-4, and 3CVL-7[65]. Among these, 3CVL-2 and 3CVL-7 act as competitive inhibitors, while 3CVL-4 functions as an allosteric inhibitor. To enhance the peptides' proteolytic stability without losing affinity or binding specificity, retro-inverso D-enantiomeric (Dri) versions (Fig. 1.3), 3CVLri-2, 3CVLri-4, and 3CVLri-7, were synthesized. Their amino acid sequences and inhibitory potential, as estimated by their half-maximal inhibitory concentration (IC_{50}), are summarized in Table 3.1.
Using molecular docking and molecular dynamics (MD) simulations, we aim to elucidate the interaction patterns between the protease and peptide to promote the rational design of optimized peptide inhibitors with improved binding affinity.

Table 3.1.: Peptide inhibitors targeting SARS-CoV-2 M^{pro}. For each peptide inhibitor the amino acid sequences are shown, with upper case letters indicating L-enantiomeric and lower case letters indicating D-enantiomeric amino acids. The table includes IC_{50} values from experimental inhibitory assays and the inhibition mode (competitive or non-competitive).

name	**sequence**	**IC_{50} (µM)**	**inhibition mode**
3CVL-2	SPHGWPSQSIEVQPQW	4.9 ± 0.8	competitive
3CVL-4	AHEGWTWDWTPQYSWK	2.4 ± 0.6	non-competitive
3CVL-7	TVAPLHAHYWDVEERH	1.0 ± 0.2	competitive
3CVLri-2	wqpqveisqspwahpsr	9.1 ± 0.5	competitive
3CVLri-4	kwsyqptwdwtwaehar	5.4 ± 1.7	non-competitive
3CVLri-7	hreevdwyhahlpavtr	1.6 ± 0.2	competitive

The results of this study have been published in Publication I (Section A.2):

Discovery of All-D-Peptide Inhibitors of SARS-CoV-2 3C-like Protease
Raphael J. Eberle, Marc Sevenich, Ian Gering, Lara Scharbert, Birgit Strodel, Nils A. Lakomek, Karoline Santur, Jeannine Mohrlüder, Mônika A. Coronado, and Dieter Willbold
ACS Chemical Biology, 2023, DOI:10.1021/acschembio.2c00735.

My contribution to this publication involves all computational simulations and analyses, as described in the following sections.

3.1. Methods

A comprehensive description of the general methods used in this project is provided in Section 2.1. In the following, we outline the project-specific procedures in detail.
All MD simulations and analyses were performed with GROMACS version 2021 [51].

Structure Preparation. The peptide models (Table 3.1) were constructed as linear chains using the Python module PeptideBuilder [279]. To obtain Dri peptides, the peptides were first generated as L-enantiomeric molecules with reversed sequence, followed by the inversion of their stereochemical configuration by flipping the PDB files along the x axis as described by Garton et al. (2018) [280] (Fig. 1.3). The crystal structure of M^{pro} was obtained from the protein data bank (PDB: 6M2N). The M^{pro} was modeled as a dimer using chain A and chain C as protomers.

Peptide Simulations. All peptides were centered in a dodecahedron box with a volume of 240.8 nm^3 - 370.5 nm^3, solvated with water and neutralised with Na^+ and Cl^- ions resulting in a system size of ∼23,500-36,000 atoms. The system setup, energy minimization, equilibration in NVT and NPT ensemble, and the production run was conducted as described in Section 2.1.3.3. We performed 1 µs MD simulations of each peptide at 310 K. Secondary structure analysis of the trajectories was performed using the dictionary of protein secondary structures (DSSP) program [281,282]. To identify the peptide structures with the most populated conformation, we applied the clustering algorithm as described by Daura et al. (1999) [283] using a 3.5 Å root mean square deviation (RMSD) cutoff. The 10 most populated clusters were then used for individual docking runs against M^{pro}.

Docking. First, a 200 ns MD simulation of the M^{pro} dimer (PDB: 6M2N) without ligand was conducted to study the behavior of the protomers and detect a potential collapse of one active site. The pocket volume of both active sites was monitored using the VMD plugin Epock version 1.0.5[284].
Prior to docking, the crystal structure of the M^{pro} dimer (PDB: 6M2N) and the peptide structures were prepared with AutoDockTools[44] from the MGLTools software suite[285]. After adding hydrogen atoms to the M^{pro} dimer and the peptides, Gasteiger partial charges were computed to obtain structure files in PDBQT format. Docking grids were generated around the active site (His41, Met49, Tyr54, Phe140, Leu141, Asn142, Gly143, Ser144, Cys145, His163, His164, Met165, Glu166, Leu167, Asp187,Arg188, Gln189, Thr190, and Gln192; box dimensions: 40 Å× 40 Å× 40 Å) and a potential allosteric binding site[286] (Lys12, Lys97, Lys100, Tyr101, Lys102, Phe103, Val104 and Arg105; box dimensions: 60 Å× 50 Å× 80 Å) of chain A. The competitive peptides, 3CVL-2, 3CVL-7, 3CVLri-2 and 3CVLri-7, were then docked against the active site, and the non-competitive 3CVL-4 and 3CVLri-4 against the potential allosteric binding site using AutoDock Vina (Vina)[32] for flexible ligands and rigid receptor docking.

Complex Simulations. The system setup, energy minimization, equilibration in NVT and NPT ensemble, and the production run was conducted as described in Section 2.1.3.3. Each M^{pro}–peptide system (∼86,000 atoms) was simulated for 100 ns at 310 K. Per peptide, one simulation was selected to be extended to 200 ns and performed in triplicate.
The MD trajectories were analyzed by calculating the RMSD and the root mean square fluctuation (RMSF) of the peptides, using the backbone atoms (Cα, C, N), to examine their flexibility within the binding site and to identify the flexible and rigid areas of the peptide. The binding affinity was estimated by computing the binding free energy (ΔG_{bind}) between the protein and the peptide with gmx_MMPBSA[59] using the molecular mechanics Poisson–Boltzmann surface area (MM/PBSA) method. The temperature was set to 310 K and the ionic strength to 150 mM.
After clustering the peptide conformations within the binding site, we determined the minimum distance between the residues of protein and ligand for the three largest clusters in order to define the binding site and to identify interacting residues.

3.2. Results

In vitro assays indicated inhibitory activity of L-enantiomeric peptides, 3CVL-2, 3CVL-4 and 3CVL-7, and their Dri counterparts, 3CVLri-2, 3CVLri-4 and 3CVLri-7, against the SARS-CoV-2 M^{pro}. To elucidate their molecular interaction pattern with the protease, we performed peptide–protein docking and subsequent MD simulations with the L- and Dri-peptides, and the M^{pro} dimer.

3.2.1. Sampling Peptide Conformations

To investigate their conformational landscape, we performed 1 µs MD simulations of the unbound L- and Dri-peptides starting from a fully extended form. Secondary structure analysis of the trajectories revealed that the peptides predominantly exhibit random coil, bend, and turn structures (>70 %), while β-strands are almost never observed (Fig. 3.1A). Helices are most often found in the non-competitively inhibiting 3CVL-4, in both L- and Dri-enantiomeric versions (∼30 %). The competitive inhibitors 3CVL-2 and 3CVL-7 sample more helices in their Dri- (∼20 %) than in their L-configuration (< 6 %). Circular dichroism (CD) spectroscopy analyses could confirm a predominant random coil structure within all 3CVL and 3CVLri peptides[65].

To obtain a set of peptide structures that represent the most populated conformations sampled during the simulations, we conducted cluster analysis using a 3.5 Å RMSD cutoff. This cutoff was applied to ensure that the first ten clusters represent >75 % of all sampled

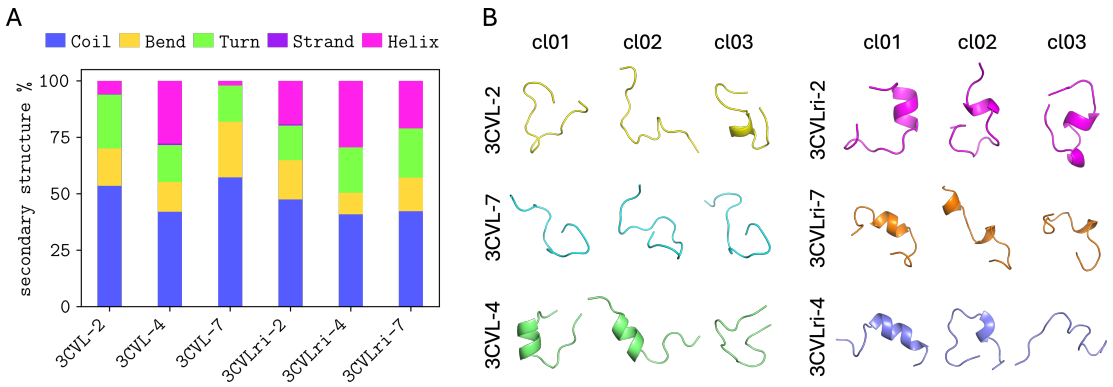


Figure 3.1.: Secondary structure analysis of the MD simulations with unbound inhibitory L- and Dri-peptides. (A) The trajectories of the peptides 3CVL(ri)-2, 3CVL(ri)-4 and 3CVL(ri)-7 were analyzed using the DSSP program to identify the percentage of secondary structure elements, summed over the trajectory and peptide residues. The secondary structure elements include random coil (irregular), bend, turn, helix (α-helix, 3_{10}-helix and π-helix) or β-strand (β-bridge and extended strand) structures. (B) The secondary structure of the peptides is shown for the three most populated conformations (cl01–cl03) obtained through cluster analysis.

conformations. The representative structures of the first ten clusters comprise a balanced mixture of folded and extended conformations, including some structures with short helices. Peptide conformations representing the first three clusters (>46 %) are shown in Fig. 3.1B. This variety of peptide structures within this set will be used as starting structures for docking against M^{pro} in the following steps.

3.2.2. M^{pro}–Peptide Complex Generation

The M^{pro} is catalytically active only in its homodimeric form[122]. It has been hypothesized that only one protomer of the dimer is active at a time[287], while the active site of the other protomer collapses[286].

To investigate the dynamic behavior of the active sites in the M^{pro} dimer, we performed a 200 ns MD simulation of the crystal structure (PDB: 6M2N) without any bound ligand. Active site volume measurements of the trajectory revealed no differences between the two protomers (Fig. A15).

Therefore, we randomly selected the active site of chain A for docking. The ten different structures obtained through clustering of the unbound peptide simulations in the previous section were docked against the active site (3CVL(ri)-2 and 3CVL(ri)-7) or a potential allosteric site[286] of M^{pro} in individual docking runs using Vina. The docking solutions for the 3CVL and 3CVLri peptides all exhibit binding affinities in the range of -9.7 to -5.6 kcal/mol (Fig. A16). The best affinities were obtained for the non-competitive peptides, 3CVL-4 and 3CVLri-4, which exhibit two distinct binding modes to the M^{pro} dimer near the potential allosteric site (Fig. 3.2). In some docking solutions, the peptides are bound to a cleft between domain II and domain III of chain A. The best performing complexes, however, are associated with extended peptide conformations, binding along the interface of both M^{pro} protomers, as observed for the cluster 8 3CVL-4 (-9.7 kcal/mol) and cluster 4 3CVLri-4 (cl04: -9.2 kcal/mol) docking poses. Docking solutions of the competitive inhibitors 3CVL(ri)-2 and 3CVL(ri)-7 demonstrate a clear preference for the S-subpocket of the M^{pro} active site, where the peptides exploit the more accessible binding cavities, in contrast to the sterically constrained S′-subsite. The competitive inhibitors are also frequently observed interacting with the other protomer (chain B).

To effectively inhibit protein activity, competitive inhibitors are required to block substrate access to the catalytic site. Thus, we introduced an additional criterion beyond predicted binding affinity, defined as low minimum distances between the peptide inhibitors and the catalytic dyad residues (Cys145 and His41). For non-competitive peptides, the distance criterion was adapted to residues belonging to the proposed allosteric binding site (Lys12, Lys97, Lys100, Tyr101, Lys102, Phe103, Val104 and Arg105).

After ranking the docking poses based on these selection criteria (Table A16), three

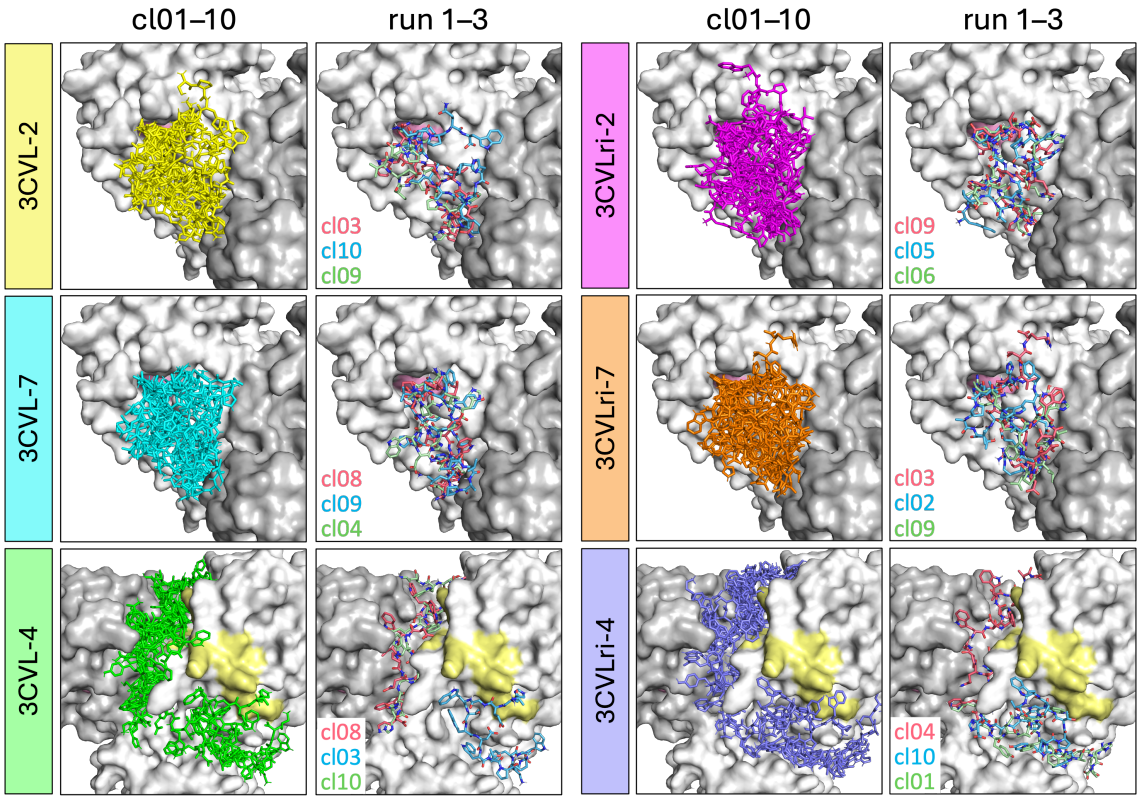


Figure 3.2.: Docking solutions and starting structures for MD simulations in complex with M^{pro}. The 3CVL and 3CVLri inhibitors were docked to M^{pro} using peptide conformations derived from cluster analysis of unbound peptide MD trajectories (cl01–10). From the resulting top-ranked M^{pro}–peptide models, three were selected for further MD simulations (runs 1–3, colored red, blue, and green). The superimposed docking poses are shown as stick representations. The M^{pro} dimer surface is illustrated with chain A in white and chain B in grey, while the catalytic dyad and allosteric binding site are highlighted in pink and yellow, respectively.

different peptide poses of each 3CVL and 3CVLri in complex with M^{pro} were submitted to MD simulations to investigate the stability and interaction modes of the generated complexes in detail. This includes one pose with the highest binding affinity (run 1), one pose with the lowest minimum distance to the binding site (run 2), and one pose that optimally combined both criteria (run 3). Their structures are shown in Fig. 3.2.

3.2.3. Simulation of M^{pro}–Peptide Complexes

For each 3CVL and 3CVLri peptide, the three different M^{pro}–peptide models, obtained from the docking approach, were submitted to 100 ns MD simulations. To estimate the stability and affinity of the M^{pro}–3CVL(ri) complexes, the trajectories were analyzed using RMSD and ΔG_{bind} calculations, respectively.
While all peptides remained bound to M^{pro} throughout the simulation, they deviated from

their initial structures to varying degrees due to internal conformational rearrangements. RMSD analyses show that the poses selected based on binding affinity (run1) generally exhibit higher stability and deviate less from the initial conformation (Fig. A17). Except for 3CVL-2 and 3CVLri-2, complexes in run1 also exhibit the most favorable ΔG_{bind} estimates (Fig. A18), supporting the binding affinity predictions from the docking approach.
In addition to assessing stability and binding affinity, we also investigated the peptides' potential to effectively block or disrupt the active sites of the M^{pro} dimer. This ability can be estimated by identifying the peptides predominant poses within the binding sites through conformational cluster analysis. In Fig. A19 the competitive inhibitors, 3CVL(ri)-2 and 3CVL(ri)-7, bound to the active site of chain A are illustrated. This visualization facilitates the evaluation of their ability to block the catalytic site, which is highlighted in pink. In contrast, the binding poses of the non-competitive inhibitor 3CVL(ri)-4 (Fig. A20) do not allow for an extrapolation of their potential effect on the M^{pro} activity. To elucidate the extent of occupation by the competitive peptides and the effect of the non-competitive peptides on the active sites, we computed the volume of the active sites of both protomers (chain A and chain B) throughout the simulations using Epock[284]. The changes in active site volume upon peptide inhibitor binding are shown for the complex MD runs 1–3 in Fig. 3.3. With the exception of 3CVL-7 in run3, which binds remotely from the catalytic site (Fig. A19), the competitive peptides effectively occupy the chain A active site, decreasing its volume by up to $\sim$600 Å^2. Interestingly, they also often show a reducing effect on the active site volume of the other protomer (chain B), which would support the hypothesis of a single active protomer with a collapsing active site in the other protomer[286,287].
The MD simulation runs of the non-competitive inhibitors in complex with M^{pro} reveal mixed effects on the active sites. Reduced active site volumes in both protomers are observed only for 3CVL-4 in run 3 and for 3CVLri-4 in run 1. In these runs, the peptides interact with both protomers, as they bind at the interface of chain A and chain B. However, when the peptides interact solely with chain A—as in run 2 of 3CVL-4 and runs 2 and 3 of 3CVLri-4—the active site volume of chain A may be reduced, however, that of chain B is even increased. To investigate the M^{pro}–peptide interactions in more detail, we selected the most promising simulation run—based on the combined criteria of affinity, stability and the effect on active site volume—to be extended to 200 ns and performed in triplicate.

3.2.4. Extensive Analysis of 3CVL and 3CVLri Binding Modes

We performed more extensive MD simulations of the most promising structures to gain deeper insights into the interaction patterns between M^{pro} and the 3CVL/3CVLri peptides. Each M^{pro}–peptide complex remained stable with larger conformational rearrangements

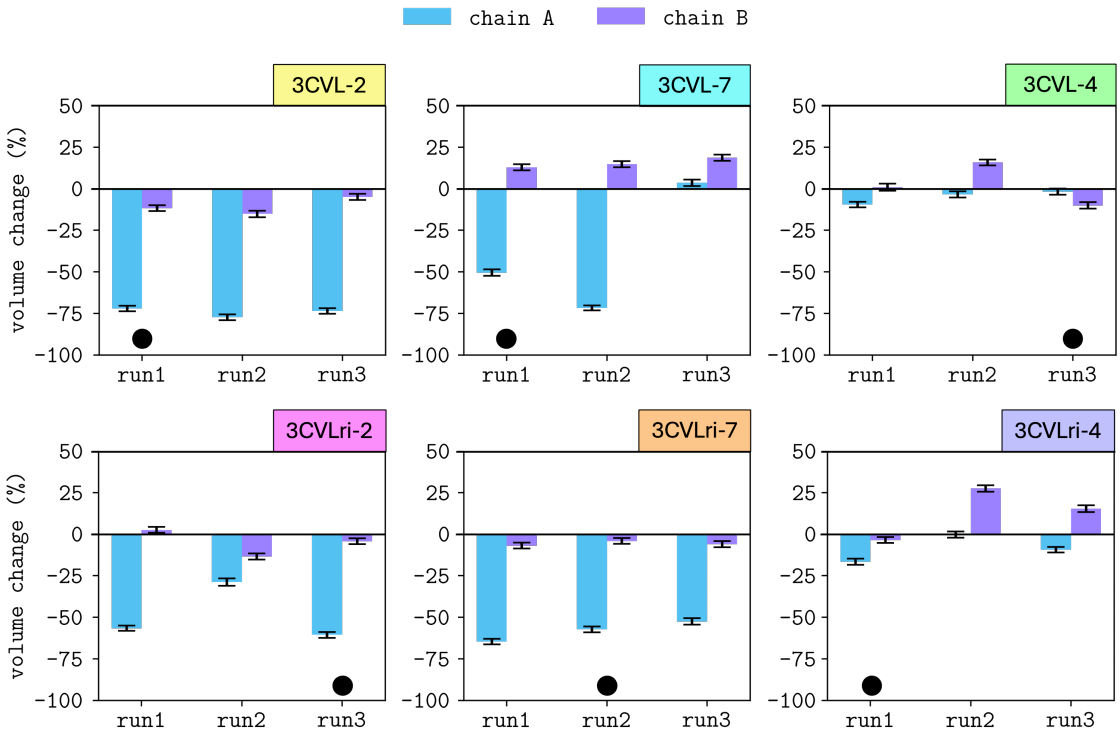


Figure 3.3.: Change of the active site volume of M^{pro} in complex with 3CVL and 3CVLri peptides. The active site volume of chain A and chain B was calculated for the MD trajectories of M^{pro} in complex with 3CVL(ri)-2, 3CVL(ri)-7, and 3CVL(ri)-4, for each run (1-3). The barplot shows the change of the active site volume upon inhibitor binding as the deviation of the averaged volume of the complex trajectory from that of the unbound (apo) M^{pro} simulation in percent: $(\mathrm{Vol}_{\mathrm{complex}} - \mathrm{Vol}_{\mathrm{apo}})/\mathrm{Vol}_{\mathrm{apo}} \times 100\,\%$. For the competitive inhibitors (3CVL(ri)-2 and 3CVL(ri)-7), the chain A bar includes the volume displaced by the peptide to compare their occupation potential. The black dots indicate runs that were selected for extended triplicate simulations.

restricted to the first 50 ns of the 200 ns MD simulations (Fig. A21). By identifying the most popular peptide poses using cluster analysis, valuable information on the binding mode of 3CVL and 3CVLri peptides was obtained.

Fig. 3.4A illustrates the distinct interaction patterns of the competitive inhibitors with the substrate-binding site. While 3CVL-2 extends between the S1 and S3 subpockets toward the S2 subsite, 3CVLri-2 spans the S1′, S1, S2, and S3 subsites.

The interaction areas of 3CVL-7 and 3CVLri-7 both involve residues of the S1′–S3 subsites. However, while 3CVL-7 engages this region through its N-terminal segment, 3CVLri-7 primarily interacts via its C-terminal region. The minimum distance between each M^{pro} binding site and peptide residue was calculated to create a contact map representing the interaction mode of each peptide with the M^{pro} binding site. The graphs in Fig. 3.4B reveal that all competitive peptides contain residues in close proximity to the catalytic dyad (His41 and Cys145), suggesting their potential to block substrate binding.

We computed the per-residue ΔG_{bind} decomposed per peptide residue and plotted it alongside the RMSF in Fig. 3.5. This representation enables the identification of peptide residues that contribute most significantly to the affinity and stability of the peptide–protein complex. For the competitive peptides, residues that bind closely to the M^{pro} active site—within the S1′–S3 region—predominantly exhibit lower ΔG_{bind} and RMSF values, indicating more favorable interactions and greater stability. Higher flexibility is primarily observed at the

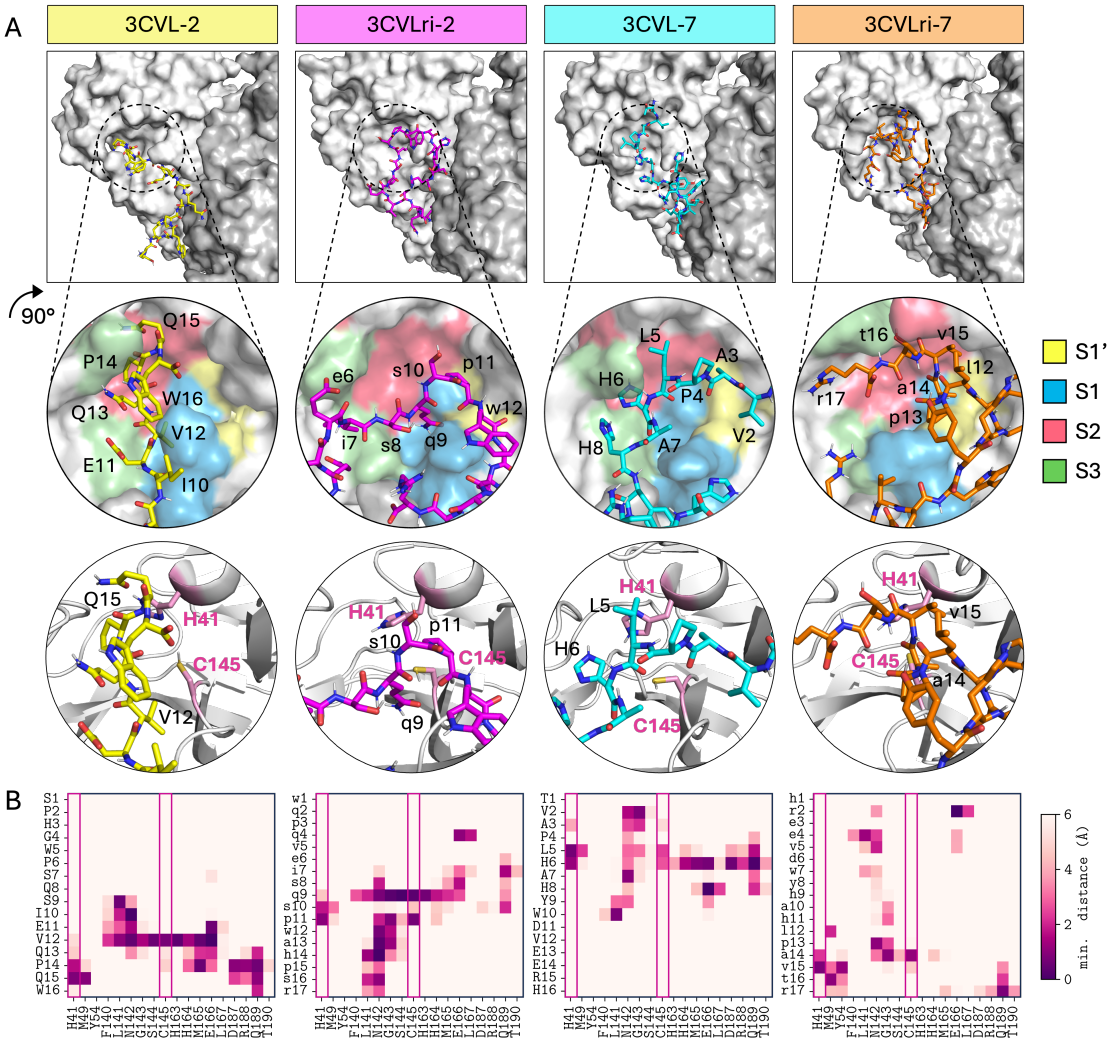


Figure 3.4.: Binding modes of the competitive 3CVL(ri)-2 and 3CVL(ri)-7 peptide inhibitors and their interactions with M^{pro} active site residues. (A) The most populated pose of the 3CVL(ri)-2 and 3CVL(ri)-7 peptides at the M^{pro} active site was computed by cluster analysis of the triplicate 200 ns MD simulations. The structure of the M^{pro} dimer is represented as surface in white (chain A) and grey (chain B). The 3CVL and 3CVLri peptides are shown as sticks. The individual subsites of the active site are marked by different colors (S1′: yellow; S1: blue; S2: red; S3: green). The catalytic dyad residues (H41 and C145) are shown as sticks (pink) in the cartoon representation of M^{pro} below. (B) The minimum distance between each peptide and M^{pro} binding site residue is represented as a color map, with H41 and C145 marked by a pink frame. Adapted figure from Eberle et al. (2023)[65]

peptide termini, which often correlates with less favorable ΔG_{bind} values compared to the tightly bound regions. An interesting observation is that a Pro residue is positioned at the active site for each competitive inhibitor, predominantly occupying the S1′ subsite, as seen for 3CVLri-2, 3CVL-7, and 3CVLri-7. In contrast, 3CVL-2 does not interact with the S1′ subsite, instead its strongest contacts are with the S1 and S2 subsites through Val12, Pro14, and Gln15. The most favorable interaction in 3CVLri-2 arises from one of its D-Gln residues occupying the S1 subsite, showing similarities to the M^{pro}-substrate interaction pattern, which is characterized by the highly conserved Gln at the P1 position. While 3CVL-2 and 3CVLri-2 exhibit strong interactions with the S1 subsite through occupation

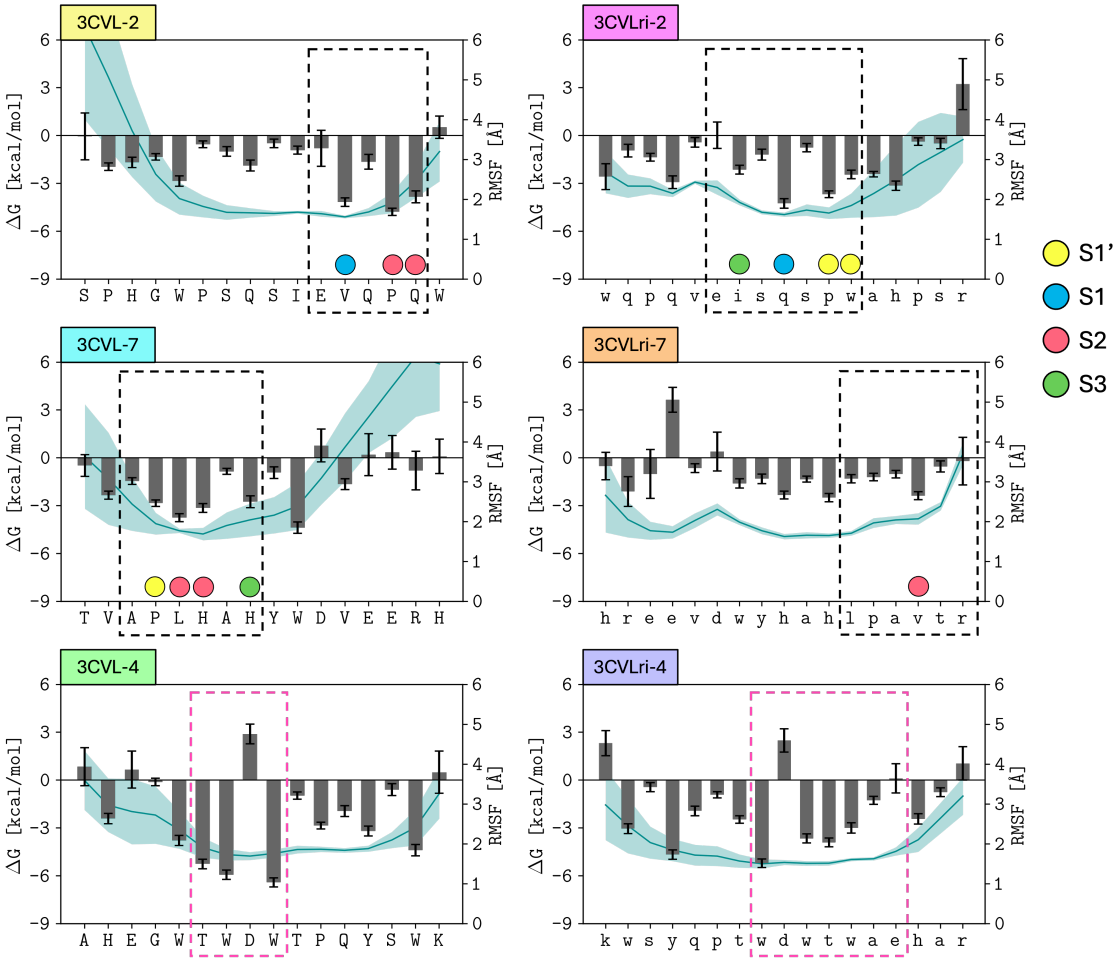


Figure 3.5.: Per-residue binding free energy decomposition of the 3CVL and 3CVLri peptides simulated in complex with M^{pro}. The binding free energy was calculated for each M^{pro}–3CVL(ri) complex. The per-residue ΔG_{bind} decomposition (grey bars) and RMSF (blue line) are shown for each peptide as the average (±SEM) of triplicate 200 ns MD simulations. Residues of the competitive peptides that are bound closely to the active site (S1′-S3) of chain A, and residues of non-competitive peptides interacting with Lys12 or Lys97 of chains A and B are indicated by black and pink dashed frames, respectively. The most favorable residues (< −2 kcal/mol) within the active site frame are marked by a yellow (S1′), blue (S1), red (S2), or green (S3) circle, corresponding to the subsite they occupy.

by Val12 and D-Gln9, respectively, 3CVL-7 and 3CVLri-7 have minimal interactions with this pocket. Instead, their binding affinity to the active site primarily stems from interactions with the S2 subsite. Interestingly, the most favorable ΔG_{bind} contribution for 3CVL-7 results from W10 binding outside the active site, in a pocket between chains A and B of the M^{pro} dimer.

Total ΔG_{bind} calculations revealed that 3CVLri-2 exhibits the highest affinity for M^{pro} among all competitive peptides (Fig. A22A). Furthermore, analysis of the active site volume of chain A shows that 3CVLri-2 displaces the largest volume, indicating a high occupancy potential (Fig. A22B). This superior binding may be attributed to favorable active site interactions, particularly through D-Gln at the S1 subsite, as well as its high stability across all replica simulations, as indicated by the lowest RMSD values throughout the simulations. Unlike the other competitive peptides, which interact with the active site primarily through their terminal regions, 3CVLri-2 engages via its central segment, likely contributing to its enhanced stability.

Fig. 3.6A depicts the non-competitive inhibitors 3CVL-4 and 3CVLri-4 bound in a fully extended conformation within a cleft between both M^{pro} protomers, with their side chains occupying cavities in chains A and B. Minimum distance analysis reveals that two residues, Lys12 and Lys97, in both chains A and B play a critical role in interacting with 3CVL-4 and 3CVLri-4, forming hydrogen bonds with their backbones and side chains (Fig. 3.6B). Per-residue ΔG_{bind} decomposition and RMSF analysis (Fig. 3.5) further suggest that residues of the non-competitive peptides interacting with Lys12 and Lys97 (pink frame) contribute very favorably to ΔG_{bind}. In particular, aromatic residues such as Trp and Tyr exhibit high affinities for the M^{pro} dimer. Notably, their strategic positioning within the 3CVL(ri)-4 peptides, particularly in an alternating arrangement, appears to facilitate their occupation of specific pockets in the opposing protomers. At the same time, their backbones remain extended within the cleft, stabilized by hydrogen bonds with Lys12 and Lys97, as demonstrated by 3CVLri-4 in Fig. 3.6C.

Despite Asp8 (3CVL-4) and D-Asp9 (3CVLri-4) appear to form hydrogen bonds with Lys97 (chain B), their ΔG_{bind} values are highly unfavorable. In general, negatively charged residues across all peptides exhibit less favorable interactions combined with higher uncertainties in ΔG_{bind}.

Both 3CVL-4 and 3CVLri-4 reduce the active site volume of chain A by 18 % and 24 %, respectively, with minimal uncertainty across all replicas, while the other protomer remains largely unaffected (Fig. A22). Furthermore, both peptides exhibit high total ΔG_{bind} values and highly stable conformations, as indicated by low RMSD and RMSF values. These characteristics support the proposed binding modes and identify the 3CVL(ri)-4 binding site as a high-potential allosteric site.

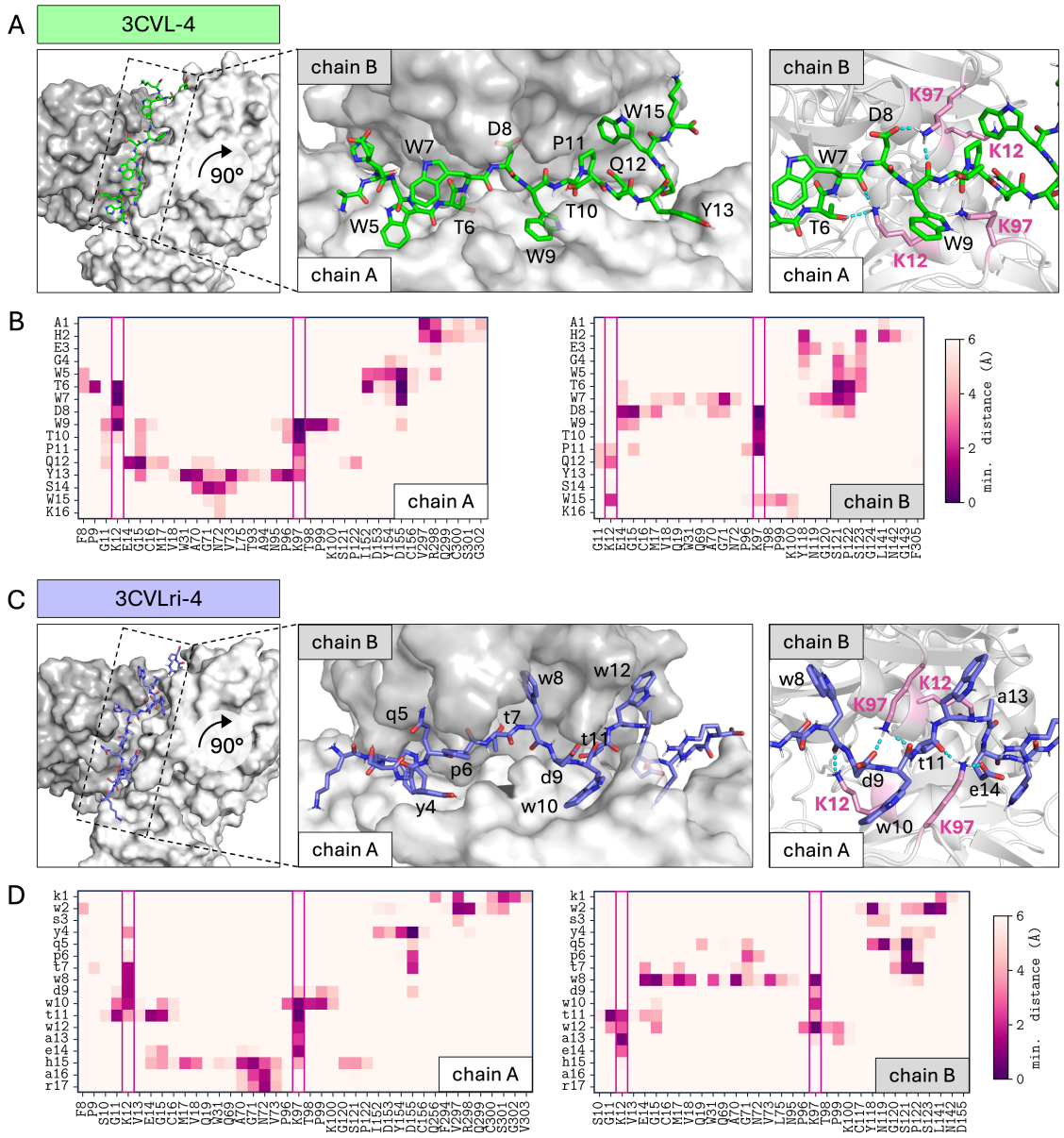


Figure 3.6.: Binding modes of the non-competitive 3CVL(ri)-4 peptide inhibitors and their interactions with M^{pro} residues of a supposed allosteric binding site. The most populated pose of the (A) 3CVL-4 and (C) 3CVLri-4 peptides bound to the cleft between both M^{pro} protomers was computed by cluster analysis of the triplicate 200 ns MD simulations. The structure of the M^{pro} dimer is shown as surface (left, center) or cartoon (right) representation, with chain A in white and chain B in grey. The peptides are depicted as sticks. In the right figure, two residues (K12 and K97) of chain A and chain B that are involved in interactions with both peptides, are highlighted in pink and depicted as sticks in the cartoon representation of M^{pro} below. Hydrogen bonds between these residues and the peptide are depicted as dashed lines (cyan). The minimum distance between each peptide and M^{pro} binding site residue is represented for (B) 3CVL-4 and (D) 3CVLri-4 as a color map, with K12 and K97 marked by a pink frame. Adapted figure from Eberle et al. (2023)[65]

3.3. Discussion

The aim of this work was to elucidate the molecular interaction patterns between the SARS-CoV-2 M^{pro} and three peptide inhibitors, 3CVL-2, 3CVL-4, and 3CVL-7, as well as their retro-inverso D-enantiomeric counterparts, 3CVLri-2, 3CVLri-4, and 3CVLri-7. The computational workflow for each peptide involved sampling different peptide conformations using MD simulations, which were then docked to the active site (for competitive inhibitors) or a potential allosteric binding site (for non-competitive inhibitors) of the M^{pro} dimer. Based on affinity and binding site distance criteria, three docking structures were selected for further evaluation by MD simulations. The most promising simulation—showing a good combination of binding affinity, complex stability, and inhibition potential—was extended to 200 ns and performed in triplicates to thoroughly investigate the binding mode and affinity.

The docking solutions for 3CVL(ri)-2 and 3CVL(ri)-7 showed predominant positioning at the S-subsite of the substrate binding area, with rare extension to the S′-subsites and very diverse poses. In contrast, the non-competitive peptides 3CVL-4 and 3CVLri-4 exhibit two binding modes, both characterized by primarily extended conformations: either binding to the interface of both protomers or to the cleft between domain I and II of chain A. MD simulations of the complex structures revealed that peptides binding to the interface of both protomers exhibit the most favorable ΔG_{bind} and have the ability to reduce the volume of both active sites in the M^{pro} dimer, potentially inhibiting substrate binding.

MD simulations of the competitive L- and Dri-peptide inhibitors revealed multiple stable binding modes, indicating diverse interaction possibilities between these large peptides and the M^{pro} active site. We also did not observe similarities in the interaction patterns in the most promising poses of the L- and Dri-versions of the same peptide. This contradicts expectations, as the Dri-peptides were specifically designed to enhance proteolytic stability while preserving the binding characteristics of their L-counterparts. However, analysis of binding modes and affinity of the extensive simulations showed some recurring patterns, such as the positioning of Pro near the active site, especially at the P1′ site. In both 3CVL-2 and 3CVLri-2, Gln residues were positioned close to the active site. In 3CVLri-2, one D-Gln9 occupied the P1 subpocket with high affinity, showing similarities to the M^{pro} substrates that have Gln conserved at position P1.

For the non-competitive peptides 3CVL-4 and 3CVLri-4, we observed that Lys12 and Lys97, located at the protomer interface, form key interactions with both the backbone and side chains of the inhibitors. With four Trp and one Tyr residues, the 3CVL(ri)-4 peptides contain the highest number of aromatic residues compared to 3CVL(ri)-2 and 3CVL(ri)-7, which each possess only two. The opposing protomers provide multiple cavities that are effectively occupied by these aromatic residues, which contribute most favorably to ΔG_{bind}.

Moreover, both 3CVL-4 and 3CVLri-4 demonstrated the ability to reduce the active site volume of chain A by 18 % and 24 %, respectively. Elucidating the molecular mechanism underlying this reduction, as well as its potential to disrupt substrate binding or protease activity, would require further investigation in future studies. However, our results strongly suggest that the allosteric binding site is located at the interface of the two protomers. The high affinity and stability of the non-competitive peptides at this binding site further support their proposed binding mode.

The computationally determined ΔG_{bind} values for the competitive peptides show no correlation with experimental IC_{50} values, suggesting that the predicted binding modes may be suboptimal and that workflow adjustments are needed. The complex generation step offers the greatest potential for improvement. In our approach, we first sampled conformations of the unbound peptides and used the ten most prominent as starting structures for docking with Vina. However, while these conformations may dominate in unbound peptides, they might not be optimal for interacting with the M^{pro} active site. Furthermore, Vina may not be the most effective tool for exploring the peptides conformational landscape within the binding site. Peptide-docking specialized programs, such as AutoDock CrankPep (ADCP)[39], have been demonstrated to outperform small-molecule docking tools like Vina, especially for peptides with more than 10 amino acids[41].

Our findings further suggest a potential correlation between peptides that bind to the active site via their terminal regions and reduced overall stability and affinity. Thus, another criterion for evaluating docking poses could be the specific region of the peptide that interacts with the active site of M^{pro}.

Nonetheless, our findings provide valuable insights for the rational design of peptide inhibitors targeting M^{pro}. For competitive inhibitors, we observed that Pro is frequently positioned at the S1′ subsite, showing potential for favorable occupancy. Unlike the conserved Gln interactions with the S1 subpocket, as observed to be highly favorable for 3CVLri-2, Pro is not known to be present at the S1′ subsite in any M^{pro} substrate sequence, offering new perspectives. For non-competitive inhibitors, the presence of multiple aromatic residues at specific positions, such as every other residue in the 3CVLri-4 peptides, appears to enhance binding affinity by facilitating the occupation of certain subpockets at the interface of both protomers. Additionally, our MD simulations generally suggest that negatively charged residues should be avoided in both competitive and non-competitive peptide inhibitors.

In summary, the insights obtained from modeling the different interaction modes of M^{pro} in complex with the L- and Dri-peptide inhibitors via MD simulations provide a strong starting point for the rational design of optimized peptide inhibitors.

4. Computer-Aided Peptide Inhibitor Design against SARS-CoV-2 Proteases

The main protease (M^{pro}) and papain-like protease (PL^{pro}) are essential to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), as they process the non-structural proteins nsp1–nsp16, which mediate critical processes in viral replication, making them promising drug targets for our computer-aided peptide inhibitor design (CAPID) study[94]. In our previous work, we investigated the interactions between M^{pro} and peptides with experimentally validated inhibitory activity against the protease (Chapter 3)[65]. Our simulations of the M^{pro}–peptide complexes revealed a high diversity of potential binding modes, complicating the prediction of reliable peptide design strategies.
To address this, we return to first principles by modeling both proteases with their natural substrates, using this information as a foundation for the design of optimized and effective peptide inhibitors. From the polypeptide cleavage site sequences of PL^{pro} and M^{pro}, three and eleven substrate peptides, respectively, can be extracted (Fig. 4.1). Modeling these substrate peptides in complex with their respective proteases provides the basis for the following *in silico* peptide inhibitor design strategies employed in this study:

- **Strategy I:** Conversion of the substrates into their retro-inverso D-enantiomeric (Dri) counterparts to enhance proteolytic stability while preserving binding affinity.
- **Strategy II:** Utilization of a library of M^{pro}–peptide molecular dynamics (MD) simulations with computed binding affinities to guide the design of peptides with optimized binding properties.
- **Strategy III:** Application of computational saturation mutagenesis for efficient screening of potential design candidates.

By exploring these strategies, we aim to establish a CAPID workflow that enables the development of optimized peptide inhibitors targeting SARS-CoV-2 proteases, while also being adaptable to other biological targets.

4.1. Methods

A comprehensive description of the general methods used in this project is provided in Section 2.1. In the following, we outline the project-specific procedures in detail.

Structure Preparation. The crystal structures of M^{pro} (PDB: 7T70) and PL^{pro} (PDB: 6WX4) were prepared using the Protein Preparation Wizard in Maestro (Schrödinger)[224]. Prior to preparation, the M^{pro} structure—a C145A mutant used to stabilize the complex with the nsp4-5 substrate—was reverted to the wild-type C145. The M^{pro} was prepared as homodimer. The substrate sequences were derived from the UniProt database (accession code: P0DTD1) by extracting the cleavage sites from polyprotein 1ab for the non-structural proteins (nsp1–nsp16).

Pose Generation. Peptide–protein complexes were generated using the docking program AutoDock CrankPep (ADCP)[39]. The docking grid was prepared with AutoGridFR (AGFR)[232], which used the co-crystallized ligands' position to set the docking box at the active site of the proteases.
Since ADCP uses peptide sequences rather than structures as ligand input and does not support docking of D-enantiomeric peptides, we addressed this limitation by converting the protein structure to its D-enantiomeric form instead. The D-protein was generated by flipping the coordinate file along the x axis to create a mirror structure as described by Garton et al. (2018)[280]. After docking L-peptides to the D-protein, the resulting complex coordinates were flipped again to obtain the desired L-protein–D-peptide complex.
The substrate peptides of M^{pro} and PL^{pro}, and their Dri-versions, were docked to the active site of the respective receptor using 250 parallel searches, all starting from a coil conformation, and 5 million evaluations per peptide residue. The same settings were applied for the M^{pro} peptide inhibitor docking in Section 4.3. All docking runs, for M^{pro} and PL^{pro}, were performed in triplicate, using different seeds for each replica.
In a consensus docking approach—combining force field-based docking using ADCP, as described above, with artificial intelligence (AI)-based structure prediction—PL^{pro} was predicted in complex with the three L-substrates (p01–p03) of different lengths (6–12 aa) using the AlphaFold 3 (AF3) server[22].
Peptide designs were generated with the mutagenesis wizard of PyMOL (Schrödinger)[240] to incorporate L- and D-residues using a protease–substrate (m01 or p02) complex structure as template. Prior to MD simulations, acetyl (ACE) and N-methyl amide (NME) caps were added to the peptides N- and C-termini, respectively.

Re-scoring. The top-10 M^{pro}–peptide and PL^{pro}–peptide docking solutions from each triplicate ADCP run were re-scored based on distance cutoffs primarily based on the hydrogen bond pattern between the binding site and the substrate (Figs. 4.5A and 4.8A). To ensure correct peptide orientation within the binding site, we defined d_{ori} as the average distance between main-chain atoms of binding-site residues and peptide positions: for M^{pro}, S3/P3 and S2′/P2′, including binding site residues Glu166 and Thr26; for PL^{pro}, S3/P3, S2/P2, and S1/P1, including Gly163, Gly271 and Tyr268.
Additionally, d_{occ} was defined to ensure occupancy of specific binding subsites. For M^{pro}, d_{occ} is the average distance between the side chains of S1 residues (Phe140, His163) and P1-Gln; for PL^{pro}, $d_{occ}$is the average distance between the side chains of S4 residues (Pro248, Thr301, Tyr264, and Tyr273) and P4-Leu. Docking solutions with $d_{ori} < 5.5\,\text{Å}$ (M^{pro}) or $d_{ori} < 4.0\,\text{Å}$ (PL^{pro}) were re-ranked: A new pose was accepted as the best pose if the docking score improved and $d_{occ,new} <$ cutoff, or if $d_{occ,new} < d_{occ,best}$ and $d_{occ,best} \geq$ cutoff. The cutoff values were 3 Å (M^{pro}) and 5 Å (PL^{pro}).

MD Simulations. All MD simulations and analyses were performed with GROMACS version 2021 [51]. The system setup, energy minimization, equilibration in NVT and NPT ensemble, and the production run was conducted as described in Section 2.1.3.3. Each peptide–protein system (M^{pro}: ∼84,000 atoms or PL^{pro}: ∼110,000 atoms) was simulated for 100 ns.
All statistical analyses were conducted starting from 20 ns to consider mostly converged peptide binding modes. To estimate peptide stability and flexibility within the binding site during the simulations, root mean square deviation (RMSD) and root mean square fluctuation (RMSF) were computed using all heavy atoms to account also for side-chain flexibility. Conformational clustering was performed as described by Daura et al. (1999) [283] to obtain the predominant peptide conformations. The binding free energy (ΔG_{bind}) between the protein and peptide was calculated for 10 % of the trajectory snapshots (500 frames of a 100 ns MD simulation) using the molecular mechanics Poisson–Boltzmann surface area (MM/PBSA) method as implemented in gmx_MMPBSA [59]. The temperature was set to 310 K and the ionic strength to 150 mM.
Structural visualization was performed using PyMOL [240]. For computing the electrostatic potential surface (EPS) of the proteases, the adaptive Poisson-Boltzmann solver (APBS) Electrostatics Plugin [288] implemented in PyMOL version 2.5 was employed.

Computational Dynamic Saturation Mutagenesis. The computational dynamic saturation mutagenesis (CDSM) was performed as described in Section 2.2 with the M^{pro}–m01 and PL^{pro}–substrates (p01–p03) complex MD trajectories.

Table 4.1.: Peptide designs purchased for testing against SARS-CoV-2 PL^{pro} and M^{pro}. Lowercase letters denote D-amino acids. Asterisks indicate issues encountered during peptide synthesis.

M^{pro}	Sequence	Solvent	PL^{pro}	Sequence	Solvent
Dri-m01	rfGsqlva	DMSO	Dri-p02	paGGkltf*	water
a05	PRWQPKWY	water	z07	PFRLIGWL	DMSO
a06	IHLQVRVF	DMSO	z08	PFRLGGWL	DMSO
a10	LRWQIVYI	DMSO			
a21	PRWQPQWY	water			
b06	PRWQpKWY	water			
b07	PRWQpQWY	water			
c04	MRLYIWFW	DMSO			
c05	MRLQIWFW	DMSO			
c07	MRLQIVFW	DMSO			

* L-Thr was incorporated instead of the intended D-Thr.

***In vitro* Inhibition Assay.** Protein expression and purification of M^{pro} and PL^{pro} were conducted as described in Section 2.3.1. All peptides (Table 4.1) were purchased from GenScript (Rijswijk, Netherlands) as lyophilized powders with >90 % purity. All peptides feature amidated C-termini and acetylated N-termini. Peptide purity was confirmed by supplier-performed Reverse-phase high-performance liquid chromatography (RP-HPLC) and electrospray ionization mass spectrometry (ESI-MS) analyses (Fig. A2–A14). Peptides were dissolved in water or dimethyl sulfoxide (DMSO) based on solubility. The inhibitory assays were conducted as described in Section 2.3.2. For the initial screening, 50 or 100 µM of the peptides were used. Following initial screening, we selected the most promising peptides and determined their half-maximal inhibitory concentration (IC_{50}) values in inhibition assays across concentration ranges of 0–400 µM and 0–100 µM for a10 and c04, respectively. As a positive control for M^{pro} inhibition, Disulfiram (GenScript, Netherlands, Rijswijk) was used in concentrations of 0–100 µM. All measurements were performed in biological duplicates. The IC_{50} values were computed by fitting a dose-response curve to the initial velocities plotted against the peptide concentrations. In detail, a 4-parameter logistic (4PL) dose-response model—also known as the Hill equation—was fitted to the data using nonlinear least-squares regression (Levenberg-Marquardt algorithm) implemented in SciPy (scipy.optimize.curve_fit)[289]. The model is described by:

$$f(x) = \text{bottom} + \frac{\text{top} - \text{bottom}}{1 + (x/\text{IC}_{50})^n}$$

where x is the inhibitor concentration, bottom and top are the lower and upper plateaus of the response, IC_{50} is the half-maximal inhibitory concentration, and n is the Hill coefficient, defining the steepness of the curve.

4.2. Strategy I: All-D Retro-Inverso Versions of Substrate Peptides

From the PL^{pro} and M^{pro} cleavage site sequences of non-structural polypeptides nsp1–nsp16, we extracted three (p01–p03) and eleven (m01–m11) substrate peptides, respectively. Each peptide comprises twelve amino acids with the LXGG or Q recognition motifs at the center for cleavage by PL^{pro} and M^{pro}, respectively. The peptide length of 12 amino acids was selected as it is short enough to be efficiently processed by the docking engine ADCP[41], while still providing sufficient coverage of the binding site to yield meaningful insights for the design process.

We used the retro-inverso method to generate Dri substrate versions and investigate their binding modes relative to their natural L-counterparts (Fig. 1.3). The amino acid sequences of the L-and Dri-peptides are shown in Fig. 4.1.

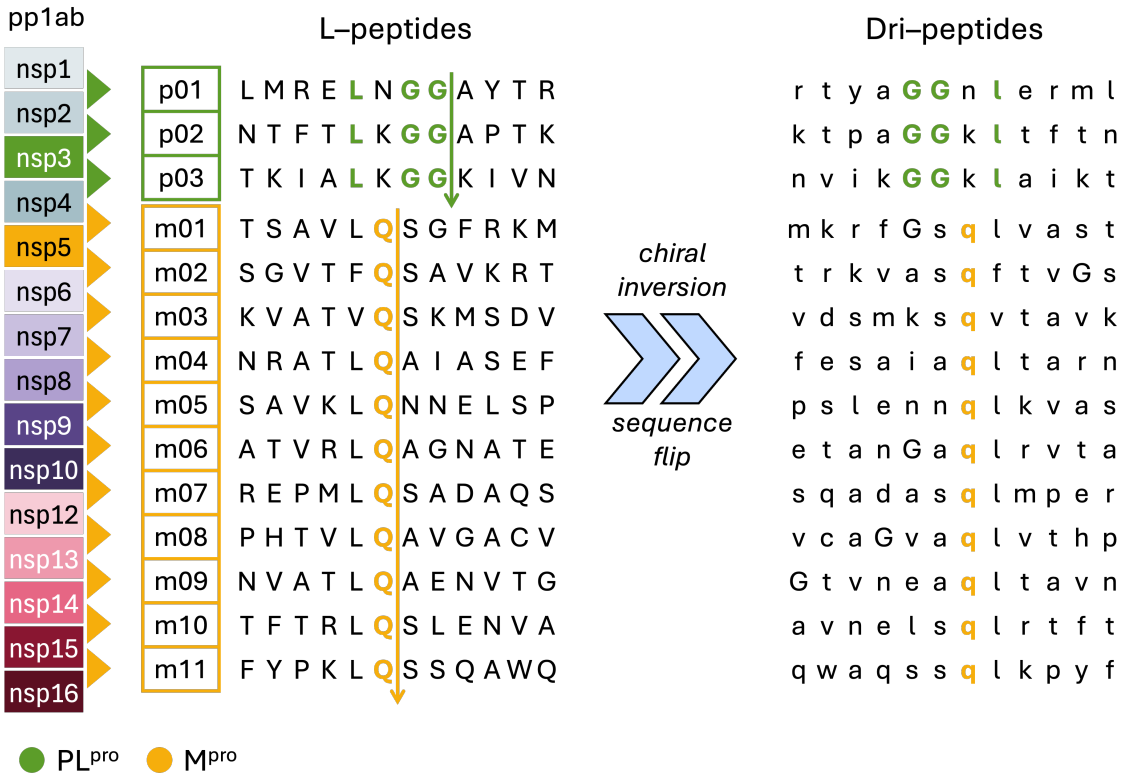


Figure 4.1.: Substrate peptides of PL^{pro} and M^{pro} extracted from polypeptide cleavage sequences. The cleavage site sequences of PL^{pro} and M^{pro} in the polypeptides pp1a and pp1ab were converted into the substrate peptides p01–p03 and m01–m11. Each of these 12-amino-acid peptides contains the respective recognition motif, LXGG for PL^{pro} (green) and Q for M^{pro} (orange), centered within the sequence. Cleavage sites are indicated by vertical arrows. Through chiral inversion and sequence flip (switching N- and C-termini), Dri-versions of the L-substrate peptides of M^{pro} and PL^{pro} are obtained. The data was derived from the UniProt database (accession code: P0DTD1) by extracting the cleavage sites from polyprotein 1ab.

4.2.1. Main Protease

Complex Generation. The L- and Dri-versions of substrates m01–m11, (L)m01–m11 and (Dri)m01–m11, were docked to the active site of one protomer in the M^{pro} dimer (PDB: 7T70) using ADCP. In this context, the flipped sequence and stereochemistry of the Dri-peptides had to be accounted for. To ensure that their binding mode at the M^{pro} active site matched that of the natural L-substrate peptides, the N- and C-termini of the Dri-peptides needed to align with the P′ and P positions, respectively, rather than with the P and P′ positions as in the L-peptides. Thus, we applied a re-scoring strategy to the top-10 docking solutions from triplicate runs to ensure correct peptide orientation within the binding site and proper positioning of the conserved Gln side chain in the S1 pocket. The correct orientation was defined by distance cutoffs between specific peptide positions and binding subsite residues (P3–S3 and P2′–S2′). Prior to re-scoring, all top-ranked docking structures of the L-substrates already exhibited the correct orientation and positioning of the P1 side chain. In this case, re-scoring merely served to select the best top-ranked structure across triplicates. In contrast, none of the Dri-peptides exhibited top-ranked structures that met the criteria, thus re-scoring was essential.

Figures 4.2A and B show the re-scored docking poses of the L- and Dri-peptides. Superimposed structures of (L)m01–m11 reveal a consistent binding mode for positions P4–P1′, which are docked to the respective S4–S1′ subpockets. The remaining peptide positions, P6–P5 and P2′–P6′, exhibit considerably higher flexibility in their positioning

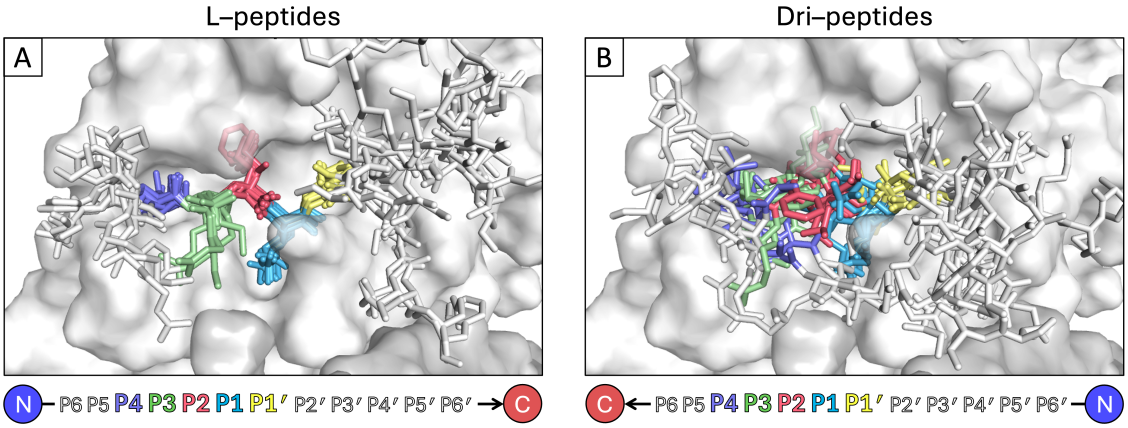


Figure 4.2.: Re-scored docking solutions of L- and Dri-peptides in complex with M^{pro}. The docking poses obtained after re-scoring are shown for (A) L- and (B) Dri-peptides (m01–m11) docked to the active site of M^{pro}. The superimposed peptide poses are displayed as sticks (white), with positions P4 (violet), P3 (green), P2 (red), P1 (blue), and P1′ (yellow) highlighted. In Dri-peptides, the N- and C-termini (blue and red circles) are flipped, extending from the P′ to the P site, whereas in L-peptides, they extend from the P to the P′ site. The M^{pro} dimer surface is illustrated with chain A in white and chain B in grey.

around the active site. This correlates with results from co-crystallization of the substrates (Fig. A23)[290], where these positions exhibit higher conformational variety. These regions are also often not fully resolved by X-ray crystallography, which may indicate high internal structural flexibility. The higher consistency of P4–P1′ poses may be attributed to favorable interactions with the S4–S1′ region, which features more defined cavities compared to the rest of the binding subsites. This is further associated with higher conservation of residues with specific features, such as size and hydrophobicity, to optimally accommodate these S4–S1′ subsites (Fig. 1.5B)[124,125]. In contrast to the L-peptides, the (Dri)m01–m11 poses exhibit significant variability in the positioning of their P4–P1′ region. They further exhibit generally less favorable affinities, with an average of −22.6 kcal/mol compared to −25.6 kcal/mol for (L)m01–m11 (Fig. A24A). This suggests that the docked Dri-peptides may lack favorable interactions critical for the optimal occupation of the binding subsites by the P4–P1′ region.

However, since the receptor was kept rigid during docking, M^{pro} was unable to adapt to the Dri-peptide structures. Thus, to dynamically investigate the interactions between the protease and the peptides, the generated complex models were submitted to MD simulations.

MD Simulations. We performed 100 ns MD simulations with the re-scored (L)m01–m11 and (Dri)m01–m11 docking poses bound to the active site of M^{pro}. Prior to the simulations, the peptide termini were capped with ACE and NME groups to neutralize terminal charges, avoid potential interactions between oppositely charged termini, and better mimic the environment of the native polypeptides from which the sequences were derived.

To estimate the stability and flexibility of the peptides within the binding site, we computed their RMSD and RMSF. The RMSD values observed for the L- and Dri-peptides over the entire trajectory generally indicate greater stability for the L-peptides compared to the Dri-peptides (Fig. 4.3). One Dri-peptide, Dri-m08, even completely dissociates from the binding site after approximately 50 ns. However, most L- and Dri-peptides stabilize their conformations after the initial 20 ns.

RMSF analysis indicates that the primary structural changes occur at the N- and C-termini. While the L-peptides exhibit generally low flexibility, with RMSF values consistently below 2 Å in the P4–P2′ region, the Dri-peptides display significant variability in RMSF across all peptides. Their reduced conformational stability suggests less favorable interactions between Dri-peptides and the binding site, which aligns with the observations from their docking results.

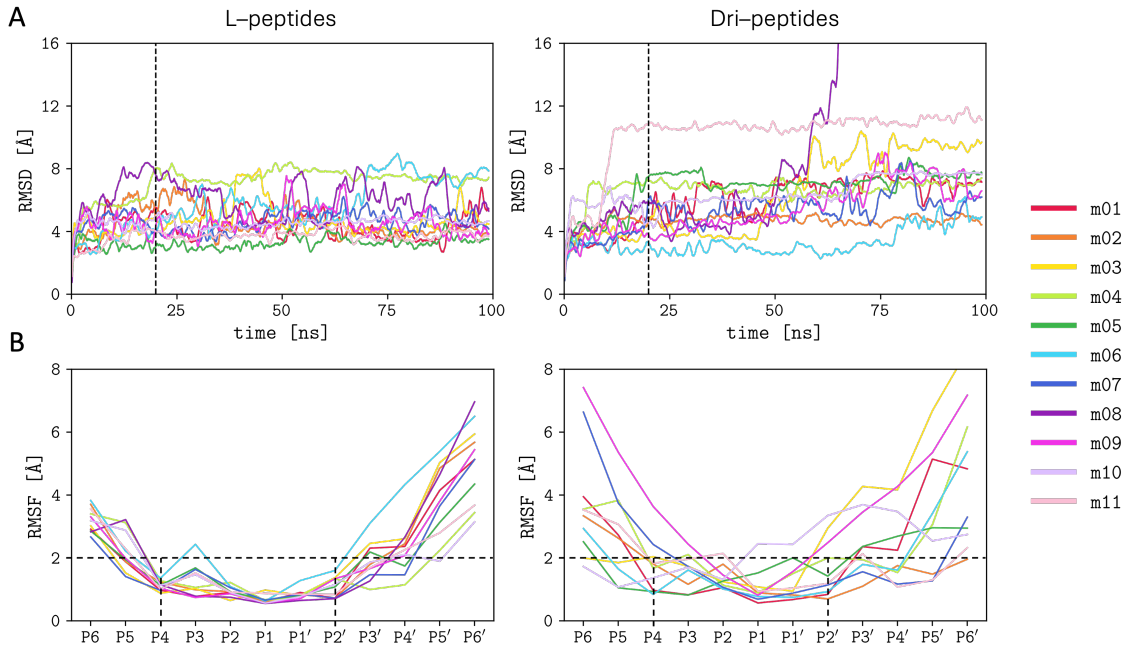


Figure 4.3.: Stability and flexibility of L- and Dri-peptides in complex with M^{pro} during MD simulations. The (A) RMSD and (B) RMSF was computed for heavy atoms of the L- and Dri-substrate peptides (m01–m11), simulated for 100 ns in complex with M^{pro}. (A) The vertical dashed line indicates the time (20 ns), after which most peptide conformations exhibit stable RMSD values. (B) The RMSF is displayed for the peptide positions P6–P6′ with the horizontal dashed line at 2 Å indicating the most rigid P4–P2′ region in the L-peptides.

Hydrogen Bonding. Previous studies have demonstrated that hydrogen bonds play a critical role in substrate binding to the M^{pro} active site, often mediated by the peptides' backbone[74,290]. To elucidate how the L- and Dri-peptides form hydrogen bonds with the protease, we analyzed the occurrence of hydrogen bonds for each peptide residue, distinguishing between mainchain amide nitrogen (MC_N), mainchain carboxy oxygen (MC_O), and variable sidechain groups (Fig. 4.4). For the L-substrate peptides, a distinct hydrogen bonding pattern emerges, with notable interactions at positions P3, P1, P2′, and P4′. Generally, the peptides exhibit frequent hydrogen bonds formed by residues P3, P1, P2′, and P4′ via MC_N, while MC_O-mediated hydrogen bonds are consistently formed at P3, P1, and P2′ throughout the simulation. Furthermore, the only persistent hydrogen bonds formed via sidechains across all peptides are observed for the Gln residue at P1, underscoring its uniquely high conservation in the cleavage site sequences.

In contrast, none of the Dri-peptides fully reproduce the hydrogen bonding pattern observed for the L-peptides. While some Dri-peptides exhibit hydrogen bonds involving MC_O at both P1 and P2′, these interactions never occur concurrently with P3. The most consistent hydrogen bonds are typically formed by the P1 Gln residue, particularly through its

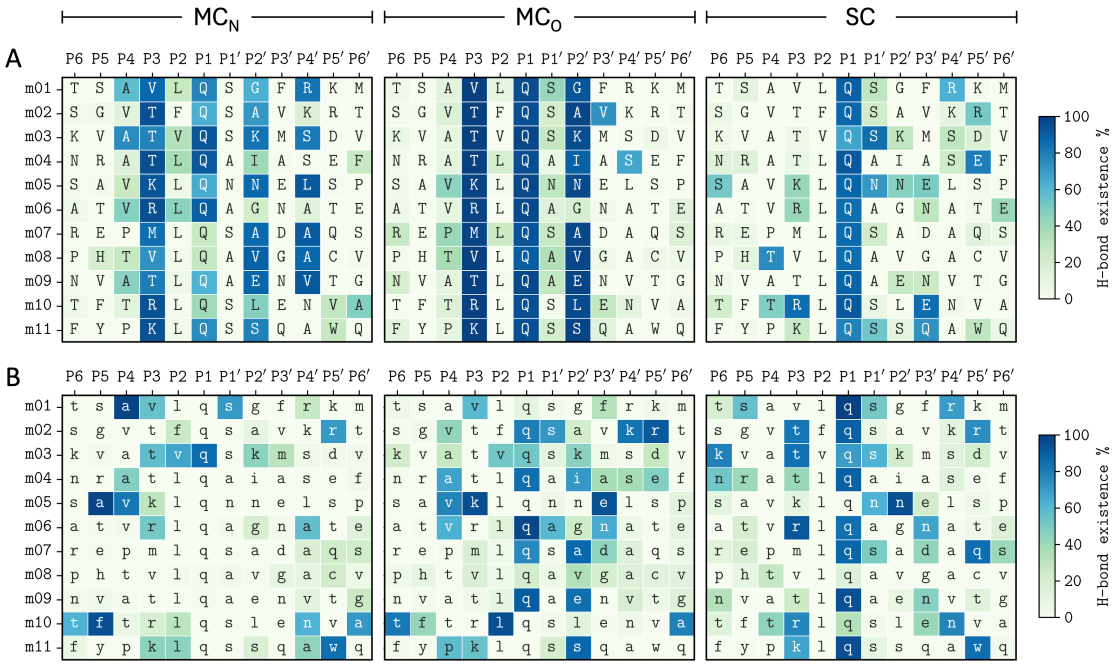


Figure 4.4.: Hydrogen bond existence for L- and Dri-peptides simulated in complex with M^{pro}. The hydrogen bond (H-bond) existence between each (A) L- and (B) Dri-peptides (m01–m11) and M^{pro} is shown for each peptide position (P6-P6′), and is further divided into mainchain (MC) hydrogen donor (nitrogen, N), hydrogen acceptor (oxygen, O) and variable sidechain (SC) hydrogen donors/acceptors. From left to right, the amino acid sequence of each L- and Dri-peptide is displayed from N- to C-terminus and from C- to N-terminus, respectively. The existence is considered 100 % if at least one H-bond was detected for the respective peptide group in all analyzed frames of the MD simulation.

sidechain, whereas hydrogen bonds via MC_N are rarely observed.

To investigate the hydrogen-bond network structurally, we first visualized the hydrogen bonds present in the crystal structure of M^{pro} bound to the co-crystallized nsp4/5 substrate (Fig. 4.5A). The crystal structure reveals typical main-chain hydrogen bonds between the protease and peptide positions P3, P1, P2′, and P4′ (Glu166: P3; His164, Cys145, Gly143: P1; Thr25: P2′; Thr24: P4′), with an additional side-chain interaction between Thr24 and P4′ MC_O. The P1 Gln side chain forms hydrogen bonds with MC_O of Phe140 and the side chains of Glu166 and His163. Additional hydrogen bonds involve Gln69 (side chain: P4′ Arg side chain), Gln189 (side chain: P2 MC_N, MC_O: P5 Ser side chain), Thr190 (MC_O: P4 MC_H), and Gln192 (MC_H: P6 MC_O). However, as residues other than Gln at P1 vary across M^{pro} substrates, hydrogen bonds involving peptide side chains (e.g., P5 Ser and P4′ Arg) are not considered conserved. To compare hydrogen bonds formed in the MD simulations of L-m01 and Dri-m01 with the co-crystallized substrate, we performed cluster analysis for peptide positions P4–P4′ (excluding flexible terminal residues) using an RMSD cutoff of 2.0 Å. The most populated binding poses of L-m01 and Dri-m01,

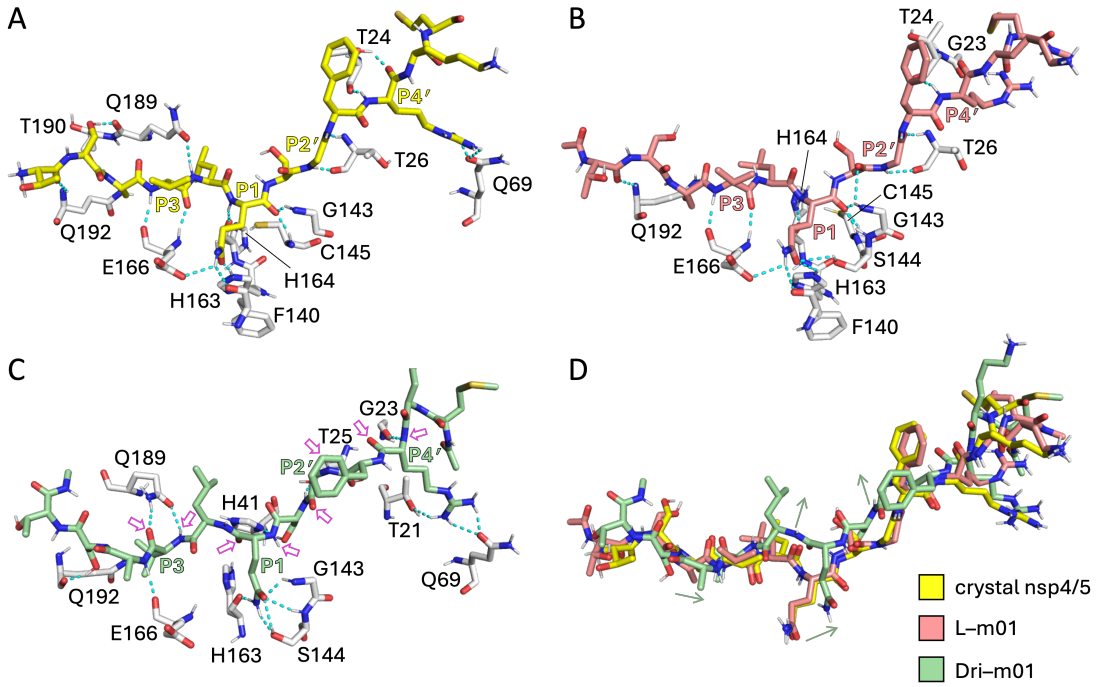


Figure 4.5.: Comparison between co-crystallized substrate peptide nsp4/5, L-m01 and Dri-m01 forming hydrogen bonds with M^{pro}. The structure of (A) the co-crystallized nsp4/5 (PDB: 7T70) and the most populated conformations of the simulations with (B) L-m01 (83 % sampled) and (C) Dri-m01 (74 % sampled) are shown in complex with M^{pro}. The peptide structures are displayed as sticks in yellow (A), salmon (B) and green (C). M^{pro} residues forming hydrogen bonds with the peptides are shown as white sticks, with hydrogen bonds shown as cyan dotted lines. Pink arrows in (C) indicate main-chain hydrogen bond acceptors and donors of P3, P1, P2′ and P3′ that are usually involved in M^{pro} substrate binding. (D) Superposition of the crystal structures of nsp4/5, L-m01, and Dri-m01 with green arrows indicating the structural deviation of Dri-m01 from the L-peptide structures.

with 83 % and 74 % probability, are presented in Fig. 4.5B and C. L-m01 reproduces the hydrogen-bonding pattern of the co-crystallized nsp4/5 peptide, with the same protease residues interacting with the respective main-chain atoms of peptide positions P3, P1, P2′, and P4′. Stabilization of Gln at P1 is similarly achieved, with an additional side-chain interaction from S144. The side-chain hydrogen bonds involving Thr24 and Gln189 are missing in L-m01, likely due to the generally higher flexibility of side chains compared to the protein backbone.

In contrast, most hydrogen bonds with the main chains at positions P3, P1, P2′, and P4′ in Dri-m01 are missing or mediated by different binding site residues. For example, the P3 main chain interacts with the Gln189 side chain instead of the Glu166 main chain. Fig. 4.5D shows that a shift of P4 (green arrow) allows Glu166 to form a hydrogen bond with P4 MC_N, replacing its interaction with P3 MC_N. In contrast, the large backbone shift in the P2–P2′ region failed to restore hydrogen bonding, as evidenced by the rare

occurrence of main-chain hydrogen bonds in this region (Fig. 4.4). Instead, this peptide region moved away from the binding site, likely due to unfavorable interactions from the amide bond flip. The P1-Gln side chain is also shifted, which resulted in the loss of hydrogen bonds with Phe140 and Glu166, however, it remained consistent interactions with Gly143, Ser144, and His163 instead. Thus, the distinct behavior of L- and Dri-peptides can be explained by the reversed N- and C-termini and amide bond orientations in Dri-peptides. Flipped positions of the MC_N and MC_O atoms in the Dri-peptides lead to unfavorable interactions with the binding site and disrupt the conserved hydrogen-bonding network that stabilizes the substrate binding to M^{pro}.

Binding Affinity. To assess the impact of Dri-conversion on individual peptide residues contributing to the binding affinity of M^{pro}–peptide complexes, we performed per-residue ΔG_{bind} decomposition using the MM/PBSA method. As shown in Fig. 4.6A, the P4–P1 positions, particularly P1, contribute most favorably to the ΔG_{bind} across all L-peptides, while the P′ site exhibits less favorable or even unfavorable interactions. This underscores the critical role of the P4–P1 region, including the highly conserved Gln at P1, in substrate binding to the M^{pro} active site. In contrast, although the Dri-peptides also display mostly favorable ΔG_{bind} values at their P2 and P1 positions, their overall per-residue ΔG_{bind} decomposition profile lacks the clear pattern observed for the L-peptides and instead shows high variability (Fig. 4.6B). This discrepancy is further reflected in their total ΔG_{bind} values, which range from −28.7 to −64.2 kcal/mol for Dri-peptides, compared to −62.5 to −77.5 kcal/mol for L-peptides (Fig. A24B). The superior binding affinity of L-peptides

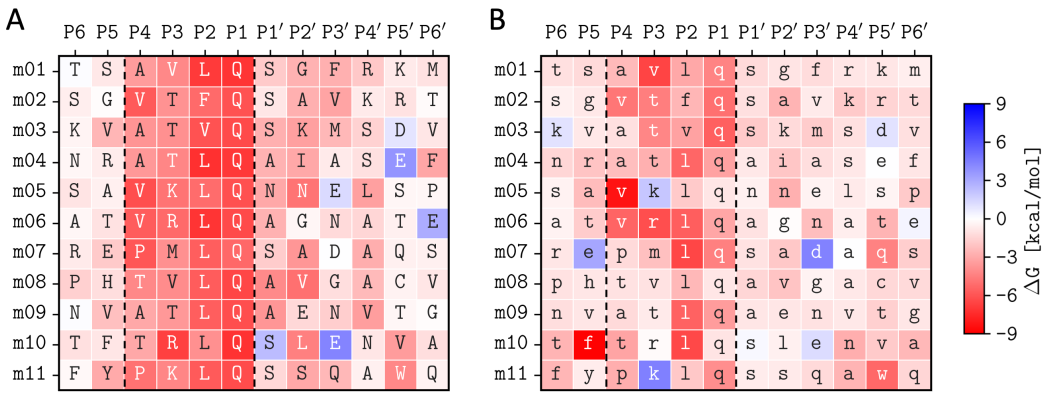


Figure 4.6.: Per-residue binding free energy decomposition of L- and Dri-peptides simulated in complex with M^{pro}. Per-residue decomposition of the binding free energy (ΔG_{bind}) for each position (P6–P6′) of the (A) L- and (B) Dri-peptides (m01–m11) MD simulations in complex with M^{pro}. From left to right, the amino acid sequence of each L- and Dri-peptide is displayed from N- to C-terminus and from C- to N-terminus, respectively. Vertical dashed lines mark the most favorably binding peptide region P4–P1.

is likely due to stronger interactions with M^{pro} facilitated by the stable hydrogen bonding via the mainchain of the L-peptides.
To summarize, our MD simulations of M^{pro} in complex with substrate peptides and their Dri-versions suggest that the Dri-peptides interact less favourably, potentially due to a disrupted hydrogen bond pattern caused by the inverted backbone of these peptides. For this reason, we will continue our peptide inhibitor design process using the L-peptides as a basis rather than the Dri-peptides. Nonetheless, Dri-m01 will be assayed for inhibitory activity against M^{pro} in Section 4.5.1 to evaluate the Dri-method and our computational predictions.

4.2.2. Papain-like Protease

Complex Generation. For structure-based docking using the ADCP software, the target structure must be selected carefully. For example, Fig. A26 shows two crystal structures of PL^{pro}: one co-crystallized with the inhibitor XR8-89 (PDB: 7LBR) and the other bound to the covalent inhibitor VIR251 (PDB: 6WX4). VIR251 is a peptidomimetic containing the GG part of the recognition motif LXGG from the natural substrates. Thus, it binds to the respective tunnel near the catalytic site, as indicated by a cyan arrow in Fig. A26B. This tunnel is closed in the other crystal structure because inhibitor XR8-89 does not engage this tunnel subsite. To enable successful docking of the L- and Dri-peptides, this tunnel must be accessible. Therefore, the PL^{pro} structure in complex with inhibitor VIR251 (PDB: 6WX4) was chosen for subsequent docking.
Docking of the 12 amino acid substrate peptides to the PL^{pro} active site (PDB: 6WX4) using ADCP yielded ambiguous solutions, with some top-ranked structures showing a solvent-exposed Leu residue at P4. This is contrary to expectations, as this residue is highly conserved in the LXGG binding motif and is expected to form specific interactions with the protease. Therefore, we decided to involve an additional tool, AF3, combining the physics-based docking method of ADCP with an AI-based structure prediction method to improve reliability of the results.
The amino acid sequences of PL^{pro} and each substrate peptide, (L)p01–p03, were submitted to AF3. We tested different lengths of the substrates, ranging from six to twelve amino acids, with each having the LXGG motif at the center. The predicted complex structures were assessed by confidence metrics such as the interface predicted Template Modeling score (ipTM). The highest-confidence pose was obtained for the 8 amino acid L-p02, with an ipTM of 0.95 (Fig. A27C), showing Leu at P4 occupying a hollow of the substrate binding site. Based on these insights, we installed a re-scoring procedure for docking with ADCP, to ensure the positioning of the P4-Leu residue at the binding subsite S4, as well as the correct orientation of L- and Dri-peptides within the tunnel via distance cutoffs for the

positions P1, P2 and P3. All re-scored docking solutions for L- and Dri-peptides against PL^{pro} using ADCP, as well as the solutions for the PL^{pro}–substrate complexes predicted by AF3 are presented in Fig. A27.

During ADCP docking, top-ranked poses frequently placed the N-terminus in a negatively charged cavity of the substrate binding site (Fig. A25B)—an unrealistic pose that also complicates the addition of terminal capping groups. Given the high confidence in the AF3 prediction of the PL^{pro}–p02 complex, we used it as a template to manually generate the L-p01 and L-p03 complexes with PL^{pro}. This issue did not arise for Dri-peptides, likely because their swapped N- and C-termini would result in unfavorable C-terminal interaction with this negatively charged cavity. Instead, we observed that the conserved Leu at P4 of the Dri-peptides is projecting into this pocket (Fig. A27D). To enable direct comparison with the 8-residue L-peptides, we also selected 8-residue Dri-peptide docking poses after re-scoring for subsequent MD simulations.

MD Simulations. We performed 100 ns MD simulations with the (L)p01–p03 and (Dri)p01–p03 bound to the active site of PL^{pro}. Prior to the simulations, the peptide termini were capped with ACE and NME groups.

To assess peptide stability and flexibility during the MD simulations, we computed the RMSD and RMSF. The Dri-peptides deviated more from the starting structure at the beginning of the simulation, reaching RMSD values of 4–8 Å, whereas the L-peptides reached RMSD values of 2–5 Å (Fig. A28A). After approximately 20 ns, all simulations converged, indicating stable binding for both L- and Dri-peptides. The RMSF values further support this observation, showing high rigidity (<2 Å) for positions P4–P2′ in all L-peptides. Similarly, the Dri-peptides exhibit low flexibility at positions P4–P1, corresponding to the LXGG binding motif.

To analyze the hydrogen bond pattern of the L- and Dri-peptides with the protease, we calculated hydrogen bond existence for each peptide residue, distinguishing contributions from main-chain amide nitrogen (MC_N), main-chain carboxy oxygen (MC_O), and variable side chain groups. Fig. 4.7A reveals that in the L-peptides, consistent interactions are mediated primarily by the main chain at positions P3–P1′, while side-chain interactions are rare and less consistent. In the Dri-peptides, this main-chain mediated hydrogen bond pattern is disrupted, resulting in a loss of interaction consistency in the P3–P1′ region; however, hydrogen bond persistence increases at other positions, as seen for MC_N of P5 across (Dri)p01–p03.

To visualize hydrogen bonding between the peptides and PL^{pro}, we performed cluster analysis using an RMSD cutoff of 2.5 Å. The representative structures of the most populated binding poses of L-p02 (75 % sampled) and Dri-p02 (53 % sampled) are shown

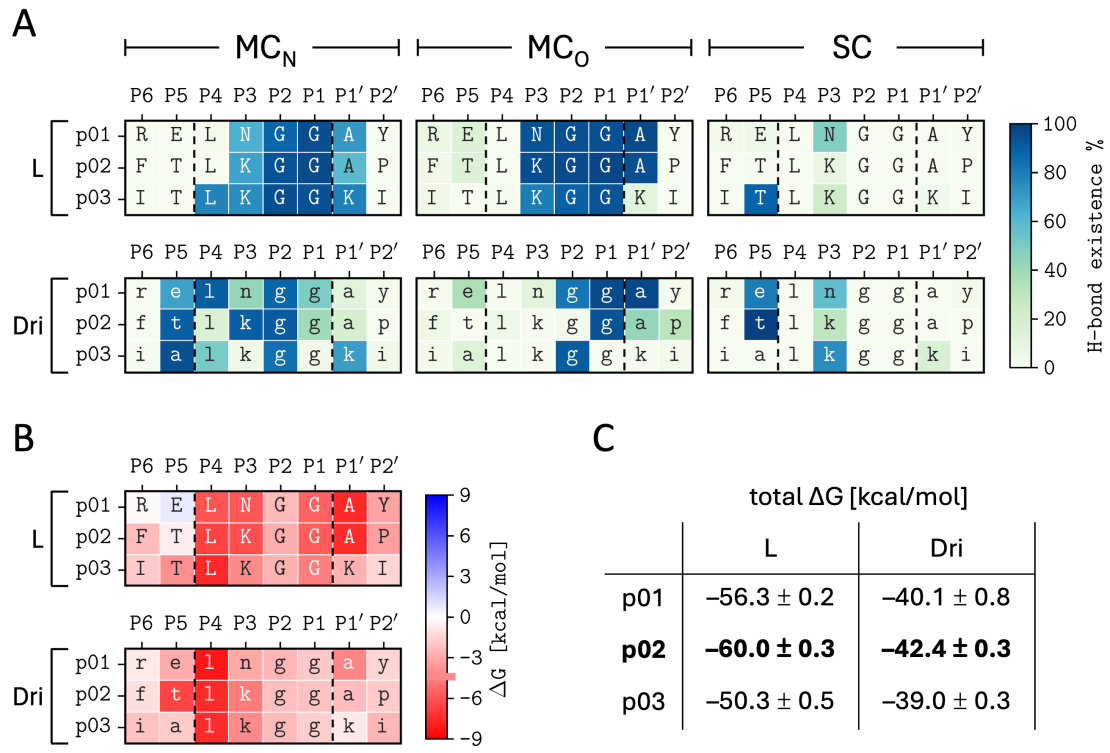


Figure 4.7.: Hydrogen bond analysis and binding free energy calculations of L- and Dri-peptides simulated in complex with PL^{pro}. The 100 ns MD simulations of L- and Dri-peptides (p01–p03) in complex with PL^{pro} were analyzed for hydrogen bonding patterns and binding free energy (ΔG_{bind}). (A) The hydrogen bond (H-bond) existence is shown for each peptide position (P6-P2′), and is further divided into mainchain (MC) hydrogen donor (nitrogen, N), hydrogen acceptor (oxygen, O) and variable sidechain (SC) hydrogen donors/acceptors. The existence is considered 100 % if at least one H-bond was detected for the respective peptide group in all analyzed frames of the MD simulation. (B) The per-residue ΔG_{bind} decomposition is displayed for P6–P2′. Vertical dashed lines mark the binding motif LXGG (P4–P1). All amino acid sequences are presented from left to right, with L-peptides oriented from N- to C-terminus and Dri-peptides from C- to N-terminus. (C) The total ΔG_{bind} was calculated for each PL^{pro}–peptide complex simulation.

in Fig. 4.8A and B, respectively. At positions P3–P1′ of L-p02, both MC_N and MC_O form hydrogen bonds with PL^{pro} binding site residues. Most binding site residues form backbone-mediated hydrogen bonds, including Tyr268 with MC_N of P3, Gly163 with both MC_N and MC_O of P2, and Gly271 with MC_O of P3, MC_N of P1, and MC_N of P1′. Additionally, MC_O of P3 interacts with the side chain of Tyr264. Further side-chain-mediated hydrogen bonds are formed by MC_O of P1 with Trp106, which has been described as an oxyanion hole-stabilizing residue in SARS-CoV PL^{pro} [291], indicating a productive substrate peptide pose. MC_O of P1′ also interacts with the side chain of Trp106 and His272, a member of the catalytic triad. Substrate peptides p01 and p03 exhibit the same hydrogen bond pattern (Fig. A29), although p03 deviates slightly at P1′, likely due to the Lys residue at this position instead of Ala present in p01 and p02.

Fig. 4.5C shows that Dri-p02 is shifted towards the P-site, such that MC_N of positions P3–P1′ in Dri-peptides aligns with those of the L-peptides. This enables Dri-peptides to retain some hydrogen bonds present in the L-peptides, including MC_N-mediated interactions between P3, P2, and P1 with Tyr268, Gly163, and Gly271, respectively. However, the flipped amide bonds in Dri-peptides disrupt the formation of other hydrogen bonds, such as MC_O-mediated interactions at P3 and P2 due to the shifted and reoriented carbonyl groups.

To assess the contribution of individual peptide residues to the binding affinity of PL^{pro}–peptide complexes, we performed per-residue ΔG_{bind} decomposition with the MM/PBSA method. Fig. 4.7B shows that in both L- and Dri-peptides, Leu at P4 contributes substantially to the ΔG_{bind}. In contrast to Dri-peptides, the L-peptides exhibit also highly favorable ΔG_{bind} values at P3 and P1, and—in the case of p01 and p02—for Ala at P1′. Interestingly, although the LXGG binding motif (spanning P4–P1) is conserved, the Gly at P2 in L-peptides exhibits relatively modest affinity ($\Delta G_{bind} > -3$ kcal/mol) compared to other residues in P4–P1 (< -4 kcal/mol). This suggests that the conservation of the GG motif occupying the narrow tunnel binding subsite of PL^{pro} may rather reflect steric

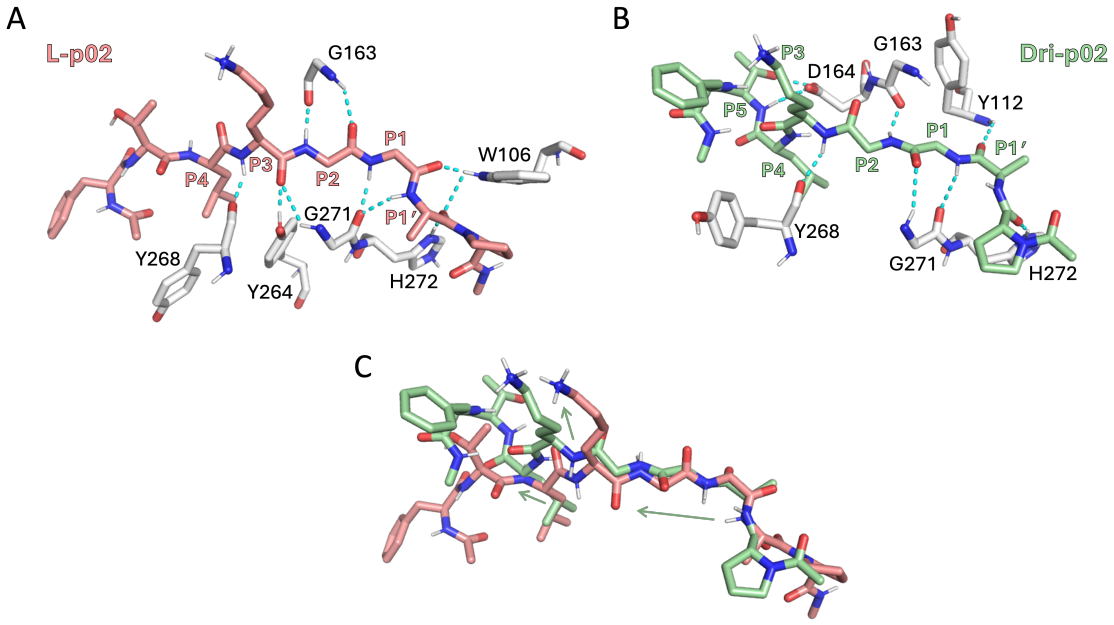


Figure 4.8.: Comparison between L- and Dri-peptide p02 forming hydrogen bonds with PL^{pro}. The structure of the most populated conformation of the simulations with (A) L-p02 (75 % sampled) and (B) Dri-p02 (53 % sampled) are shown in complex with PL^{pro}. The peptide structures are displayed as sticks in (A) salmon and (B) green. PL^{pro} residues forming hydrogen bonds with the peptides are shown as white sticks, with hydrogen bonds displayed as cyan dotted lines. (C) Superposition of L-p02 and Dri-p02 with green arrows indicating the structural deviation of Dri-p02 from the L-peptide structure.

selectivity than binding affinity.
The total ΔG_{bind} values for the PL^{pro}–peptide complexes (Fig. 4.7C) show that L-peptides exhibit affinities >10 kcal/mol more favorable than those of Dri-peptides, with L-p02 exhibiting the highest affinity overall (-60.0 ± 0.3 kcal/mol). Although Dri-peptides exhibit lower affinities than L-substrates, Dri-p02 will be tested for inhibitory activity against PL^{pro} in Section 4.5.3 to validate our computational predictions.

4.3. Strategy II: Peptide–Protease Complex Library

In Section 4.2, we simulated all substrates of M^{pro} extracted from the pp1a and pp1ab cleavage sites. To further expand the library of M^{pro}–peptide MD simulations, the next step is to include peptides that have been experimentally validated as M^{pro} inhibitors, providing additional insights to guide the design of effective peptide inhibitors against M^{pro}.

4.3.1. Expanding the Library with M^{pro} Peptide Inhibitors

Inhibitors p12–p16. In a computational peptide design approach against the M^{pro}, Chan et al. (2021)[74] have developed the inhibitors p12–p16 inspired by M^{pro}–substrate MD simulations. Their study demonstrated inhibitory activity against M^{pro} for p12, p13, p15, and p16.
We docked the peptides p12–p16 against M^{pro} using ADCP, followed by re-scoring based on the distance cutoffs described in Section 4.1, and performed 100 ns MD simulations of the top-ranked complex models.
The peptides display stable RMSD and RMSF values, indicating stable binding modes with M^{pro} (Fig. A30). Per-residue ΔG_{bind} decomposition was computed using the MM/PBSA method, and the results for p12–p16 are displayed as a color map in Fig. 4.9A. The results indicate improved binding affinities for the peptides P′ region compared to the substrates m01–m11. However, the presence of Lys in p15 and p16 appears to be unfavorable based on our simulations. This contradicts the authors design predictions, which were based on computational saturation mutagenesis analysis of the M^{pro}–substrate models using the BUDE_SM algorithm[75]. However, their study does not mention any MD simulations of M^{pro} in complex with p15 or p16. Therefore, a direct comparison of these results is not possible.

Competitive 3CVL and 3CVLri. In Chapter 3, we modeled M^{pro} with the peptides 3CVL and their Dri-variants, 3CVLri, whose inhibitory activity was validated in *in vitro* assays (Table 3.1)[65]. The simulations indicated that the peptides adopted various binding

modes that were, however, rarely similar to those of the substrate peptides. Given that ADCP has proven to be an effective docking engine also for larger peptides[41], we used this program to repeat the docking process for the competitive 3CVL(ri) peptides against the active site of M^{pro}. Subsequently, the top-3 ranked structures (run1–run3) of 3CVL-2, 3CVL-7, 3CVLri-2, and 3CVLri-7 in complex with M^{pro} were simulated for 200 ns.
Fig. A31 illustrates the per-residue ΔG_{bind} decomposition alongside the RMSF for each peptide residue, with the P and P′ positions shown to the left and right side of the pink line. For the L-peptides, we expect the P-region to be positioned at the N-terminus (left) and the P′-region at the C-terminus (right). To adopt the same binding mode as the L-peptides, the orientation of the Dri-peptides must be reversed, with the P′ and P sites at the N- and C-termini, respectively.
The peptides 3CVL-2, 3CVL-7, and 3CVLri-7 demonstrated binding to the subsites of the M^{pro} active site in the expected orientation in at least one run. However, during docking, the peptides were often positioned in the reversed orientation or exhibited binding modes that completely deviated from the typical substrate binding pattern extending from S to S′ subsites. Only 3CVL-7 consistently displayed expected orientations across all three MD runs. Furthermore, three neighbouring residues of 3CVL-7 were found to bind to the S1 site each in one of the three MD runs. Notably, all three runs exhibited relatively low RMSF values for positions surrounding the P1 site, indicating stable binding to the active site of M^{pro}. Interestingly, the binding mode of 3CVL-7 in run 3 matches that of its Dri-version in run 1. Both also indicate low total ΔG_{bind} values of -73.5 and -67.0 kcal/mol.
In contrast, 3CVLri-2 does not adopt the binding mode of its L-enantiomeric counterpart. In a single MD trajectory where its conformation resembles the substrate binding pattern, the N-terminus occupies the P-region, resulting in binding as a conventional D-peptide rather than the expected retro-inverso mode. However, with the D-enantiomeric Ile exhibiting highly favorable ΔG_{bind} values at the P1 position, this binding mode could be interesting for peptide design.
Overall, finding an optimal binding pose for the 3CVL(ri) peptides proved again challenging due to their long length of 16–17 amino acids, which allows for numerous binding possibilities to M^{pro}. However, using ADCP as the docking program, we were able to find binding poses that resemble those of the substrates, making these peptides a valuable addition to our M^{pro}–peptide library for peptide inhibitor design. The best-performing MD simulations, with the 3CVL(ri)-peptide poses occupying the subsites S6–S6′ of the M^{pro} active site, were considered for the peptide design process, with their per-residue ΔG_{bind} decomposition shown in Fig. 4.9A.

4.3.2. Design of Novel M^{pro} Peptide Inhibitors

Based on the ΔG_{bind} values obtained for each P and P′ position of the peptide–protein simulation library, comprising the M^{pro} substrates, the peptide inhibitors p12–p16, and the most promising 3CVL(ri) binding modes, we designed a total of 21 L-peptides (a-designs) and 7 L/D-mixed peptides (b-designs). The peptide designs were generated by combining residues with the most favorable ΔG_{bind} values from the library, occasionally incorporating random modifications, such as D-amino acids in the b-designs, and reducing the peptide length to eight amino acids spanning P4–P4′. The structures of the peptide designs were obtained through modification of the m01 substrate peptide, followed by 100 ns MD simulations of the M^{pro}–design complexes. The a-designs remain generally stable during the 100 ns simulations, whereas the b-designs are more flexible, indicating that individual D-amino acids may introduce unstable interactions A32.

The per-residue ΔG_{bind} decomposition was computed for all a- and b-designs using the MM/PBSA method. The a-designs generally indicate a considerable enhancement in binding affinity, particularly in the P′ region (Fig. 4.9B). The amino acids that proved most favorable at certain peptide positions include Leu and Ile at P4, Arg at P3, Trp and Leu at P2, Trp and Gln at P1, and Pro at P1′. It appears that excluding charged residues (e.g. Lys, Glu and Asp) from positions near the catalytic site, while generally introducing mostly non-polar residues, yields highly favorable interactions. This aligns with the EPS of the M^{pro} substrate binding site, revealing a minimally charged surface with only a slight negative charge, which explains the unfavorable binding affinity of negatively charged peptide residues (Fig. A25A). In contrast, the spacious, moderately more negatively charged S3 subsite explains the favorable ΔG_{bind} values for Arg at peptide position P3.

The b-designs are mostly derived from the a-designs, while incorporating up to four D-enantiomeric amino acids. Notably, an increasing number of D-residues in the peptide design correlates with progressively more unfavorable ΔG_{bind} values at the respective peptide positions (Fig. 4.9B).

Although the total ΔG_{bind} values exhibit some variability across both peptide design types, the majority demonstrate binding affinities comparable to or exceeding those of the substrate peptides m01–m11 and the peptide inhibitors p12–p16 (Fig. 4.9D). This observation is particularly notable given the reduced length of the designed peptides.

For *in vitro* validation of inhibitory activity against M^{pro} (Section 4.5.1), the most promising candidates were selected. This involves the designs a05 ($\Delta G_{bind} = -86.5 \pm 0.5$ kcal/mol), a06 (-81.1 ± 0.4), a10 (-86.0 ± 0.3), and a21 (-77.7 ± 0.4), which exhibit exceptional binding affinities. Additionally, b06 (-68.3 ± 0.7) and b07 (-81.4 ± 0.4) were included due to their structural identity to the a05 and a21 designs, with the sole modification of a D-amino acid at position P1′.

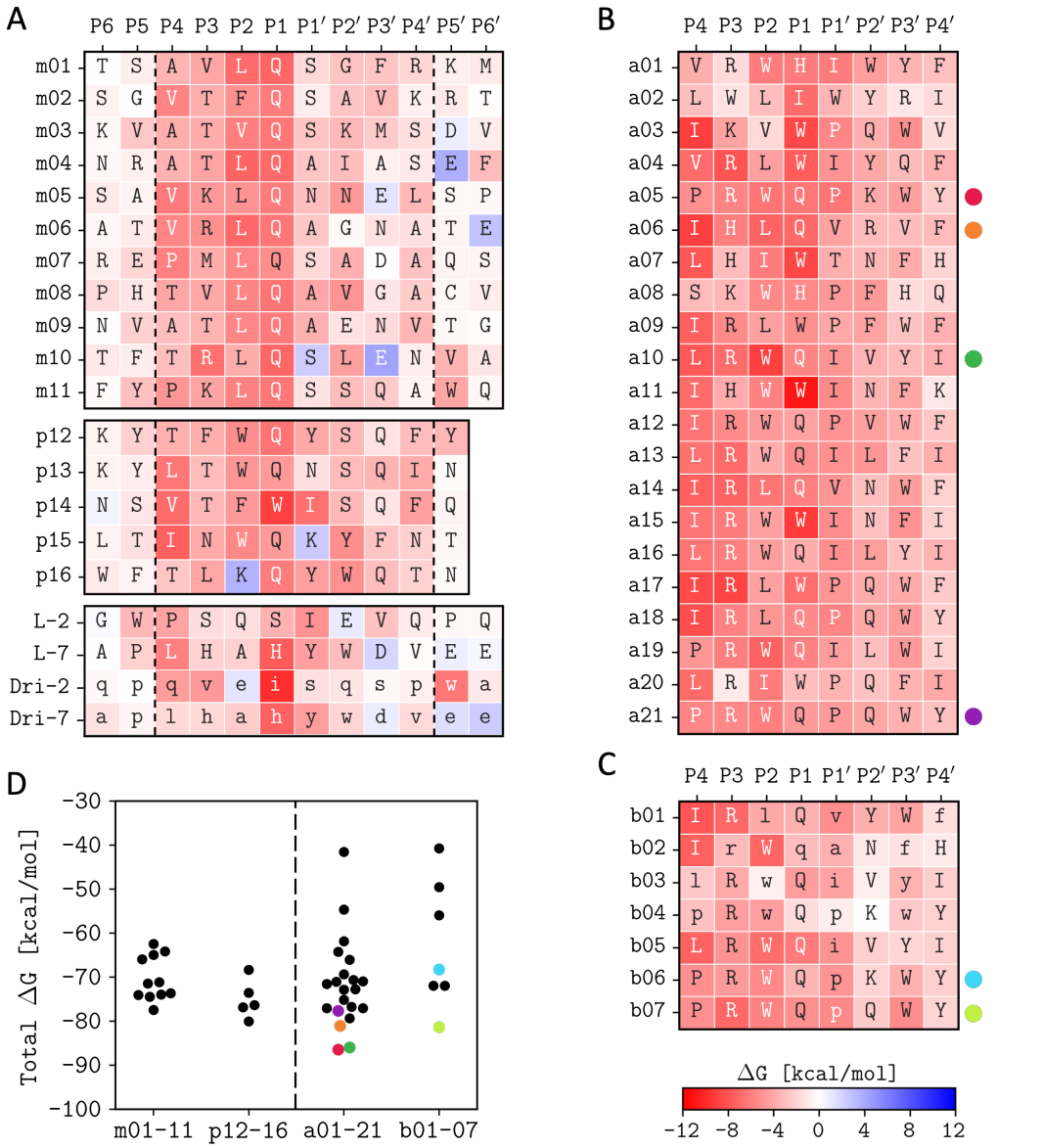


Figure 4.9.: Peptide design process based on binding affinities of M^{pro}–peptide library MD simulations. The per-residue binding free energy decomposition was computed for each peptide position (P6–P6′) of M^{pro} simulated in complex with: (A) the peptide library, including the substrates (L)m01–m11, the inhibitors p12–p16[74], and the L- and Dri-peptides, L-2 (3CVL-2, run1), L-7 (3CVL-7, run3), Dri-2 (3CVLri-2, run1), and Dri-7 (3CVLri-7, run1)[65]; (B) L-peptide designs (a-designs); and (C) mixed designs containing L- and D-amino acids (b-designs). The vertical dashed lines in (A) mark the P4–P4′ region, corresponding to the 8-residue peptide designs (B, C). All amino acid sequences are displayed from N- to C-terminus, except for Dri-7, which is displayed from C- to N-terminus. In all color plots, the ΔG_{bind} values range from −12 (red) to +12 kcal/mol (blue). (D) The total ΔG_{bind} values are shown for comparison between complexes of M^{pro} with the peptide library (left) and peptide designs (right). Designs in (B) and (C) selected for inhibitory assays in the wet lab are marked by colored dots and are also highlighted in the same color in plot (D).

4.4. Strategy III: Computational Dynamic Saturation Mutagenesis

To optimize the peptide design process by reducing the computational time associated with large peptide–protein simulation libraries while simultaneously maximizing the information obtained per peptide position, we developed an automated protocol for *in silico* saturation mutagenesis. Here, we aimed to provide peptide design suggestions based on differences in binding affinity following systematic substitution of each residue in a peptide bound to a receptor. To ensure applicability to peptide design, we designed the protocol to account for the flexibility of peptide–protein interactions. Therefore, the protocol was implemented to be compatible with GROMACS, using a structural ensemble extracted directly from an input trajectory. Fig. 4.10 illustrates the procedure of the Computational Dynamic Saturation Mutagenesis (CDSM) program. First, the user specifies the number of frames (n) to be extracted from the input trajectory of a peptide–protein complex MD simulation. Each peptide residue is then individually mutated to each of the other 19 standard proteinogenic amino acids (excluding the parent amino acid), while the rest of the peptide sequence remains unchanged. Following mutation, a short MD simulation is performed to allow both the protein and peptide to adapt to the substitution and resolve potential steric clashes, as exemplarily shown in Fig. 4.10 for the mutation from alanine to the bulkier tryptophan. Finally, the change in binding affinity between parent and mutant peptides is calculated as the difference in per-residue ΔG_{bind} decomposition using the molecular mechanics generalized Born surface area (MM/GBSA) method for each peptide residue and visualized as a color map.

4.4.1. M^{pro} Inhibitor Design.

For peptide inhibitor design against M^{pro}, we employed CDSM on the 100 ns MD simulation of M^{pro} in complex with the substrate peptide m01, randomly selected from Section 4.2. To assess reproducibility, we performed the protocol in triplicates, each using five snapshots from the trajectory. Fig. 4.11A shows the $\Delta\Delta G$ values for each peptide residue and substitution as the mean of triplicates, alongside color maps displaying the respective standard error of the mean (SEM) derived via error propagation and the standard deviation (SD) across replicates. Given that SD values are predominantly below 0.8 kcal/mol—exhibiting only occasional outliers with up to 1.2 kcal/mol—the results can be considered reproducible, thereby underscoring the stability of the CDSM protocol (Fig. 4.11C).
To evaluate the reliability of the $\Delta\Delta G$ predictions, the SEM was determined across all snapshot simulations and replica runs via error propagation (Fig. 4.11B). This represen-

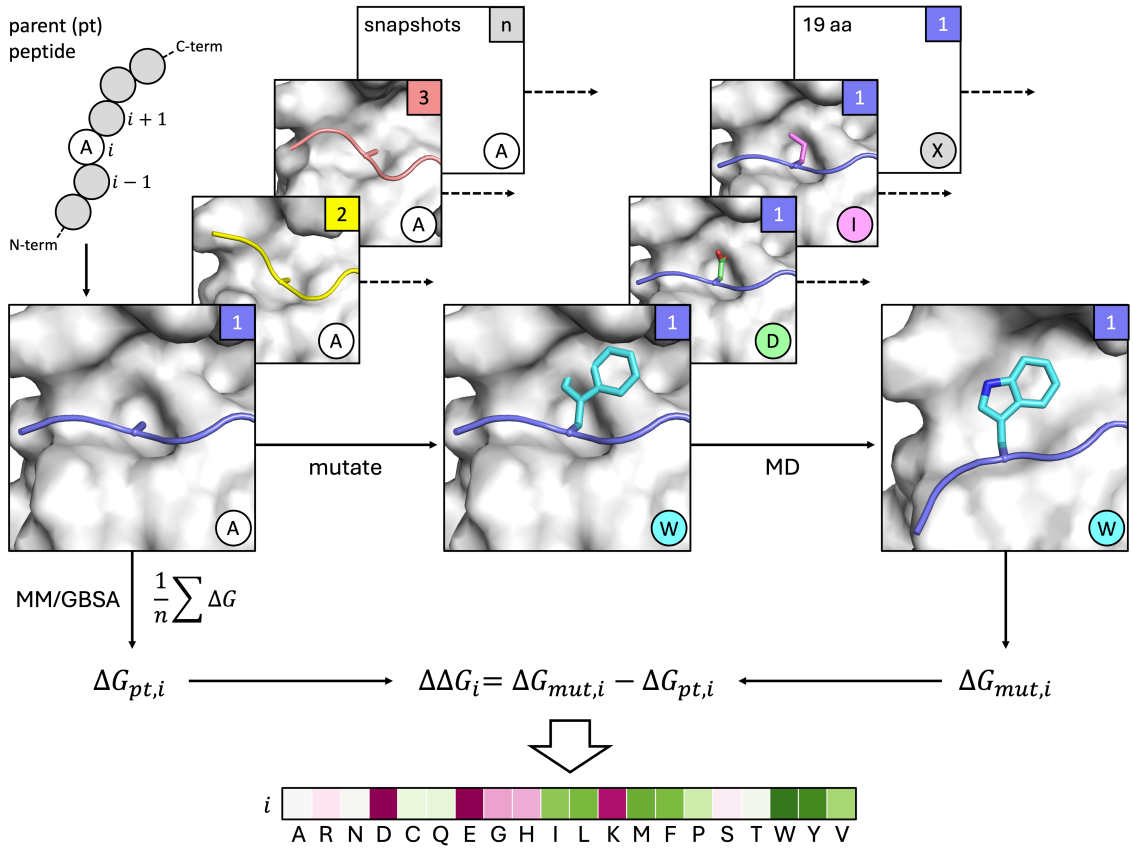


Figure 4.10.: Computational Dynamic Saturation Mutagenesis protocol for peptide design involving peptide–protein MD simulations. A user-defined number (n) of snapshots is extracted from the MD trajectory of a peptide–protein complex. Each peptide position (i) is then systematically mutated to all other 19 standard proteinogenic amino acids (aa) for each snapshot, followed by a short MD simulation to relax the peptide–protein complex. In each frame, the amino acid types are indicated by their one-letter code in the bottom-right corner, and the snapshot number is displayed in the top-right corner. The complete process is exemplified for the substitution of the parent Ala residue with Trp in snapshot number 1. The final structures from all snapshot simulations are then combined and used for per-residue ΔG_{bind} decomposition. The change in per-residue binding free energy decomposition upon mutation ($\Delta\Delta G_i$) for each position i is calculated as the difference between the mutant (mut) and parent (pt) ΔG_{bind} values ($\Delta\Delta G_i = \Delta G_{mut,i} - \Delta G_{pt,i}$). Finally, the $\Delta\Delta G$ values are visualized in a color map to highlight favorable (green) and unfavorable (pink) mutations at each peptide position i. In total, peptide length $\times$ n frames $\times$ 19 aa MD simulations are performed. The structures of peptide and protein are represented as cartoon and surface view, respectively. Peptide residue i is shown as sticks.

tation informs the reader about the relative trustworthiness of the predictions. The SEM plot reveals high uncertainty, particularly for charged residues such as Arg, Asp, Glu, and Lys. High SEM values indicate that the corresponding $\Delta\Delta G$ values must be interpreted carefully.

With this in mind, the $\Delta\Delta G$ results can now be analyzed and transformed into new peptide designs. The plot reveals a trend already observed in the simulation of the peptide–protein

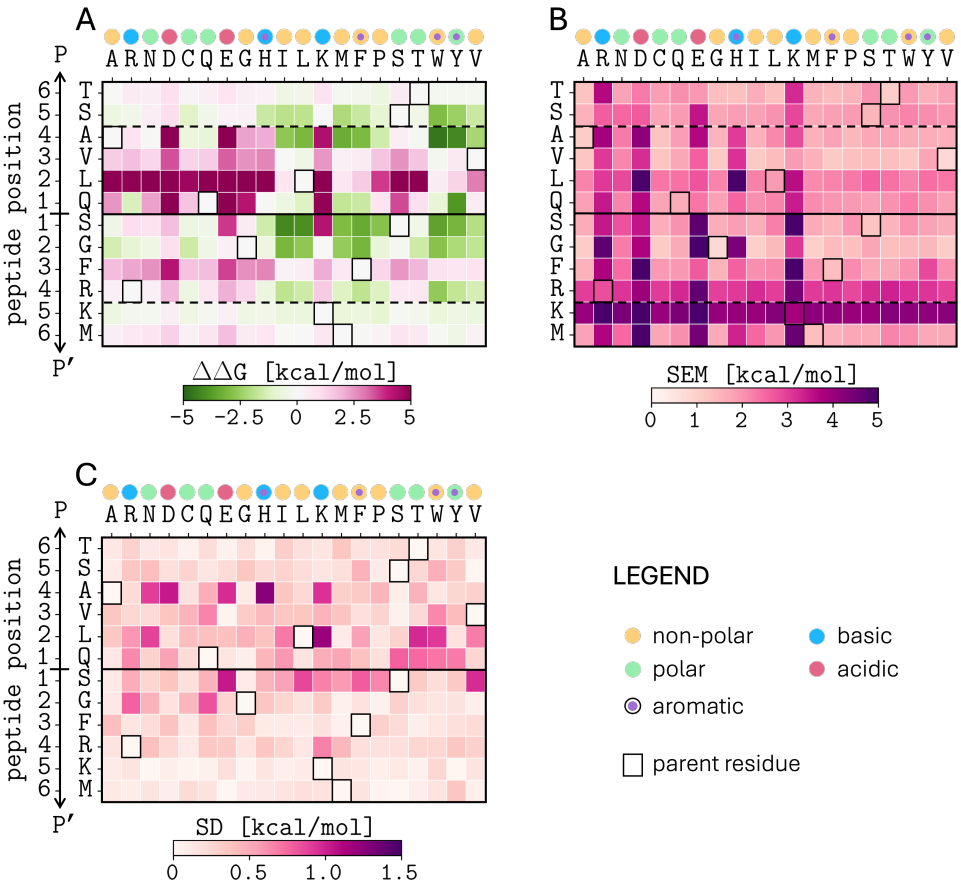


Figure 4.11.: Computational Dynamic Saturation Mutagenesis of the M^{pro}–m01 complex MD simulation. The 100 ns MD simulation of M^{pro} in complex with the substrate peptide m01 was used as input for the CDSM, performed in triplicate using five snapshots each. (A) The ΔΔG values for each peptide residue and substitution are shown as the mean of the triplicates, with favorable and unfavorable mutations in green and pink, respectively. (B) The respective standard error of the mean (SEM) was determined across the five snapshot simulations and replica runs through error propagation. (C) The standard deviation (SD) was computed for the triplicate means. The peptide sequence on the y-axis is further separated into P and P′ peptide positions. The mutated amino acids on the x-axis are classified based on their side chain properties: non-polar (yellow), polar (green), basic (blue), and acidic (red). Aromatic residues are additionally marked by a purple dot. When the parent residue matches the mutated amino acid, the field is marked by a black frame.

complex library in Section 4.2: mutations to negatively charged residues (Asp and Glu) as well as the positively charged Lys lead to highly unfavorable interactions, particularly in the most conserved peptide region P4–P1′. Together with the basic Arg, these mutated amino acids are further associated with high uncertainties, making them unsuitable for peptide design.

In contrast, the most favorable mutations involve non-polar and aromatic residues, particularly at the peptide position P4 (Ala), the highly conserved P1 (Gln) and the P′ site. This aligns with the results from Section 4.2, where the most favorable peptide designs

predominantly feature non-polar residues at all positions, except for Arg at P3.
Interestingly, mutations of Val at P3, Leu at P2, and Phe at P3′, resulted in the most unfavorable ΔG_{bind} values, with no substitution yielding distinct improvement, indicating that the parent amino acids are already optimal for their respective positions.
To obtain c-designs with a length of eight amino acids, we combined the most favorable predictions for the P4–P4′ positions. All designs were generated by modifying the m01 substrate peptide, followed by 100 ns MD simulations of the M^{pro}–design complexes. All c-designs remained bound to M^{pro}, with the majority (except c02 and c06) exhibiting considerably high stability throughout the simulation, as shown by RMSD and RMSF analysis (Fig. A33).
Fig. 4.12A shows the per-residue ΔG_{bind} decomposition for the c01–c03 designs, computed using the MM/PBSA method. The peptides c01–c03 were generated based solely on insights obtained from CDSM of substrate m01. While Trp was predicted to be highly favorable at multiple positions, we intentionally introduced sequence variability by selecting other favorable amino acids at these sites to avoid over-representation of this large aromatic residue. Among the c01–c03 designs, c01 exhibited the best total ΔG_{bind} with -74.6 kcal/mol (Fig. 4.12D), which is already in a good range compared to the substrate

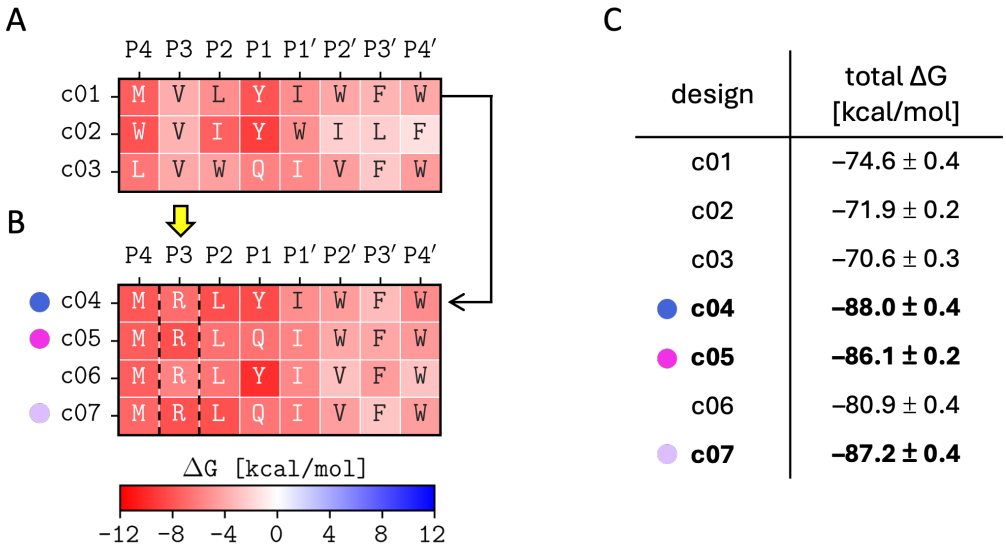


Figure 4.12.: CDSM-guided rational peptide inhibitor design against M^{pro}. The per-residue binding free energy decomposition (ΔG_{bind}) was computed for each peptide position (P6–P6′) of M^{pro} simulated for 100 ns in complex with the peptide designs (A) c01–c03 and (B) c04–c07. By introducing Arg at peptide position P3 (marked by a yellow arrow) in c01, c04 was obtained through rational design. All amino acid sequences are displayed from N- to C-terminus. In all color plots, the ΔG_{bind} values range from -12 (red) to $+12$ kcal/mol (blue). (C) The total ΔG_{bind} values are summarized for all c-designs. Designs selected for inhibitory assays in the wet lab are marked by colored dots.

and peptide inhibitors in Section 4.3. However, we identified potential for improvement, particularly at position P3. While the CDSM did not suggest any mutation that would significantly improve binding affinity over Val at P3, our design simulations in Section 4.3 indicated that Arg would be a highly favorable residue at this position. Furthermore, c01 is a highly hydrophobic peptide, which could lead to solubility issues. Thus, introducing Arg at P3 not only greatly improved the total ΔG_{bind} to -88.0 kcal/mol—thereby surpassing the binding affinity of all peptide inhibitors and designs in Section 4.3—but also introduced charges into the peptide, enhancing its pharmacokinetic properties.
Alongside c04, we selected the promising designs c05 and c07, with ΔG_{bind} values of -86.1 kcal/mol and -87.2 kcal/mol, for *in vitro* testing of inhibitory activity against M^{pro} in Section 4.5.1. The designs c05 and c07 differ from c04 by a Gln at P1, and in the case of c07, also by a Val at P2′. The introduction of Gln in c05 increases polarity, thereby improving solubility, whereas c07 was designed to study the effect of reducing the number of aromatic residues on binding affinity.

4.4.2. PL^{pro} Inhibitor Design.

In Section 4.2, we performed 100 ns MD simulations of PL^{pro} in complex with the L-peptides p01–p03 (Fig. 4.7). We used these simulations for CDSM to obtain design suggestions for the generation of novel PL^{pro} peptide inhibitors. Fig. 4.13A presents the computed $\Delta\Delta G$ values and their respective SEM for each parent peptide position and mutation, averaged across all three CDSM runs with p01–p03 (Fig. A34). The plot reveals that mutations to negatively charged residues (Asp and Glu) at all peptide positions (P6–P2′) and to bulky aromatic amino acids (Phe, Trp, and Tyr) at P2 lead to highly unfavorable interactions, which are associated with high uncertainties. Increased uncertainties were observed at positions P6 and P5 across all mutations due to the parent residues Arg and Glu in p01 (Fig. A34A). Mutation of the conserved Leu at P4 generally results in unfavorable ΔG_{bind} values, with Arg as the exception. The most favorable mutations involve non-polar and aromatic residues, especially at positions P3–P1′, with P2 showing the greatest potential for improved binding affinity upon mutation to Ile, Leu, or Met.
We combined promising design suggestions for positions P6–P2′ from CDSM to create z-designs, which were generated by modifying the p02 substrate peptide. The designs were then submitted to 100 ns MD simulations in complex with PL^{pro}. Fig. 4.13B presents the per-residue ΔG_{bind} decomposition, computed using the MM/PBSA method, for the z04–z08 designs and the reference peptide p02 for comparison.
The designs with the most favorable total ΔG_{bind} values (Fig. 4.13B) are z07 (-83.0 ± 0.2 kcal/mol) and z08 (-83.3 ± 0.3 kcal/mol), which exceed the ΔG_{bind} of p02 by 23 kcal/mol. They both feature an Arg at P4, which exhibits considerably improved affinity compared

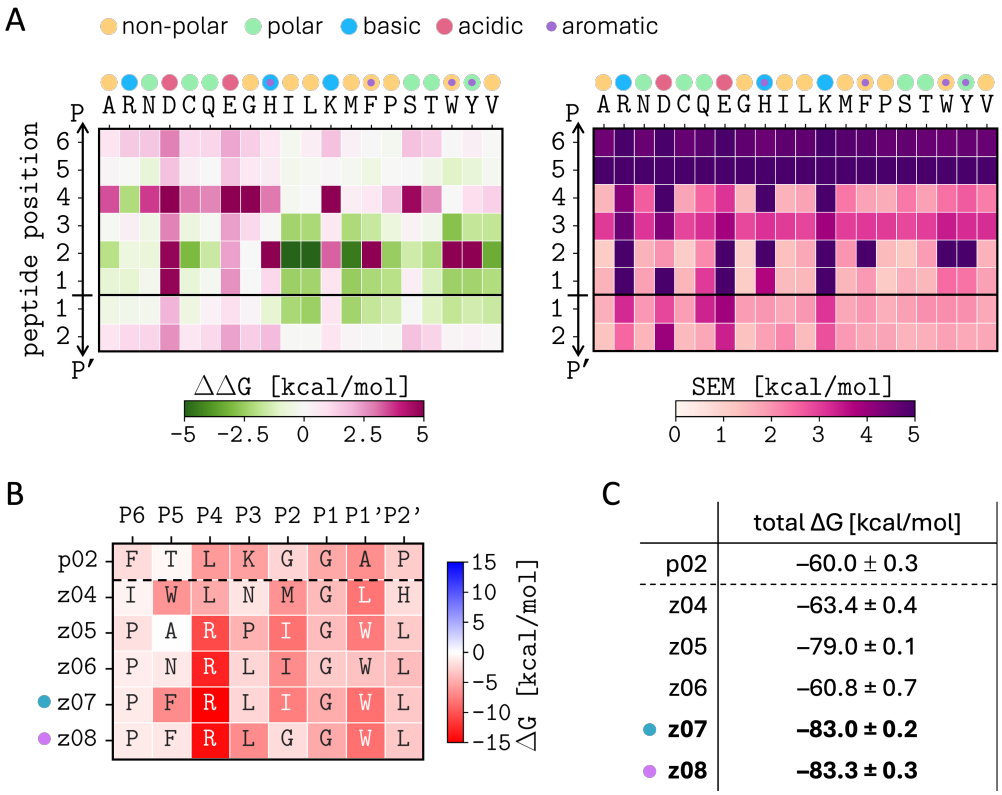


Figure 4.13.: CDSM-guided rational peptide inhibitor design against PL^{pro}. (A) The 100 ns MD simulations of PL^{pro} in complex with the substrate peptides p01, p02 and p03 were used as input for CDSM using five snapshots each. The ΔΔG values for each peptide position of P6–P2′ and substitution are shown as the mean of p01–p03. The respective standard error of the mean (SEM) was determined across the three runs and the five snapshot simulations through error propagation. The mutated amino acids on the x-axis are classified based on their side chain properties: non-polar (yellow), polar (green), basic (blue), and acidic (red). Aromatic residues are additionally marked by a purple dot. (B) The per-residue ΔG_{bind} decomposition was computed for each peptide position (P6–P2′) of PL^{pro} simulated for 100 ns in complex with the peptide designs z04–c08. (C) The total ΔG_{bind} values are summarized for all z-designs. The p02 substrate peptide is included as reference. Designs selected for inhibitory assays in the wet lab are marked by colored dots.

to the parent Leu residue, which is conserved at this position across PL^{pro} substrates. The improved affinity can be attributed to a negatively charged cavity in the PL^{pro} substrate binding site that is occupied by Arg, as shown in Figures 4.14B and C for z07 and z08, respectively. Arg interacts with this cavity primarily through hydrogen bonding with the binding site residues Asp164, Ala246 (backbone-mediated), Asp302, Tyr273, and Thr301. Fig. 4.14A shows that this cavity is small when p02 is bound but expands upon Arg binding. The Ile substitution at P2 in z07 improves binding affinity compared to the conserved Gly in the parent peptide and also seems to influence the ΔG_{bind} of other positions, particularly Leu at P3 and Phe at P5. This is evident from a comparison of z07 and z08, which differ only at P2 (Ile vs. Gly). Because Ile is bulkier than Gly occupying the narrow S2 subsite, it

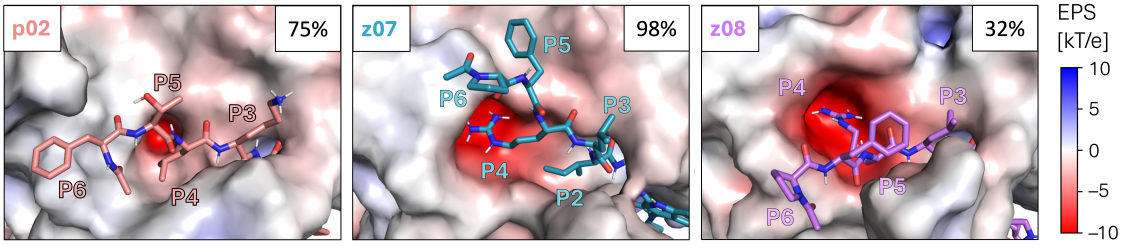


Figure 4.14.: Electrostatic interactions of the peptide designs z07 and z08 with PL^{pro}. The most populated conformations of the peptides simulated in complex with PL^{pro} are shown for (A) p02 (75 % sampled, reference), (B) z07 (98 % sampled), and (C) z08, with the first and second clusters each representing 32 % of the peptide poses. The EPS was computed for the PL^{pro} structures and is shown at 30 % transparency, with red and blue indicating negatively and positively charged areas on the PL^{pro} surface, respectively. The peptide structures are displayed as sticks, with positions P6–P2 marked.

opens the tunnel that usually harbors the conserved GG substrate motif, increasing solvent exposure of Leu at P3 in z07 (Fig. 4.14B) compared to z08 (Fig. 4.14C), explaining the reduced binding affinity of z08. Additionally, a slight backbone shift in z07 may enable stable binding of Phe at P5 to a subsite above the negatively charged cavity. In contrast to z07, P5 and P6 in z08 are highly flexible, while Arg at P4 remains very rigid in both designs (Fig. A35). To assess how these structural differences affect inhibitory potency, we evaluated both z07 and z08 against PL^{pro} in *in vitro* assays (Section 4.5.3).

4.5. Experimental Validation of the Peptide Designs

4.5.1. Inhibitory Activity of Peptide Designs against M^{pro}

Production of Natural M^{pro}. To obtain reliable inhibition results for the designed peptides, M^{pro} had to be produced with high purity and activity. However, M^{pro} presents a special case, as both its N- and C-termini are close to its substrate binding site. It was shown that the presence of excess residues at the protease termini impedes its activity[292,293]. Thus, the preparation of natural M^{pro} (natM^{pro}) with authentic N- and C-termini is required. Zhao et al. (2024) presented a method in which the His-tag is located at the N-terminus, followed by a Tobacco Etch Virus (TEV) protease recognition sequence (TEV-site)[274]. The optimum recognition site of TEV protease is ENLYFQ↓G/S, with cleavage occurring between the Gln and Gly or Ser[294]. To obtain a natural N-terminus in M^{pro}, a Ser must occupy the P1′ position, as seen in the cleavage site sequence of nsp4/5 (Fig. 4.1). To optimally avoid auto-cleavage by M^{pro}, Zhao et al. tested various amino acids at the P2 position of the TEV-site, with Ser proving to be the most efficient. Thus, to yield natM^{pro},

we used the sequence ENLYSQS as the TEV-site at the N-terminus and introduced a termination codon after the last M^{pro} residue to also ensure an authentic C-terminus. Successful recombinant expression of His_6-tagged M^{pro} in *Escherichia coli* (*E. coli*), followed by His-tag cleavage to produce $natM^{pro}$, including multiple purification steps using immobilized metal ion affinity chromatography (IMAC), was verified by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (Fig. A36A and B).

Initial Inhibition Assays. The inhibitory activity of the peptide designs against M^{pro} was determined using an *in vitro* assay with the fluorogenic substrate Dabcyl-KTSAVLQ↓SGFRKME-Edans. First, an initial screen was performed with all selected peptide designs from Strategy I–III: Dri-m01 (Section 4.2); a05, a06, a10, a21, b06 and b07 (Section 4.3); and c04, c05 and c07 (Section 4.4). M^{pro} was incubated with 100 µM of each peptide for 15 min, after which its activity was measured. The results are shown in Fig. 4.15. Peptide designs Dri-m01, a05, a06, a21, b06, and b07 showed no or only minimal inhibitory activity against M^{pro}. In contrast, designs c05 and c07 inhibited M^{pro} by approximately 90 %. Designs a10 and c04, which achieved complete inhibition, were selected for subsequent IC_{50} estimation.

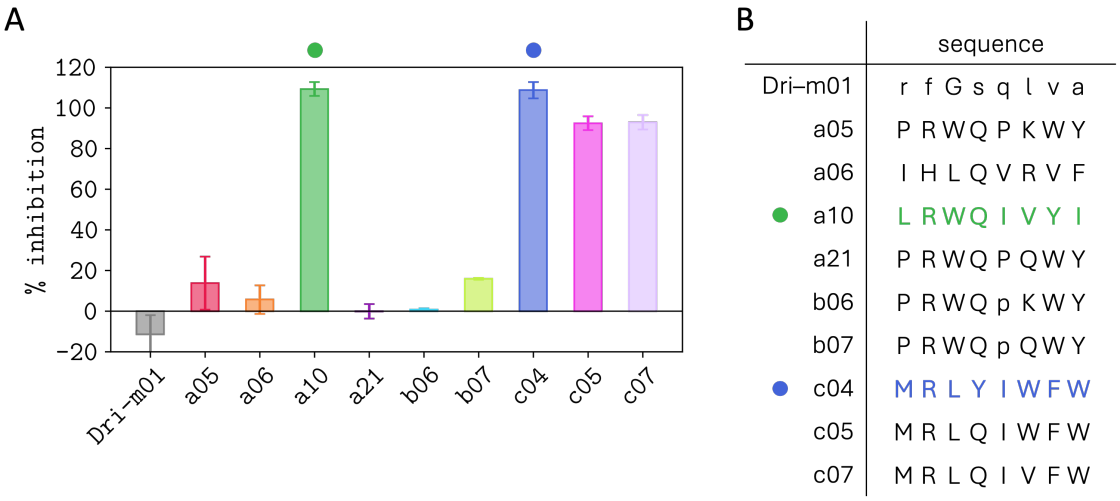


Figure 4.15.: M^{pro} inhibition after incubation with peptide designs. The activity of M^{pro} was measured using an *in vitro* fluorogenic assay following incubation with 100 µM of each peptide design. The assay conditions were 0.5 µM $natM^{pro}$, 10 µM substrate in 20 mM Tris-HCl, pH 7.2 and 200 mM NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA) and 1 mM tris(2-carboxyethyl)phosphine (TCEP). Measurements were performed in biological duplicates. (A) The M^{pro} inhibition data relative to the control (no inhibitor) are presented as a barplot, and (B) the amino acid sequences of the peptide designs are summarized. Peptide designs that inhibited M^{pro} to 100 % were selected for IC_{50} estimation and are marked with colored dots.

IC_{50} Estimation. As the standard metric for drug potency, the IC_{50} enables direct comparison of inhibitors within this study and with literature. IC_{50} values can be determined by constructing dose-response curves that evaluate the effect of varying inhibitor concentrations on protein activity. In this manner, peptides a10 and c04 were assayed across concentration ranges of 0–400 µM and 0–100 µM, respectively, using the *in vitro* fluorogenic assay with $natM^{pro}$ (Fig. 4.16A,B). Disulfiram was used as a standard, assayed across a 0–100 µM concentration range (Fig. 4.16C).
The a10 and c04 peptide designs demonstrate different potencies with IC_{50} values of 22.3 ± 4.7 and 4.8 ± 0.5 µM, respectively. Compared to Disulfiram with an IC_{50} of 4.6 ± 0.7 µM, c04 presents a promising drug candidate.

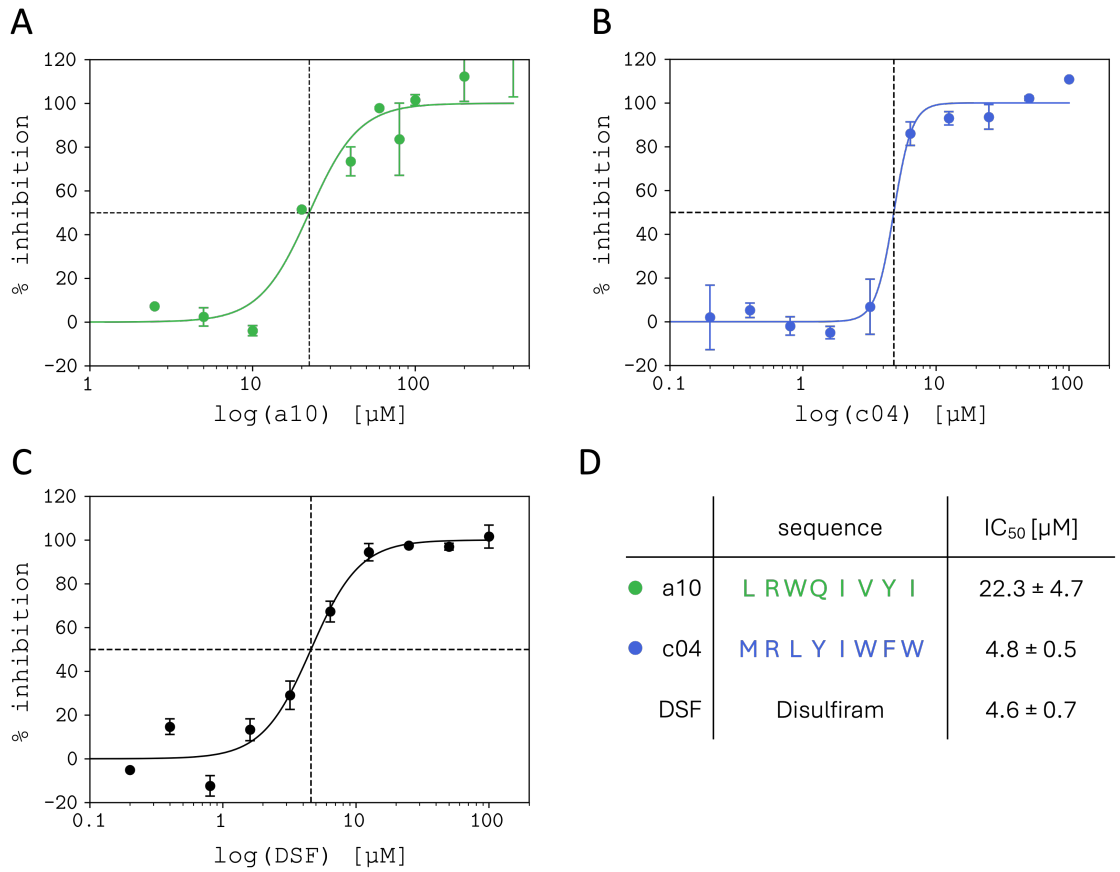


Figure 4.16.: Estimation of IC_{50} for the designs a10 and c05. The inhibition (%) of M^{pro} was determined using an *in vitro* fluorogenic assay across a range of concentrations for peptide designs (A) a10, (B) c04, and, (C) as a standard, the inhibitor Disulfiram. The assay conditions were 0.5 µM $natM^{pro}$, 10 µM substrate in 20 mM Tris-HCl, pH 7.2 and 200 mM NaCl, 1 mM EDTA and 1 mM TCEP. Measurements were performed in biological duplicates. IC_{50} values were estimated through dose-response fitting curve data to estimate IC_{50} values for each measurement, (D) summarized in atable.

4.5.2. Correlations between *In Vitro* and *In Silico*

To evaluate our design predictions, we assessed the correlation between ΔG_{bind} values from 100 ns MD simulations of peptide–M^{pro} complexes (Sections 4.2–4.4) and experimental inhibition potency against M^{pro} from initial screening assays (Fig. 4.17A).
Peptide designs with moderate ($> 90\,\%$) or even complete inhibition also exhibited the most favorable ΔG_{bind} values (< -85 kcal/mol), with a05 being the only outlier. Consistently, peptides with low or negligible inhibition ($< 20\,\%$) showed less favorable predicted binding affinities. Most interestingly, for c04, the peptide design with the highest predicted binding affinity ($\Delta G_{bind} = -88.0 \pm 0.4$ kcal/mol), the highest inhibitory potency ($IC_{50} = 4.8 \pm 0.5\,\mu M$) was determined. Despite the limitations of single-concentration inhibition data and insufficient points for linear regression, the results demonstrate correlations between *in vitro* and *in silico* findings supporting our CAPID methods.
Comparison of amino acid sequences of the designs, sorted by inhibition potency, reveals specific amino acids and features at defined peptide positions that are unique to designs with moderate to high inhibition potency (Fig. 4.17B). For example, the top-four designs all include an Ile at P1′, a non-polar residue at P2′, and predominantly non-polar and

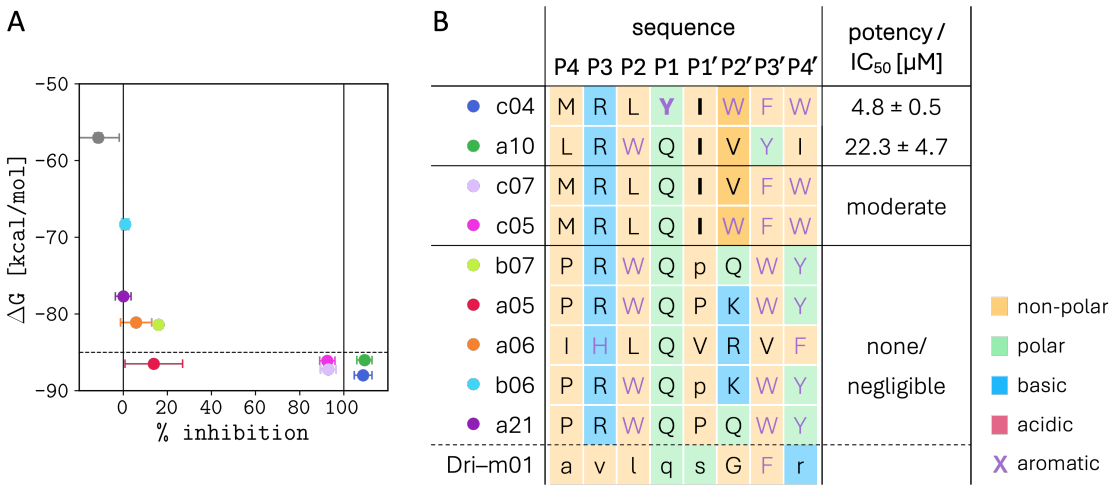


Figure 4.17.: Correlation between *in silico* ΔG_{bind} estimation, *in vitro* inhibition potency and peptide design features. (A) Inhibition (%) of M^{pro} from initial *in vitro* fluorogenic assays using 100 µM peptide plotted against ΔG_{bind} from 100 ns MD simulations in complex with M^{pro}. (B) Peptide designs sorted by inhibition potency at 100 µM inhibitor concentration: c04 and a10 (complete inhibition, with IC_{50} determined), c07 and c05 (moderate, > 90 % inhibition), and the remainder (none or negligible, < 20 % inhibition). Amino acid sequences are shown with backgrounds colored by side chain properties: non-polar (yellow), polar (green), basic (blue), and acidic (red). Aromatic residues are additionally indicated with purple characters. All sequences are displayed from N- to C-terminus, except for Dri-m01, which is shown from C- to N-terminus. Peptide positions P4–P4′ are indicated. Amino acid types or features unique to the peptides with moderate to high inhibiton potency are highlighted in bold or with a stronger background.

aromatic amino acids at P2′–P4′.

We designed c07 based on c05 to assess the effect of reducing aromatic residues in the designs. The similar moderate inhibitory potency of both peptides shows that changing P2′ from aromatic Trp to aliphatic Val had no effect on activity towards M^{pro}, aligning with our computational ΔG_{bind} values for Trp and Val at this position.

Notably, the most potent design, c04, is distinguished from all others by a Tyr residue at P1. This is also the only feature that distinguishes it from the less potent design c05, highlighting its importance for enhanced binding affinity.

Fig. 4.18 shows a representative structure of the most popular conformations from 100 ns MD simulations of M^{pro} with peptide designs a10 and c04. Interestingly, the structure of c04 looks more stretched than that of a10, assuming a more relaxed conformation of c04. However, both peptide designs exhibit the conserved main-chain mediated hydrogen bonds at P3, P1, P2′ and P4′.

The Gln at P1 in a10 forms many hydrogen bonds with the side chains of Glu166 and His163 and the main chain of Phe140, as observed for the m01 substrate. In contrast, no hydrogen bond formation is observed for Tyr at P1 in c04. However, hydrogen bond existence analysis shows that the side chain of Tyr forms hydrogen bonds to the same percent (∼75 %) as the Gln in a10 (Fig. A37). Analysis of the trajectory of the M^{pro}–c04 complex shows that Tyr forms hydrogen bonds via its hydroxy group either with the side chains of Glu166 or His172, or sometimes even both. The hydrogen bond existence analysis further shows that the P2 Trp side chain in a10 forms highly consistent hydrogen

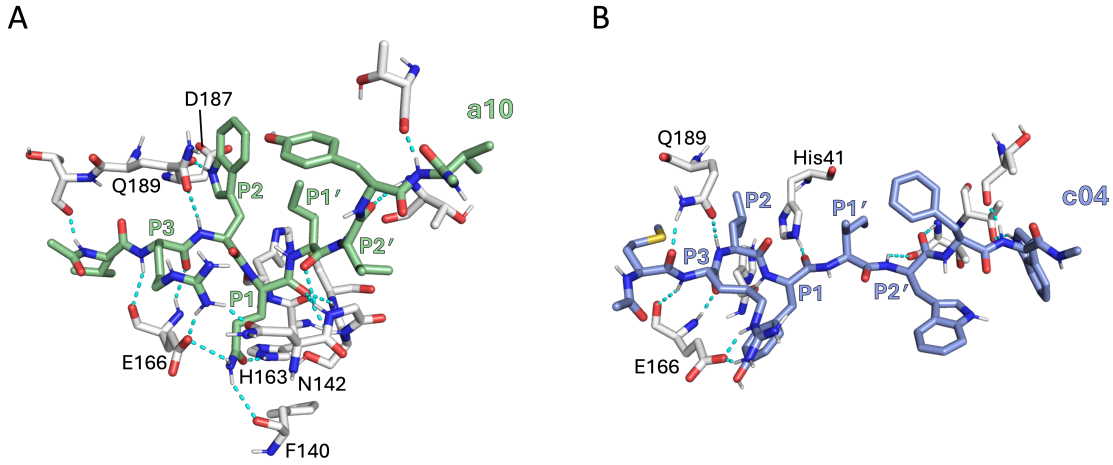


Figure 4.18.: Interaction pattern of peptide designs a10 and c04 with M^{pro}. The most populated conformations of the simulations with (A) a10 (84 % sampled) and (C) c04 (99 % sampled) are shown in complex with M^{pro}. The peptide structures are displayed as sticks in green (A) and blue (B). M^{pro} residues forming hydrogen bonds with the peptides are shown as white sticks, with hydrogen bonds shown as cyan dotted lines.

bonds with the main chain of Asp187, as displayed in Fig. 4.18A. Both peptide designs contain an Arg at P3, which forms moderately consistent hydrogen bonds, usually with the side chains of Glu166 and Asn142.

4.5.3. Inhibitory Activity of Peptide Designs against PL^{pro}

Unlike M^{pro}, where excess terminal residues considerably impair catalytic activity[292,293], previous studies have shown that a terminal His_6-tag has only a minimal effect on the catalytic efficiency of PL^{pro} [143,295]. Thus, PL^{pro} was expressed with a C-terminal His_6-tag in *E. coli* and purified by IMAC (Fig. A36C).

We tested the peptide designs Dri-p02 (Section 4.2), z03 and z04 (Section 4.4) against PL^{pro}, using an *in vitro* assay with the fluorogenic substrate Dabcyl-FTLKGG↓APTKVTE-Edans. Fig. 4.19 shows the inhibition of PL^{pro} with 50 µM of each peptide after incubation for 10 min or direct measurement (0 min).

No inhibitory activity was observed for the peptide designs after 10 min incubation. However, direct measurement upon peptide addition revealed inhibition of up to 30 % (Dri-p02), 20 % (z07), and 11 % (z08). These findings suggest that the peptides initially interact with the protease to some extent, but their potency decreases over time. This loss of potency could result from peptide cleavage by PL^{pro} or—particularly for the hydrophobic designs z07 and z08—from aggregation over time. Given the negligible inhibitory activity, these designs exhibit insufficient potency for lead compound consideration. Thus, no further IC_{50} determination assays were performed.

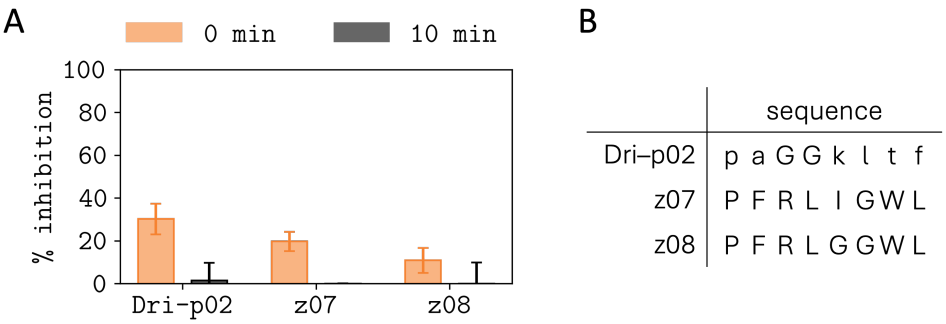


Figure 4.19.: PL^{pro} inhibition after incubation with peptide designs. (A) The activity of PL^{pro} was measured using an *in vitro* fluorogenic assay following addition of 50 µM of each peptide design (with 0 min or 10 min incubation). The M^{pro} inhibition data relative to the control (no inhibitor) are presented as a barplot. The assay conditions were 0.5 µM PL^{pro}-His_6, 10 µM substrate in 50 mM HEPES, 5 mM DTT, 0.01 % Triton X-100, pH 7.5. Measurements were performed in biological duplicates. (B) The amino acid sequences of the peptide designs are summarized.

4.6. Discussion

We explored three strategies for computer-aided peptide inhibitor design (CAPID) against two proteases of SARS-CoV-2, M^{pro} and PL^{pro}, all based on MD simulations of the proteases with their natural substrates. In strategy I, we investigated whether the retro-inverso method could be applied to the substrate peptides to yield inhibitors with similar affinity but improved proteolytic stability. To accomplish this, we first prepared models of the proteases in complex with the L- and Dri-peptides, which served as starting structures for MD simulations.

For M^{pro}, available crystal structures of the protease in complex with its substrates—though not fully resolved—enabled validation of our L-substrate docking results using ADCP. The L-substrates exhibited highly stable binding at peptide positions P4–P1′, consistent with the co-crystallized substrates. In contrast, the Dri-peptides showed reduced binding stability and low binding affinities. This instability was attributed to the switched termini and amide bonds in the Dri-peptides, which likely disrupt the conserved hydrogen bonds essential for M^{pro}–substrate binding. Most of these hydrogen bonds are backbone-mediated in both the peptide and the protein. MD simulations showed that protein adaptation to these peptides is complicated by the intrinsic rigidity of backbone-mediated hydrogen bonds compared to side-chain-mediated interactions. Consequently, the Dri-peptides were expected to lack inhibitory potential against M^{pro}, which was confirmed by *in vitro* inhibition assays with Dri-m01 exhibiting no inhibition of M^{pro}.

To generate PL^{pro} complexes with L- and Dri-peptides p01–p03, we used a consensus docking approach, combining physics-based docking (ADCP) with AI-based structure prediction (AF3) to improve reliability, as no experimentally resolved PL^{pro}–substrate complex structures are available. We obtained a highly confident AF3 pose for the 8-residue L-p02, which we used as a template for distance cutoffs in re-scoring ADCP docking solutions to generate Dri-peptide poses. MD simulations of L- and Dri-peptides in complex with PL^{pro} revealed reduced affinity for Dri-peptides, although they were able to adapt to the binding site environment. A shift towards the P-site enabled retention of specific backbone-mediated hydrogen bonds, and the peptides exhibited relatively high stability and rigidity throughout the simulations and across all peptide positions compared to the L-peptides. Nevertheless, Dri-p02 exhibited low inhibitory potency in *in vitro* assays against PL^{pro}. These results raise the question of why the Dri-method succeeded for the competitively inhibiting peptides 3CVL and 3CVLri, whose binding modes with the active site of M^{pro} we had previously characterized in Chapter 3[65].

In Strategy II, we repeated docking and modeling of the peptides 3CVL-2, 3CVLri-2, 3CVL-7, and 3CVLri-7, aiming to identify a binding mode analogous to that of the substrate peptides to obtain valuable information for peptide design. Therefore, we employed

the docking tool ADCP[39], which was specifically designed for peptide docking. However, the results remained ambiguous, often showing higher binding affinities for peptide conformations that deviated from the substrate binding mode, thereby binding either in the wrong orientation or completely randomly. In addition, given that our computational and experimental findings indicate that the retro-inverso method is ineffective for M^{pro} substrates, one could conclude that the 3CVLri peptides—and possibly their L-counterparts—do not bind in the same manner as the substrate peptides. Our modeling results from Chapter 3 and this work further suggest that these peptides may also interact with M^{pro} in multiple, yet individually stable, binding modes. For example, 3CVL-7 contains two His residues in close proximity and exhibits favorable binding affinities for both conformations, with either residue occupying the P1 position.

Indeed, multiple stable binding modes of peptides have been reported. For example, the short LPATG sorting signal has been shown to adopt multiple metastable states in complex with a sortase enzyme, as verified by nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallization[296].

However, as the 3CVL(ri) peptides are relatively large (16–17 amino acids) and linear, they possess a high number of degrees of freedom, which makes it challenging for docking programs to identify the optimal binding pose. In comparison to other peptide docking engines, ADCP performs optimally for peptides of 5–10 amino acids and remains effective for peptides of 11–20 amino acids but with decreased success rates[41]. Furthermore, standard MD simulations often struggle to efficiently sample transitions between conformational states in peptide–protein complexes, associated with high energy barriers.

To address this, enhanced MD sampling techniques, such as replica exchange MD (REMD), accelerated MD (aMD), and Gaussian accelerated MD (GaMD), have been developed to improve the sampling efficiency of MD simulations and have been successfully applied to peptide–protein modeling[297,298]. Wang et al. (2020) introduced Peptide Gaussian Accelerated Molecular Dynamics (Pep-GaMD), a GaMD variant specifically tailored for modeling the high flexibility of peptides in peptide–protein interactions (pepPIs) using a dual-boost algorithm that applies two distinct boost potentials on the potential energy of the peptide and the remaining system[299]. Their PepBinding workflow, which combines peptide docking, enhanced sampling using Pep-GaMD, and structural clustering, could be a valuable tool for predicting the optimal binding poses of the 3CVL and 3CVLri peptides[300]. For experimental verification, NMR or X-ray crystallization would be the methods of choice.

Our Strategy II, using a library of M^{pro}–peptide MD simulations that includes natural substrates and inhibitors, has shown potential for developing peptides with improved binding affinities. Additionally, it identified a peptide design, a10, with inhibitory activity against

M^{pro}, exhibiting an IC_{50} value of 22.3 ± 4.7 µM. However, the M^{pro}–peptide library lacks exhaustive coverage, limiting peptide design efficiency. Thus, we employ Strategy III, using computational saturation mutagenesis to systematically evaluate all possible amino acids at each peptide position. Several computational saturation mutagenesis methods have already been developed, each with distinct advantages and limitations. For example, MutateX[301] uses the FoldX energy function[302] and provides an automated pipeline for ΔΔG calculations over protein ensembles from MD simulations. However, as it does not relax the protein backbone upon mutation, steric clashes may occur when replacing a residue with a larger side chain, potentially resulting in artificially inflated ΔΔG values. To avoid these artifacts, the predictive saturation variation scanning (PreSaVS) method[75] employs the Bristol University Docking Engine (BUDE) force field[76], which avoids penalizing small geometric overlaps in the modeled structures, enabling faster computations but at the cost of reduced accuracy.

The Rosetta software package includes a protocol for site saturation mutagenesis through point mutation evaluation[303]. After mutation, its FastRelax tool performs a minimization step that allows side-chain and backbone movement for residues near the mutation site, enabling more realistic structural adaptation and more accurate energy predictions. However, this protocol was specifically designed to predict stabilizing mutations in single proteins to enhance thermodynamic stability, not for peptide–protein interactions. Instead, Rosetta provides another tool, FlexxDDG[304], which models changes in protein–protein or peptide–protein interactions upon mutation and computes the resulting interface ΔΔG values. However, this program lacks automated tools for saturation mutagenesis.

The advantage of our CDSM approach is its compatibility with the widely used MD simulation program GROMACS, while enabling ΔG_{bind} calculations with the MM/GBSA method using the gmx_MMPBSA software, which supports Amber[52], OPLS[305], and CHARMM[248] force fields, allowing users to adapt the protocol to their specific MD simulation system.

Unlike other protocols, our approach allows both the protein and the peptide to adopt their conformations without constraints on the mutated side chains. This is achieved through a short MD simulation (100 ps) following energy minimization and equilibration. Although this method may be computationally more expensive, the required resources scale with peptide length and the number of snapshots. For example, a CDSM calculation for a 10-amino acid peptide using 5 snapshots, as in our work, requires computational resources comparable to a 100 ns simulation (10 positions $\times$ 5 frames $\times$ 20 aa $\times$ 100 ps $=$ 100 ns). This corresponds to a single trajectory of our 20 peptide–protein MD simulations for library generation (Section 4.3), while providing more holistic design suggestions for each peptide position.

We have applied CDSM to the M^{pro}–m01 complex MD simulation. Reasonable standard deviations across triplicates validated the protocol's reproducibility. The results for m01 revealed a clear preference for non-polar and aromatic residues, whereas charged residues such as Asp, Glu, and Lys were highly unfavorable, especially at the more conserved positions P4–P1′. These findings aligned with those from Strategy II and previous studies[65,74]. However, our CDSM protocol failed to predict Arg as a more favorable residue than Val at position P3, whereas 100 ns simulations of peptides with this substitution clearly demonstrated an improvement in ΔG_{bind}. Observation of the CDSM MD simulation snapshots after mutation to Arg at P3 revealed that the residue rarely adopted a conformation required for favorable interactions with the Glu166 side chain. This suggests that the simulation time may be insufficient for larger substitutions such as Arg. The simulation time could be adapted according to the size and flexibility of the side chains, thus maintaining moderate computational costs by retaining shorter simulation times for smaller, more rigid residues that require less sampling to adapt.

For PL^{pro}, CDSM was able to identify Arg as the only mutation at P4 that improved ΔG_{bind}, a position that typically contains the conserved Leu of the LXGG binding motif. 100 ns MD simulations of the peptide designs with Arg at P4 validated the enormous improvement in ΔG_{bind} compared to Leu in substrate p02, by binding to a negatively charged cavity within the PL^{pro} substrate binding site. Combined with other favorable mutations suggested by CDSM, this yielded a 20 kcal/mol improvement in total ΔG_{bind} for peptide designs z07 and z08. However, these designs exhibited negligible inhibitory activity against PL^{pro} in *in vitro* assays. Nonetheless, our CDSM approach provided valuable mutation suggestions that, in combination with rational design, resulted in our most potent M^{pro} peptide inhibitor, c04, as verified by *in vitro* fluorescence-based inhibition assays.

Our experimental setup had some limitations, as measurements were performed in duplicate rather than in triplicate, reducing statistical robustness. However, this approach was primarily adopted due to the high cost of the fluorogenic substrate and the limited amount of $natM^{pro}$ produced. M^{pro} further exhibits inherent instability due to its tendency to interconvert between its active dimeric form and an inactive monomeric form through redox modifications of its cysteine residues[306]. Oxidative stress can induce dissociation into monomers, leading to a loss of enzymatic activity. Although TCEP was included in the buffer to maintain a reducing environment and multiple aliquots were prepared to minimize exposure to air, the loss of M^{pro} activity could not be entirely avoided.

To validate our experimental setup, we used disulfiram as a standard. Disulfiram is a cysteine protease inhibitor and an alcohol-aversive drug approved by the Food and Drug Administration (FDA). Previous studies suggested that disulfiram targets and conjugates to Cys44, located at the edge of the M^{pro} substrate binding pocket, thereby interfering

with substrate binding to the active site[307]. In the same study, the IC_{50} was determined to be $3.79 \pm 0.23\,\mu M$ for SARS-CoV-2 in a fluorescence-based peptide cleavage assay, comparable to our result of $4.6 \pm 0.7\,\mu M$ for disulfiram, supporting the validity of our experimental methods and results.

In initial inhibitory tests with M^{pro}, we identified two peptide designs, c04 and a10, that completely inhibited M^{pro} at 100 µM inhibitor concentration. Subsequent IC_{50} determination yielded values of $4.8 \pm 0.5\,\mu M$ for c04 and $22.3 \pm 4.7\,\mu M$ for a10. The IC_{50} value of c04 is comparable to those of other M^{pro}-targeted peptide inhibitors, such as p12, p13, p15, and p16, presented by Chan et al. (2021)[74], which range from 3.11 to 5.36 µM. Notably, our c04 peptide, containing eight amino acids, is shorter than these eleven-amino-acid peptides. A shorter length while exhibiting similar inhibition potency indicates improved affinity and may also impact pharmacokinetics; for example, reducing peptide size can increase cell permeability[308].

Peptide designs Dri-p02, z07, and z08 showed no inhibitory activity against PL^{pro} at 50 µM after 10 min incubation. However, without incubation (0 min)—i.e., direct measurement after peptide addition—low inhibition (up to 30 % for Dri-p02) was observed. This suggests that the peptides can interact with PL^{pro} to some extent but lose potency over time, potentially due to cleavage by the protease. The Dri-method was used to prevent peptide cleavage; however, the Dri-peptides of PL^{pro} still contain two Gly residues, which lack chiral centers and thus cannot be converted to D-amino acids. Our MD simulations suggested that the Dri-peptides are shifted, thereby disrupting important interactions with active site residues, including the oxyanion hole-stabilizing Trp106[291], suggesting that catalytic cleavage of these peptides is not possible. However, to investigate this question, specific cleavage experiments with PL^{pro}, followed by analysis via RP-HPLC for example, would be needed.

Another explanation for the loss of potency could be peptide aggregation, particularly for the hydrophobic designs z07 (PFRLIGWL) and z08 (PFRLGGWL). Tools for predicting peptide aggregation have been introduced, including the widely used algorithm TANGO[309] and a recently developed machine learning (ML)-based tool PALM[310], which enables the identification of aggregation-promoting mutations in Aβ peptides. Experimental validation of this hypothesis could be done using aggregation assays, such as a Thioflavin T fluorescence assay[311] or other spectroscopic methods like dynamic light scattering[312].

Substrate library screening studies of SARS-CoV and SARS-CoV-2 PL^{pro} for peptide positions P4–P1 revealed that Gly is required at positions P1 and P2 for protease activity. While P3 accepts a variety of amino acids, including positively charged and hydrophobic residues, P4 only tolerates hydrophobic residues, with Leu being the sole natural amino

acid accepted[313]. Peptide z07 contains Ile at P2, and both z07 and z08 contain the positively charged Arg at P4. Although MD simulations predicted high binding affinity for these substitutions, the substrate library screening studies suggest that such modifications may hinder effective binding to the PL^{pro} binding site. However, these studies assessed only catalytic activity, not binding affinity, and thus could not detect modifications that bind with high affinity while inhibiting protease activity. Nevertheless, our computational strategy failed to produce effective PL^{pro} peptide inhibitors. The specific reasons why the high predicted affinities for z07 and z08 were not reflected in *in vitro* experiments require further computational and experimental investigation.

Comparison of the experimental and computational results for the M^{pro} peptide designs suggests a correlation between the determined inhibition and predicted binding affinity, supporting our CAPID strategies, which are based on ΔG_{bind}-guided design suggestions using the gmx_MMPBSA program for MM/PBSA and MM/GBSA calculations. These findings suggest that peptide designs with a ΔG_{bind} value below -85 kcal/mol are likely to be potent in *in vitro* inhibition assays against M^{pro}, providing a useful benchmark for future designs targeting this protease. However, these binding free energies should be interpreted as relative values rather than absolute binding free energies and are therefore not universally applicable. This is because we computed the effective free energy, an approximation of the binding free energy that omits the entropic term, which allows comparison of similar peptides while being computationally efficient[59]. Comparing the more potent peptide designs with the less potent designs further reveals notable distinctions, such as the enrichment of non-polar and aromatic amino acids, particularly at peptide positions P2′–P4′. However, the most interesting distinction is the Tyr residue at P1 in c04. In M^{pro} substrates, this position is usually occupied by the conserved Gln. As it is the only residue differing from the design c05, which has reduced inhibition potency, this substitution highlights the superior binding affinity of Tyr compared to Gln. This substitution was also strongly suggested by CDSM, and its enhanced binding affinity over Gln was further validated in subsequent 100 ns simulations of the c-designs through per-residue ΔG_{bind} decomposition.

Several publications describe peptide inhibitors designed against M^{pro} with experimentally determined activities, however binding mode investigations typically show that they retain the conserved Gln residue at P1[74,314,315]. The peptide inhibitor p14 from the M^{pro}–peptide library in Strategy II contains the aromatic Trp at P1—which showed promising binding affinities, superior to the conserved Gln—however, this peptide was not tested for its inhibitory activity against M^{pro} in experimental assays[74]. Peptidomimetics with high inhibitory potency against M^{pro} often contain a 5-membered lactam at P1, thereby resembling the Glu residue at P1 while also enabling hydrophobic interactions[316,317]. One such

example is nirmatrelvir (Paxlovid, Pfizer), a reversible covalent inhibitor carrying a nitrile warhead, which is the only currently approved drug targeting M^{pro} with a reported IC_{50} of $0.022 \pm 0.004\,\mu M$[318]. However, there are some peptidomimetics that contain an aromatic residue at P1. The inhibitor ML188 possesses a pyridine ring at P1 that interacts with His163. Originally designed to inhibit the SARS-CoV M^{pro}, ML188 was reported to have inhibitory activity against SARS-CoV-2 M^{pro} ($IC_{50} = 2.5 \pm 0.3\,\mu M$)[319]. Based on ML188, Kitamura et al. (2022) designed and synthesized the non-covalent inhibitor 23R[320]. While maintaining the pyridine ring at P1, they introduced additional aromatic substitutions at P2 and P3, leading to a highly potent M^{pro} inhibitor ($IC_{50} = 0.20 \pm 0.01\,\mu M$), underlining the preference of M^{pro} for hydrophobic and aromatic groups. While most M^{pro} inhibitors focus on engaging the S3–S1′ subsites of the active site[317], our approach has shown that the S′ site also has potential for inhibitor design.

Finally, the pharmacokinetic aspects of our peptide designs should be discussed. The two main limitations of therapeutic peptides are membrane impermeability and short half-lives due to degradation and rapid clearance. In Strategy I, we attempted to address the proteolytic instability of natural peptides using the Dri-method, but this approach was unsuccessful. Some of our designs also include single D-amino acids at the P′ position (b06 and b07). However, as these designs and their L-counterparts (a05 and a21) showed no significant inhibition against M^{pro}, their effect was not further investigated. If drugs are administered orally, their stability needs to be investigated in metabolic stability experiments, for example, using solutions simulating gastric and intestinal fluids, followed by analysis via RP-HPLC to detect potential cleaved peptide fragments[65]. Similarly, hydrolysis by M^{pro} itself would need to be assessed to estimate the inhibitor's stability. Since c04 contains only L-amino acids, it is probably vulnerable to proteolytic cleavage, and such tests would be required.

To improve its proteolytic stability, modifications such as D-amino acids or other non-natural amino acids could be introduced. Holden et al. (2022) have proposed a computational site saturation mutagenesis protocol that includes non-natural amino acids using the Rosetta platform[321]. Incorporating this approach into our CDSM protocol could provide valuable additional insights to complement our successful findings. To increase proteolytic stability, cyclization of promising linear peptides is a popular approach. Several macrocycles with inhibitory activity against M^{pro} have been reported[314,315]. Macrocyclization not only provides greater proteolytic stability but also improves membrane permeability[322]. Johansen-Leete et al. (2022) designed a cyclic peptide inhibitor whose cellular uptake was further improved by conjugating it to a 6-amino-acid cell-penetrating peptide, resulting in improved antiviral activity against M^{pro} in cell tests ($EC_{50} = 6.6 \pm 0.4\,\mu M$)[315]. Our linear c04 peptide is highly hydrophobic, with its Arg at P3 conferring a positive

charge. Positively charged peptides can enter cells through clathrin-mediated endocytosis or clathrin-independent pathways. While a certain degree of hydrophobicity improves peptide membrane permeability, excessive hydrophobicity can lead to aggregation and cellular toxicity[308,323]. Therefore, it would be advisable to consider macrocyclization and other methods to improve both proteolytic stability and cell penetration.
In addition to these potential pharmacokinetic limitations, protease inhibitors with high potency typically exhibit IC_{50} values in the nanomolar to low micromolar range[324]. Thus, although c04 appears to be a promising candidate, further modifications are required to enhance its drug potency and pharmacokinetics. Beyond that, our proposed strategies, particularly the CDSM protocol, have demonstrated great potential for CAPID and broad applicability to diverse peptide–protein complexes.

5. A Modeling Workflow for PSMA-targeted Radiopharmaceuticals

The prostate specific membrane antigen (PSMA), which is generally overexpressed in prostate cancer cells, represents an attractive target for radiopharmaceuticals[178]. Rational design of PSMA-targeted radioligands requires understanding the structure–activity relationship. We therefore aim to develop a computational workflow that accurately predicts PSMA–ligand interactions and elucidates correlations between ligand structural features, their binding affinity and pharmacokinetic properties. This workflow builds upon our initial docking studies of two PSMA radioligands, [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA and [^{99m}Tc]Tc-PSMA-GCK01, published in Publication II (Section A.3):

The Theranostic Optimization of PSMA-GCK01 Does Not Compromise the Imaging Characteristics of [^{99m}Tc]Tc-PSMA-GCK01 Compared to Dedicated Diagnostic [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA in Prostate Cancer
Eduards Mamlins, Lara Scharbert, Jens Cardinale, Maria Krotov, Erik Winter, Hendrik Rathke, Birgit Strodel, Alfred O. Ankrah, Mike Sathekge, Uwe Haberkorn, Clemens Kratochwil, and Frederik L. Giesel
Molecular Imaging and Biology, 2023, DOI:10.1007/s11307-023-01881-y.

Our computational workflow involves docking a selection of clinically relevant PSMA ligands into the PSMA binding site, followed by molecular dynamics (MD) simulations to verify and evaluate ligand poses and their interactions with the binding site. One limitation of PSMA radiopharmaceuticals is their relatively high kidney uptake, driven by the abundant expression of PSMA in this non-target organ[325]. By comparing the binding modes of fast-, slow-, and practically non-clearing PSMA ligands, we seek to identify the chemical and structural properties underlying these distinct behaviors. The findings of this work will serve as the foundation for future computer-aided radiopharmaceutical design (CARD) studies.

5.1. Methods

A comprehensive description of the general methods used in this project is provided in Section 2.1. In the following, we outline the project-specific procedures in detail.

Structure Preparation. The PSMA crystal structures were prepared with the co-crystallized ligands PSMA-1007 (PDB: 5O5T) and PSMA-617 (PDB: 8BOW) present, using the protein preparation wizard in Maestro (Schrödinger)[224]. The structures of the residual PSMA ligands, PSMA-GCK01, DCFPyL, PSMA-11, PSMA-I&T, PSMA-I&S and rhPSMA-7 and PSMA-BB, were either derived from the PubChem database[326] or generated manually using the Schrödinger 2D Sketching tool implemented in Maestro (Fig. 5.5). To ensure optimal bond lengths and angles within the three-dimensional metal-coordinating chelator, we obtained the crystal structures of gallium(III) 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid (DOTA)[327] (CSD entry: DEVHIW) and technetium(V) mercaptoacetyl triglycine (MAG_3)[328] (CSD entry: YOMMUI01), and the nuclear magnetic resonance (NMR) structure of iron(III) N,N′-Bis(2-hydroxybenzyl)ethylenediamine-N,N′-diacetic acid ($HBED^-$) (CSD entry: KEKVOL)[329] from the Cambridge Structural Database (CSD)[330]. The MAG_3 structure was manually modified to mercaptoacetyl tri-D-serine (MAS_3) by substituting all glycine residues with D-serine. The DOTA and $HBED^-$ structures were used as basis for 1,4,7,10-tetraazacyclododececane,1-(glutaric acid)-4,7,10-triacetic acid (DOTAGA) and N,N′-Bis[2-hydroxy-5-(ethylene-β-carboxy)benzyl]ethylenediamine-N,N′-diacetic acid (HBED-CC), respectively.

Pose Generation. The selection of water molecules to be maintained in the PSMA binding pocket during docking was based on their conservation across different crystal structures. The PSMA structures in complex with urea-based inhibitors (PDB: 5O5T, 8BOW, 5O5U, and 5O5R) were first prepared with Maestro[224] and then superimposed. Only water molecules consistently observed at the same positions were retained, as shown for the PSMA–PSMA-1007 complex in Fig. 5.1A.
Different hydration levels were generated by progressively adding conserved water molecules to the binding site of each protein structure (PDB: 5O5T and 8BOW). Level 0 includes only the Zn^{2+}-coordinating water molecule. In level 1, the area surrounding the vector—the catalytic chamber—is solvated. Level 2 adds water molecules to the hollow cavity between the S1 glutamate pocket and the S1 accessory lipophilic region. Level 3 further includes water molecules between the S1 lipophilic pocket and the arene-binding site, and level 4 incorporates multiple water molecules at the entrance of the binding funnel.
Since the AutoDock4 (AD4) force field does not support elements such as Ga (Gallium) and Tc (Technetium), they were substituted with Fe (iron) and Mn (Manganese), respectively,

before preparing the structures for docking. The protein structures at varying hydration levels and the ligand structures were converted to the PDBQT format using the AutoDock 'prepare_receptor' and 'prepare_ligand' tools, respectively. For affinity map generation using AutoGridFR (AGFR), the docking box was centered on the co-crystallized ligand with a 4 Å padding. Docking was performed with AutoDockFR (ADFR) in triplicate, conducting 120 independent searches per replicate, each involving 5×10^6 evaluations of the scoring function.

Re-scoring. The top-10 docking solutions from each triplicate run were re-scored using seven key distances $d_{\text{vec},i}$ ($i = 1-7$), measured between selected atoms of the ligands' vector and specific binding-site residues in the PSMA crystal structures (Fig. 5.1B). Based on the mean distance $\langle d_{\text{vec}} \rangle = 4.5$ Å, observed for both co-crystallized ligands, PSMA-617 (PDB: 8BOW) and PSMA-1007 (PDB: 5O5T), a cutoff of 5.5 Å was applied for the re-scoring process. Docking solutions with $\langle d_{\text{vec}} \rangle < 5.5$ were then re-ranked according to their docking scores.

Parameterization. We performed different parameterization procedures for ligands with and without metals, and for the binuclear Zn^{2+} active site of PSMA. All quantum mechanical (QM) simulations were performed with Gaussian 09[331]. For the ligands without metal sites, we performed Merz-Kollman restrained electrostatic potential (RESP) charge calculations at the HF/6-31G* level of theory. For metal center parameterization we used

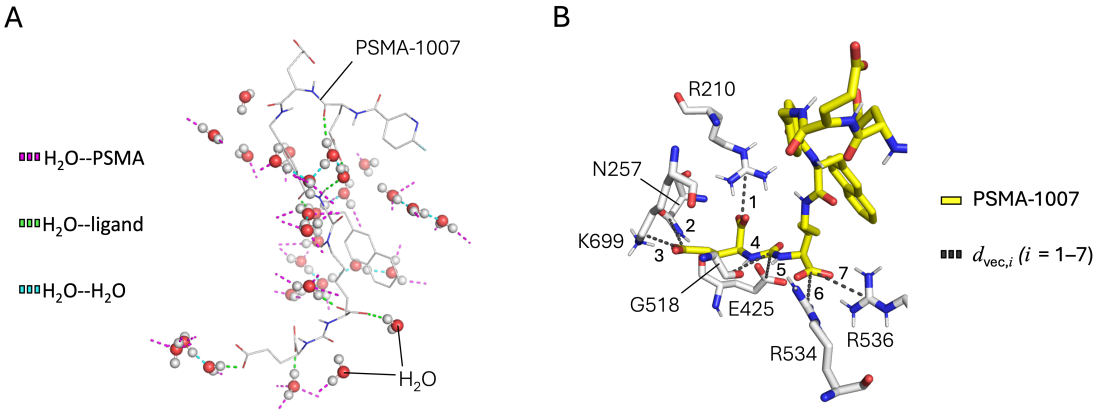


Figure 5.1.: Docking conditions and re-scoring parameters for PSMA: binding site water molecules and vector-to-residue distances. (A) Representative hydrogen bond network of binding site water molecules (spheres representaion) in the PSMA crystal structure (PDB: 5O5T) in complex with PSMA-1007 (line representaion), illustrating interactions with other water molecules, binding-site residues, and the ligand. (B) The seven distances $d_{\text{vec},i}$ (grey dashed lines) used for re-scoring are shown at the example of PSMA-1007 and PSMA binding-site residues (PDB: 5O5T), shown as yellow and white sticks, respectively.

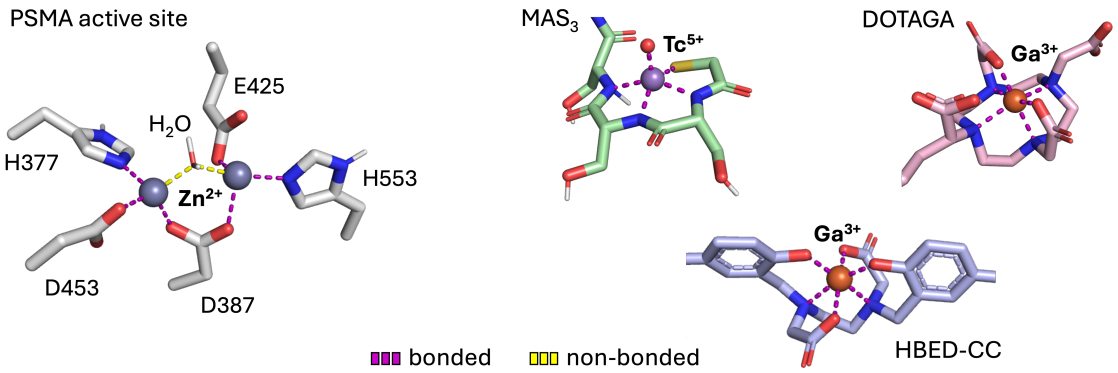


Figure 5.2.: Bonded and non-bonded modeling of the metal sites. The coordination of the metal centers are displayed for the binuclear Zn^{2+} active site of PSMA and chelators either labeled with Tc (MAS_3) or Ga (DOTAGA and HBED-CC). For the parameterization, we employed bonded (violet) and non-bonded (yellow) model approaches to describe the interactions between the metal ions and their ligands.

the MCPB.py[332] tool of AmberTools 23[333]. We employed the bonded model approach to describe the metal site in ligands labeled with Ga or Tc. For the parameterization of the binuclear Zn^{2+} active site in PSMA, we utilized a hybrid modeling strategy. Specifically, the interactions between the Zn ions and the coordinating PSMA residues (H377, D387, E425, D453 and H553) were treated as bonded, whereas the interactions involving the water molecule and the Zn ions were modeled as nonbonded (Fig. 5.2). Oxidation states of +5 for Tc, +3 for Ga, and +2 for each Zn ion were employed. QM geometry optimizations, force constant calculations, and Merz-Kollman RESP charge derivations were carried out at the B3LYP/6-31G* level of theory for all atoms, including Ga and Zn ions, with Tc as the sole exception using the LANL2DZ basis set. Finally, AMBER topology and coordinate files were generated using the LEaP module from AmberTools and subsequently converted into GROMACS-compatible input files with ACPYPE[334].

MD Simulations. The system setup, energy minimization, equilibration in NVT and NPT ensemble, and the production run was conducted as described in Section 2.1.3.3. Each PSMA–ligand system (∼97,000 atoms) was simulated for 100 ns using GROMACS version 2023[51,262]. Conformational clustering was applied with an root mean square deviation (RMSD) cutoff of 2 Å using the algorithm described by Daura et al. (1999)[283] to obtain the predominant ligand structure within the PSMA binding site. The binding free energy (ΔG_{bind}) between the ligands and PSMA (including Zn^{2+} ions) was calculated for 500 trajectory snapshots using the molecular mechanics Poisson–Boltzmann surface area (MM/PBSA) method as implemented in gmx_MMPBSA[59]. The temperature was set to 310 K and the ionic strength to 150 mM. Visualization and distance measurements of static structures, like crystal structures, were performed using PyMOL (Schrödinger)[240].

5.2. Results

5.2.1. Conformational Differences of Two PSMA Crystal Structures

A critical step of our modeling workflow is the selection of an appropriate target structure. Among our selection of clinically relevant PSMA radioligands, high-resolution crystal structures in complex with PSMA are available for PSMA-617 (PDB: 8BOW) and PSMA-1007 (PDB: 5O5T). Superposition of the three-dimensional structures reveals that the overall protein conformation is nearly identical between the two complexes (RMSD = 0.105 Å, aligned by C_α atoms), with the exception of a single loop region (residues Trp541–Gly548) that adopts a different orientation (Fig. 5.3).
The co-crystallized ligands PSMA-1007 and PSMA-617 share a similar structure, both featuring a Glu-ureido-based binding motif and a naphthyl linker group. The primary difference between them lies in their radiolabel, resulting in distinct interaction patterns with PSMA. While PSMA-1007 engages Trp541 within the arene-binding site (ABS) of PSMA through π-π stacking interactions via its FNA moiety, this binding site remains unoccupied in the case of PSMA-617, which features a DOTA chelator. The conformational rearrangement of the loop region spanning residues Trp541–Gly548, which leads to the opening of the ABS upon binding of an aromatic, π-π stacking-capable ligand such as PSMA-1007, has been described previously[186]. Since subsequent docking experiments

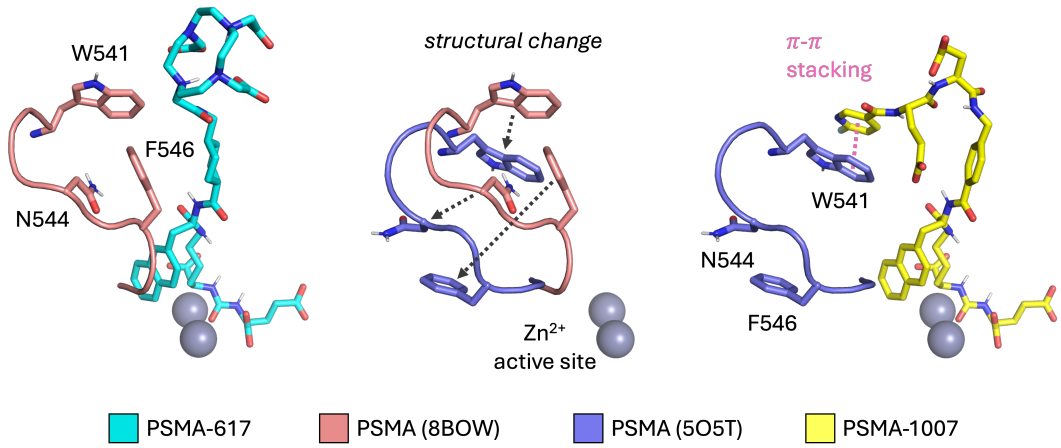


Figure 5.3.: Loop rearrangement in PSMA induced by PSMA-1007 opening the arene-binding site. The loop region (residues Trp541–Gly548) is shown for PSMA structures in complex with PSMA-617 (PDB: 8bow) and with the inhibitor PSMA-1007 (PDB: 5o5t). Arrows indicate the conformational shift of representative residues (Trp541, Asn544, and Phe546) in the superimposed PSMA structures. In the PSMA-1007 complex, the repositioned Trp541 enables π-π stacking with the 6-fluoronicotinic acid (FNA) moiety of the inhibitor.

will be carried out using a rigid receptor model, this structural difference in the loop region is important to take into account when evaluating the binding modes of PSMA ligands.

5.2.2. The Effect of PSMA Binding Site Hydration on Docking

In our primary studies, we used docking to identify the binding poses of two PSMA radioligands, [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA and [^{99m}Tc]Tc-PSMA-GCK01. We found that retaining water molecules within the binding site of PSMA facilitated the correct integration of the binding motif and linker group, resulting in relative docking scores that correlated well with empirical binding affinities[221].
The binding site of PSMA contains several water molecules that mediate hydrogen bonds, either bridging interactions between the ligand and protein residues or stabilizing intramolecular interactions within the protein (Fig. 5.1A). To further investigate the role of water molecules in ligand binding, evaluate the performance of the docking engine, and establish an optimized docking protocol for subsequent studies, we performed redocking experiments at varying hydration levels within the binding site of PSMA using the ADFR docking engine.
Two PSMA crystal structures in complex with PSMA-617 (PDB: 8BOW), with a non-labeled chelator, and PSMA-1007 (PDB: 5O5T) were used for the redocking approach. The co-crystallized ligands were prepared within their respective receptors, assigning protonation states as shown in the chemical structures in Fig. 5.5. To generate hydration levels 0 to 4, the PSMA binding site was systematically hydrated through the step-by-step addition of conserved water molecules. Level 0 contains no water molecules except for the Zn-coordinating water, while in level 1 the vector-accommodating site—the catalytic chamber—was filled. This procedure was continued up to level 4, in which nearly the entire binding funnel was filled with conserved water molecules.
The results of redocking PSMA-617 and PSMA-1007 at hydration levels 0 to 4 are shown in Fig. 5.4. In the following analysis, we focus on three regions of the ligand that are particularly relevant to their positioning within the PSMA binding site: (i) the Glu-ureido-based vector, which interacts with the catalytic chamber, comprising the S1′ glutamate pocket, the binuclear Zn^{2+} active site, and the S1 subsite with the arginine patch; (ii) the naphthyl linker group that occupies the S1 lipophilic region; and (iii) the chelator located at the entrance of the binding funnel or in the case of PSMA-1007, the FNA moiety, which interacts with the ABS.

The Vector. The majority of the top-ranked docking poses in Fig. 5.4 resemble the reference crystal structures, with the vector accommodating the catalytic chamber and the glutamic moiety of the Glu-urea-Lys motif positioned close to the S1′ glutamate pocket.

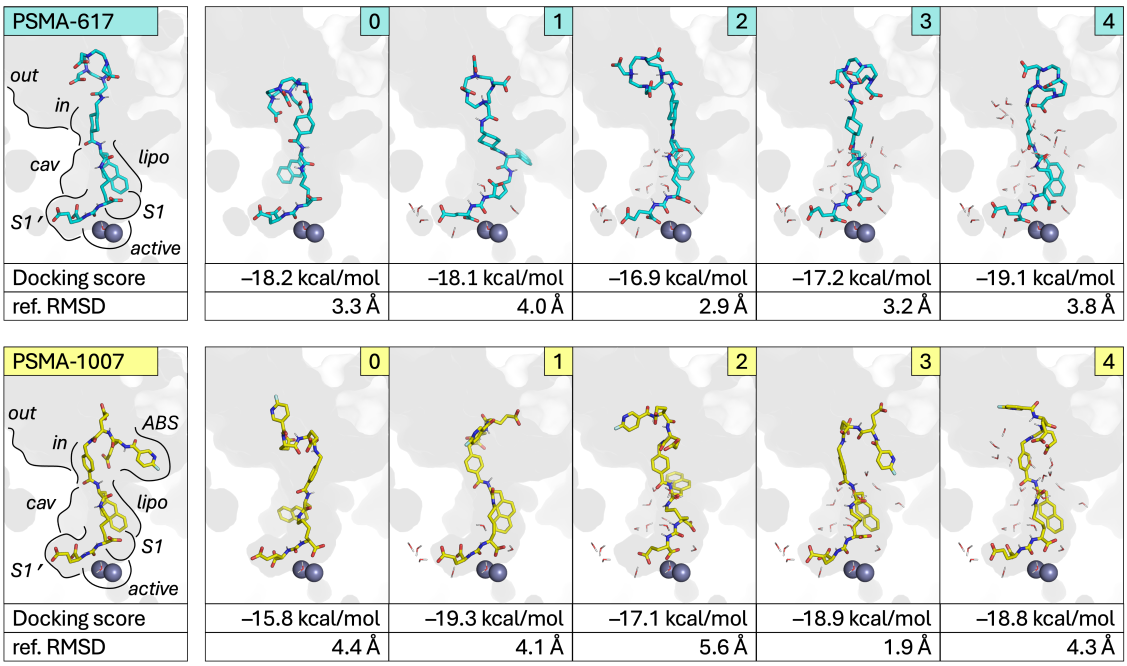


Figure 5.4.: Redocking of PSMA-617 and PSMA-1007 to the crystal structure of PSMA with different levels of hydration in the binding pocket. The non-labeled PSMA-617 (PDB: 8BOW) and PSMA-1007 (PDB: 5O5T) were redocked to the respective PSMA crystal structures. Top-ranked docking poses are shown, with binding site hydration levels indicated by numbers 0–4. For reference, the co-crystallized ligand structures are shown on the left, with labeled subsites: S1′ glutamate pocket (S1′), Zn^{2+} active site including residues interacting with the urea moiety (active), S1 subsite including arginine patch (S1), S1 accessory lipophilic region (lipo), hollow cavity between S1′ and lipo (cav), arene-binding site (ABS), and inner (in) and outer (out) area of the binding funnel entrance. Docking scores and RMSD values relative to the reference structures are provided below the docking models.

Solely PSMA-1007 at level 2 represents an exception: the vector is shifted toward the S1 region, and the glutamic moiety is positioned closer to the Zn ions. Otherwise, only minor variations within the binding motif are observed, including slight rotations of the carboxyl groups or the urea moiety, with the most pronounced rotation shown in PSMA-1007 at level 0, where the entire urea group is rotated by 180°.

The Naphthyl Group. For both PSMA-617 and PSMA-1007, we observed an influence of the hydration level on the positioning of the naphthyl group. At higher levels, the naphthyl group occupies the S1 lipophilic region, consistent with the reference structures. At lower hydration levels, however, it shifts toward the hollow cavity or the binding funnel entrance. When these alternative binding sites are occupied by water molecules, the naphthyl group is effectively forced into the vacant S1 lipophilic pocket. Moreover, PSMA-617 adopts a more extended conformation—similar to the reference structure—when the naphthyl group occupies the correct pocket, such as at levels 2–4.

The Radiolabel. In contrast to PSMA-1007, which occupies the well-defined ABS via its FNA moiety, the DOTA chelator of PSMA-617 is largely solvent-exposed at the spacious entrance of the binding funnel. Furthermore, the DOTA moiety could not be resolved in the crystal structure, indicating distinct flexibility[335]. This flexibility in the chelator region is reflected in the redocking results, which exhibit a high degree of pose variability. However, a consistent trend is observed across all water-content levels, with the top-ranked poses placing the chelator's aspartate moieties oriented in the same direction as in the crystal structure. This can be explained by the presence of three basic residues (K514, R511, and K610), which cause the surface of this PSMA region to be positively charged and enable the formation of hydrogen bonds with the DOTA chelator via its aspartate moieties.

Across all redocking solutions of PSMA-1007, the correct binding pose, with the FNA moiety engaging the ABS, was observed only once at level 3. The other poses show the FNA moiety adopting various orientations outside the binding funnel. The spacious entrance, offering numerous alternative poses of PSMA-1007, may explain why the correct positioning was generated so rarely.

Overall, the docking strategy using varying levels of hydration in the binding site of PSMA allowed the identification of at least one top-ranked structure that closely resemble the co-crystallized ligand pose. In particular, the positioning of the naphthyl linker group of both PSMA ligands was strongly influenced by the hydration levels. For PSMA-617, the vector and naphthyl group were both accurately positioned at the higher levels 2, 3, and 4, with the highest predicted affinity observed at level 4. For PSMA-1007, close alignment of the vector, naphthyl group, and chelator region with the reference structure was achieved exclusively at level 3. Notably, even though the difference in docking scores is minimal (0.4 kcal/mol), level 3 only has the second-highest predicted affinity, after level 1. Moreover, since the linker of clinically relevant PSMA ligands varies, and not all include a naphthyl group, the hydration level could affect each ligand differently. Thus, selecting a promising pose for subsequent MD simulations will require critical reassessment of the models when applying this docking strategy to the PSMA ligands discussed in the following section.

5.2.3. Pose Generation of Clinically Relevant PSMA Ligands

Following the successful redocking of PSMA-617 and PSMA-1007, we applied the same docking strategy to a selection of clinically relevant PSMA ligands with unknown experimental structures: PSMA-GCK01, DCFPyL, PSMA-11, PSMA-I&T, PSMA-I&S and rhPSMA-7. We also included PSMA-BB, a ligand with low inhibitory activity against PSMA (IC_{50}=108 nM), for comparison. Their chemical structures are shown in Fig. 5.5.

The ligands were docked in triplicates to both PSMA structures—co-crystallized with PSMA-617 (PDB: 8BOW) or PSMA-1007 (PDB: 5O5T)—at hydration levels 0–4 using the docking software ADFR.

The Re-scoring Strategy. During the redocking of PSMA-617 and PSMA-1007, the majority of top-ranked ligand poses positioned the binding motif correctly—resembling the crystal structure pose—within the catalytic chamber (Fig. 5.4). However, when docking

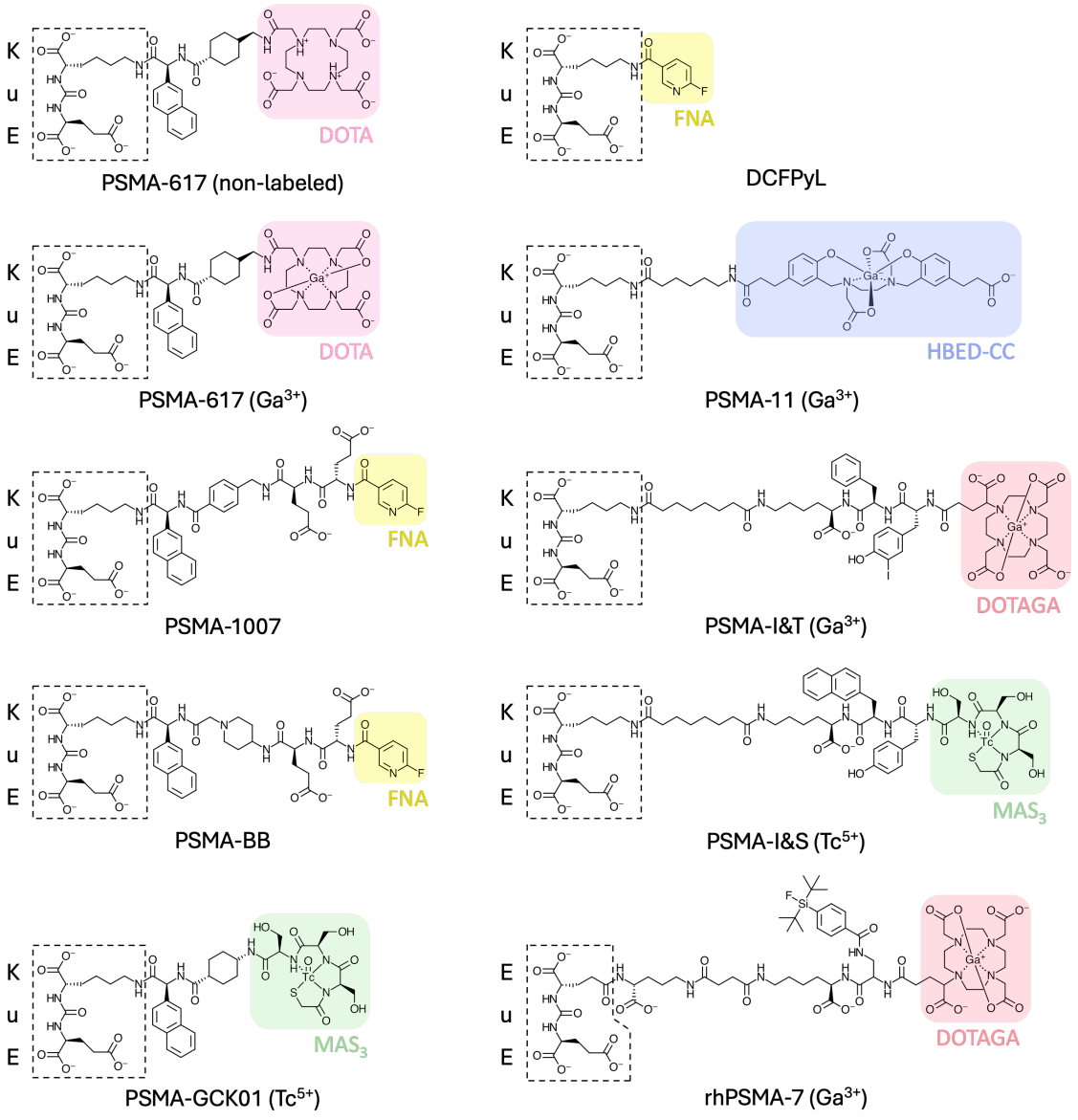


Figure 5.5.: Chemical structures of modeled PSMA ligands. The two-dimensional chemical structures of each modeled PSMA ligand are illustrated. The Glu-ureido based vectors, Glu-urea-Lys (KuE) or Glu-urea-Glu (EuE), are framed by a dashed line. The Ga-binding chelators are highlighted in pink (DOTA), blue (HBED-CC) and red (DOTAGA). The Tc-labeled chelator, MAS_3, and the F-labeled prosthetic group, FNA, are marked in green and yellow, respectively.

the other PSMA ligands, a correctly positioned vector was not consistently represented among the highest-scoring poses.
Therefore, the top-10 docking solutions from each run of the triplicate docking approach were subjected to a re-scoring procedure to determine the overall highest-ranked poses in which the Glu-ureido-based motif was correctly positioned within the catalytic chamber. To identify the desired binding pose, we defined seven distances between selected atoms of the binding motif and key binding-site residues of PSMA (Fig. 5.1B). During re-scoring, only poses with distances below a predefined cutoff were considered for re-ranking based on their docking scores.

The Docking Results. An overview of the top-ranked docking solutions after rescoring is shown in Fig. 5.6, with the corresponding docking poses displayed in Figs. 5.7–5.8 and A38–A43. The affinities for each PSMA ligand docked against both PSMA structures (PDB: 8BOW and 5O5T) at different hydration levels (L0–L4) are represented as a color map, including the redocking solutions from Section 5.2.2 for comparison. White tiles indicate conditions for which no pose within the top-10 solutions correctly positioned the binding motif in the catalytic chamber of PSMA. Comparing the results between both PSMA structures showed that the ligands docked against the PSMA structure co-crystallized with PSMA-617 tend to obtain slightly higher affinities compared to the structure that was co-crystallized with PSMA-1007 (Fig. 5.6).
The loop rearrangement that forms the ABS in PSMA (PDB: 5O5T) presents a dual effect: it may enhance affinity for specific aromatic moieties engaging this subsite, yet it also displaces Phe546, potentially reducing the hydrophobicity of the binding funnel entrance (Fig. 5.3). This observation may explain why larger ligands with aromatic moieties distant from the binding motif—such as PSMA-11, PSMA-I&T, PSMA-I&S, and rhPSMA-7—exhibit slightly improved docking performance in the PSMA structure lacking the ABS (PDB: 8BOW). In contrast, the smallest ligand, DCFPyL, which does not interact with the entrance region, demonstrates comparable performance in both structures.
The docking scores further indicate no discernible pattern between binding site hydration and ligand affinity; variations in water content—whether lower or higher—do not correlate with improved or reduced affinities.
With values ranging from -14.1 to -21.2 kcal/mol (for 8BOW), the highest affinities across all hydration levels were observed for the redocked ligands, PSMA-617 and PSMA-1007, PSMA-BB, PSMA-GCK01 and DCFPyL. Given that DCFPyL is the smallest ligand and the other ligands share a naphthyl-containing linker structure analogous to the co-crystallized ligands in the receptor, suggests that these 'favorable ligands' were able to effectively utilized the pre-formed binding pockets. Despite its structural differences from

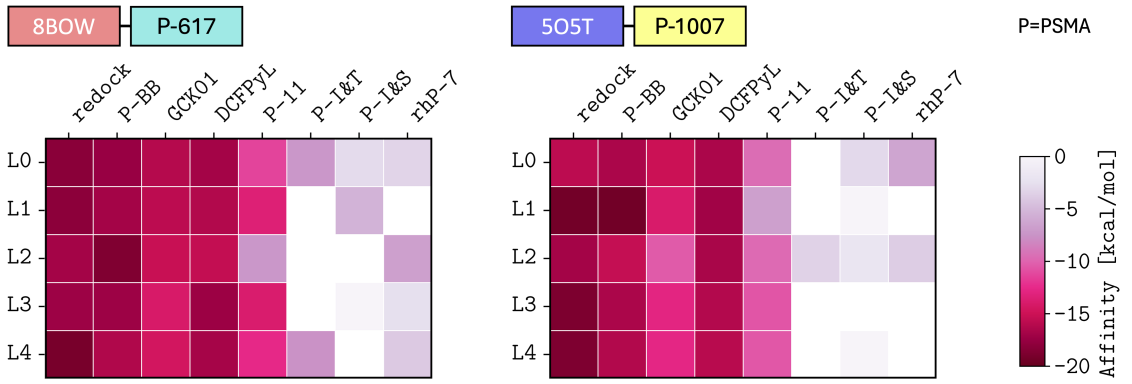


Figure 5.6.: Top-ranked docking scores of PSMA ligands. A selection of PSMA ligands (Fig. 5.5) were docked to the PSMA crystal structure, co-crystallized with PSMA-617 (PDB: 8BOW) and PSMA-1007 (PDB: 5O5T), using ADFR. The top-ranked docking scores after re-scoring are shown for each PSMA ligand and hydration level (L0–L4) as a color map. The results of the redocking (redock) of the co-crystallized ligands are shown for comparison. White fills indicate that no docking pose was found after re-scoring.

the 'favorable ligands', the minimalistic linker design of PSMA-11—composed of a single 6-aminohexanoic acid spacer, connecting the HBED-CC chelator to the Glu-urea-Lys motif—may explain the moderate affinities (from -13.9 to -7.3 kcal/mol for 8BOW).
The ligands PSMA-I&T, PSMA-I&S, and rhPSMA-7 yielded the poorest docking scores ranging from -7.6 to $+1.8$ kcal/mol (for 8BOW). For some hydration levels, not even a single docking pose in the top-10 solutions correctly positioned the Glu-ureido-based moiety within the catalytic chamber. These 'larger ligands' feature considerably longer and more complex linkers compared to the co-crystallized inhibitors. The expanded conformational search space, resulting from the greater degree of freedom in rotatable bonds, combined with an increased susceptibility to steric clashes, likely challenges the docking algorithm's ability to identify favorable binding poses. Optimal accommodation of these ligands may also require conformational rearrangements of the protein, which cannot be modeled in rigid-receptor docking.
In the following, we further discuss the conformational aspects of the docking results, focusing on the ligands interactions with distinct subsites, as part of the process to select the most promising ligand poses for subsequent MD simulations.

5.2.4. Evaluation of Ligand Poses for Subsequent MD Simulations

The selection of a docking pose as starting structure for MD simulations is a critical step. Starting structures that are already near the global free energy minimum enhance complex stability throughout the simulation, thereby facilitating accurate analysis of protein–ligand interactions. Moreover, standard MD simulations often struggle to overcome high energy barriers associated with larger conformational changes that may be required to refine unfavorable poses.

In order to find a favorable starting structure, we first performed docking at various hydration levels in the PSMA binding site to enhance the chances of producing the optimal ligand pose, while re-scoring of all docking poses ensured that only those with the vector correctly positioned within the catalytic chamber remained. The final selection process is now primarily guided by the docking score, prioritizing ligand poses with the highest predicted binding affinity, while also incorporating a rational assessment of ligand conformations.

A general observation of the docking poses reveal that the radiolabel—with the exception of the smallest ligand, DCFPyL—is consistently positioned at the entrance of the binding funnel (Fig. 5.7 & 5.8, Fig. A38–A43). This observation aligns with expectations, as the linkers were intentionally designed to be sufficiently long, while the binding funnel is too narrow to accommodate especially the larger chelators.

Having already confirmed through re-scoring that the catalytic chamber is occupied by the ligands binding motif, we will focus on two accessory pockets that are regularly targeted by PSMA ligands. Specifically, the S1 lipophilic region, which can for example be occupied by a naphthyl linker group, and the ABS, which can be engaged by the FNA moiety of PSMA-1007 (Fig. 5.4).

The S1 lipophilic pocket. Ligands sharing a linker structure similar to PSMA-617 and PSMA-1007, including the accessory naphthyl group adjacent to the binding motif, are expected to interact with the S1 lipophilic pocket.

As a representative example, Fig. 5.7 illustrates the docking results of PSMA-GCK01 at the PSMA binding site (PDB: 8BOW) across hydration levels 0–4. The highest predicted affinity of -16.1 kcal/mol was obtained at level 0. However, consistent with the redocking results in Section 5.2.2 and the docking poses of PSMA-BB (Fig. A39), positioning of the naphthyl group within the S1 lipophilic pocket occurs with increasing hydration levels. Among these poses, the level 2 structure exhibited the highest affinity (-15.3 kcal/mol) and was thus selected (labeled with a 'B'), alongside the top-ranked level 0 pose (labeled with an 'A'), as a starting structure for subsequent MD simulations.

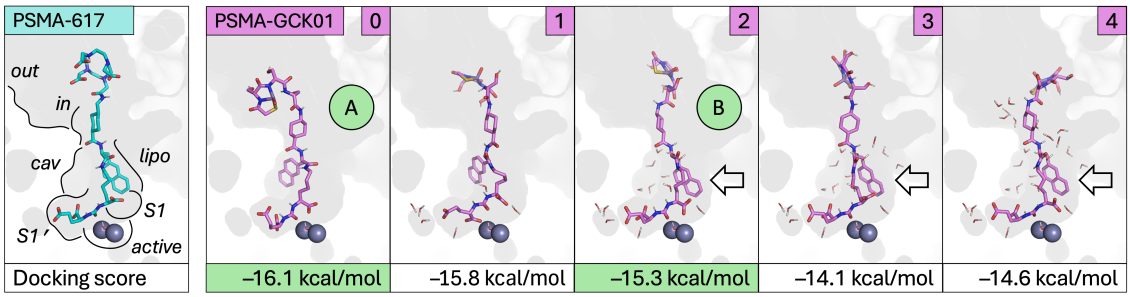


Figure 5.7.: Docking solutions of PSMA-GCK01. The ligand PSMA-GCK01 was docked into the binding site of the PSMA crystal structure (PDB: 8BOW) using ADFR. Re-scored top-ranked docking solutions are shown for each binding site hydration level ranging from 0–4. The docking scores are provided below the docking poses for each level. For reference, the co-crystallized ligand structure of PSMA-617 is displayed on the left including labeled binding subsites (Fig. 5.4). Poses selected for subsequent MD simulations are marked in green by (A) the best affinity and (B) a favorable conformation of the naphthyl group. The arrow indicates the naphthyl group occupying the S1 lipophilic pocket.

As the smallest ligand, comprising only a hydrophobic FNA moiety linked directly to the a Glu-urea-Lys binding motif, DCFPyL has the potential to interact with the S1 lipophilic pocket. However, none of the docking poses, positioned the FNA moiety at this subsite (Fig. A40). Instead, DCFPyL predominantly adopts an extended conformation within the binding funnel.

An additional S1 hydrophobic accessory pocket was identified through X-ray crystallography of PSMA in complex with the inhibitor DCIPzL (PDB: 3D7H)[185]. This 'R-patch pocket' is located within the S1 arginine patch, framed by residues Arg463, Arg536, Arg534, and Asp465, and accommodates the iodobenzene moiety of DCIPzL. Given the structural similarity between DCIBzL and DCFPyL, we hypothesize that DCFPyL may also engage this pocket via its FNA moiety.

However, DCFPyL docking yielded only extended conformations, with no poses occupying the R-patch pocket. This observation may be explained by structural alignment of the PSMA–DCIPzL crystal structure (PDB: 3D7H) with the docking receptors (PDB: 8BOW, 5O5T) via their C_α atoms (Fig. A44). The 3D7H structure exhibits high similarity to 8BOW (RMSD: 0.178 Å) and 5O5T (RMSD: 0.16 Å). The R-patch pocket is inaccessible to DCFPyL during docking because it is closed in the receptors used. A minor conformational shift of the Arg463 side chain appears sufficient to open the pocket and accommodate an aromatic moiety.

The Arene-Binding Site. PSMA-BB not only shares the naphthyl linker group with PSMA-1007 but also the FNA moiety, suggesting that it should occupy the ABS. However, none of the docking poses in Fig. A39 reflect this expected behavior. The challenge of

docking the FNA moiety within the ABS was previously observed during the redocking of PSMA-1007 (Fig. 5.4). This difficulty may arise from the spacious entrance of the binding site, which allows multiple ligand orientations. Another possible explanation lies in the structural differences between the linkers of PSMA-1007 and PSMA-BB, particularly in the region surrounding the piperidine moiety. These differences may prevent PSMA-BB from adopting the optimal conformation required for the FNA moiety to effectively engage the ABS.

The ligands PSMA-11, PSMA-I&T, PSMA-I&S, and rhPSMA-7 lack functional groups—such as the naphthyl linker moiety—near the binding motif capable of engaging S1-site proximal accessory pockets. Instead, they feature aromatic groups designed to interact with the remote ABS.

For PSMA-11, its lipophilic HBED-CC chelator is proposed to engage this pocket via its phenol rings[198,336]. However, the docking results of PSMA-11 exclusively showed extended poses, with the chelator oriented away from the binding entrance. Notably, no poses positioned the HBED-CC moiety within the ABS (Fig. A41).

We assume that the HBED-CC chelator, including its propionic acid moiety attached to the outer phenol ring, requires more space than the pocket defined by the co-crystallized PSMA-1007 structure offers (Fig. A45). The FNA moiety of PSMA-1007 is flanked by Arg511 and Trp541 on opposite sides, while the fluoric site is constrained by a β-sheet structure comprising residues 536–538 and 461–463. Accommodation of larger groups, such as a HBED-CC chelator, would thus necessitate backbone adaptations of the receptor.

PSMA-I&T and PSMA-I&S both feature a long D-amino acid-substituted peptidyl spacer linking the vector to a DOTAGA and MAS_3 chelator, respectively. These aromatic residues were introduced distal to the binding motif to enhance interaction with the ABS[337]. The docking algorithm generally struggled to generate poses of PSMA-I&T with the vector positioned within the catalytic chamber (Fig. A42), likely due to its bulky structure—comprising a long spacer, two aromatic residues, and a large DOTAGA chelator. The limited number of poses obtained after re-scoring further reduced the possibility to position any aromatic residue of the ligands at the ABS.

In contrast, re-scoring yielded several poses for PSMA-I&S, including one in which the D-Tyr residue extends into the ABS. The docking results for PSMA-I&S are presented exemplarily in Fig. 5.8. The level 0 pose, featuring this promising conformation, was selected for further analysis via MD simulations as run B, alongside the level 1 pose (run A), which exhibited the highest docking affinity.

For rhPSMA-7, the silicon–fluoride–acceptor (SiFA) moiety was intended to engage the ABS[338]. While the docking engine generated some poses with the ligands vector within the catalytic chamber, the SiFA did not occupy the ABS (Fig. A43).

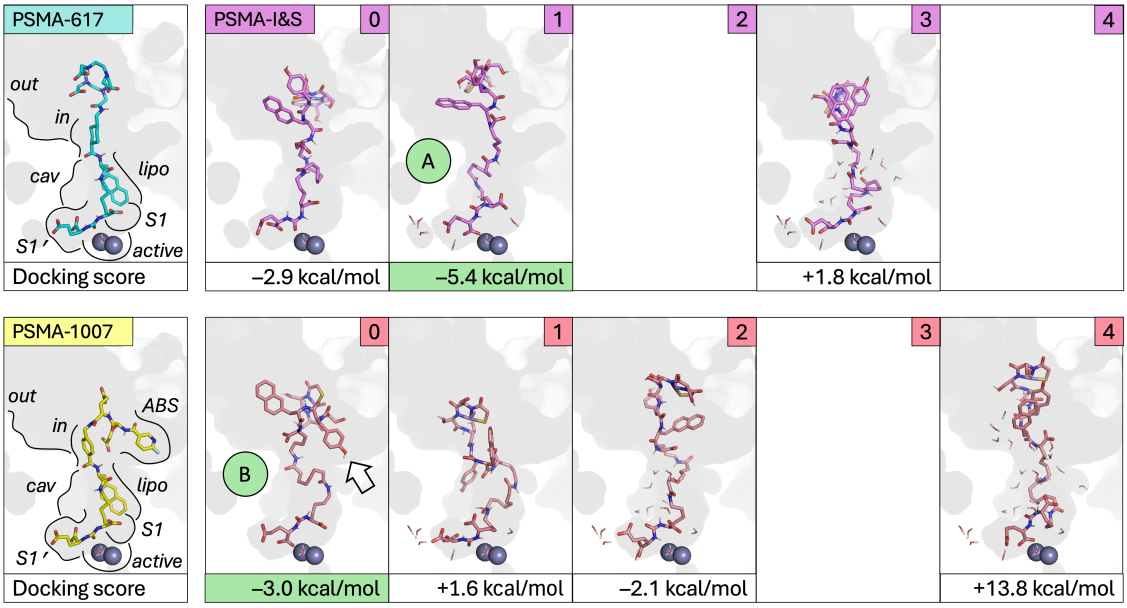


Figure 5.8.: Docking solutions of PSMA-I&S. The ligand PSMA-I&S was docked into the binding site of two different PSMA crystal structures (PDB: 8BOW (top row), 5O5T (bottom row)) using ADFR. Re-scored top-ranked docking solutions are shown for each binding site hydration level ranging from 0–4. The docking scores are provided below the docking poses for each level. For reference, the co-crystallized ligand structures of PSMA-617 and PSMA-1007 are displayed on the left including labeled binding subsites (Fig. 5.4). Poses selected for subsequent MD simulations are marked in green by (A) the best affinity and (B) a favorable conformation. The arrow indicates the tyrosine residue at the arene-binding site.

We assume that larger aromatic residues in the ligands—such as the HBED-CC chelator in PSMA-11, the iodine-substituted D-Tyr in PSMA-I&T, the naphthyl group in PSMA-I&S, or the SiFA moiety with its voluminous dimethylsilyl fluoride group—may induce steric clashes when positioned within the confined ABS (Fig. A45). Combined with the spacious entrance of the binding funnel and the high rotational freedom of these bulky ligands, this likely explains the rare occurrence of poses where the aromatic moieties occupy the ABS. By docking PSMA with clinically relevant ligands at different hydration levels in the binding site, we identified at least one pose for each ligand that met the criteria for a correctly positioned binding motif. The smaller, 'favorable' ligands, which share key features with the co-crystallized ligands in the PSMA structures we used for docking, showed particularly strong affinities. In contrast, re-scoring revealed that larger ligands produced fewer poses that meet the criteria and generally exhibit lower affinities. This suggests that the binding site of the PSMA crystal structures may be too narrow or not well-adapted to accommodate them. In the following section, we address the limitations of rigid receptor docking by employing MD simulations to study protein–ligand interactions dynamically, providing insights into complex stability and affinity.

5.2.5. Dynamic Investigation of PSMA–Ligand Interactions

To assess the previously generated PSMA–ligand complexes and investigate their molecular interactions, MD simulations were employed. These simulations enable the evaluation of ligand dynamics within the binding site, thereby providing critical insights into the stability and affinity of the complexes. MD simulations of PSMA in complex with radiopharmaceuticals present some unique challenges. Since the PSMA ligands bind to the active site, the binuclear Zn^{2+} center must be modeled accurately. This involves parameterization of the two Zn ions, each coordinated by a histidine (His553 or His377), an aspartate (Asp453) or glutamate (Glu425), along with two bridging ligands—an aspartate (Asp387) and a water molecule. Similarly, ligands containing chelators such as MAS_3, DOTA, DOTAGA, and HBED-CC also required parameterization. For the metal sites in the ligands, we employed a bonded model approach, whereas a hybrid model was used for the Zn^{2+} active site in PSMA. In the hybrid approach, interactions with coordinating residues were treated as bonded, while interactions with the water molecule were described as nonbonded (Fig. 5.2).

5.2.5.1. Evaluation of PSMA–Ligand Interactions

The co-crystallized structures of PSMA-617 (PDB: 8BOW) and PSMA-1007 (PDB: 5O5T), as well as the most favorable ligand docking poses from the previous section were submitted to 100 ns MD simulations (Table 5.1). The protonation states assigned to the ligand structures prior to simulation are shown in Fig. 5.5.
To characterize PSMA–ligand interactions, we first identified residues located within 3.5 Å of the ligand for at least 80 % of the simulation time. The residues collected across all MD simulations were defined as the PSMA binding site. Further analysis of these binding-site residues and their interaction with the ligand included per-residue decomposition of ΔG_{bind} using the MM/PBSA method, hydrogen bonding, and distance measurements. The results for each PSMA–ligand complex are shown in Fig. 5.9.
The per-residue ΔG_{bind} decomposition reveals that residue with basic side chains contribute most favorably to binding affinity, whereas acidic residues are most unfavorable, with values ranging from -100 to $+100$ kcal/mol, as shown in Fig. A52. This observation aligns with expectations, as all ligands carry a net negative charge (Table 5.1), making electrostatic interactions with positively charged PSMA residues energetically favorable. Charged residues are mainly located in the catalytic chamber and at the outer entrance of the binding funnel, with acidic residues in the majority. The S1 lipophilic pocket, the hollow cavity and the inner entrance, in contrast, mostly contain polar and non-polar residues, which contribute little to the binding energy and rarely form direct hydrogen

Table 5.1.: Summary of PSMA–ligand MD simulations: Starting conditions and results of binding free energy calculations.. For each MD simulation, the following details are provided: ligand name, bound metal, PDB ID of the PSMA crystal structure, selected starting pose (co-crystallized or docked at varying hydration levels), total ligand charge, and the binding free energy (ΔG_{bind}), including standard deviations, of the PSMA–ligand complex. MD runs of the same ligand started from docking poses with (A) the best affinity and (B) a favorable ligand conformation. The asterisk (*) denotes the modified co-crystallized PSMA-617 structure with the non-labeled (nl) chelator replaced by Ga(III) DOTA. The final column references the figure displaying the MD simulation results.

run	ligand name	metal	PSMA (PDB)	ligand pose	ligand charge	ΔG_{bind} (kcal/mol)	Fig.
nl	PSMA-617	none	8BOW	crystal	−4	−101.7 ± 4.6	A46
	PSMA-617	Ga	8BOW	crystal*	−3	−103.8 ± 3.5	A46
	PSMA-1007		5O5T	crystal	−5	−94.6 ± 4.4	A47
	PSMA-BB		8BOW	level 2	−5	−107.1 ± 2.4	A48
A	PSMA-GCK01	Tc	8BOW	level 0	−3	−115.1 ± 2.6	5.10
B	PSMA-GCK01	Tc	8BOW	level 2	−3	−100.7 ± 2.4	5.10
	DCFPyL		8BOW	level 3	−3	−66.7 ± 0.6	A49
	PSMA-11	Ga	8BOW	level 3	−5	−89.0 ± 0.9	A50
	PSMA-I&T	Ga	8BOW	level 4	−5	−118.3 ± 5.4	A51
A	PSMA-I&S	Tc	8BOW	level 1	−4	−119.0 ± 4.1	5.11
B	PSMA-I&S	Tc	5O5T	level 0	−4	−111.7 ± 4.6	5.11
	rhPSMA-7	Ga	8BOW	level 2	−6	−145.3 ± 2.6	5.12

bonds with the ligand.

Across all PSMA–ligand complexes, residues within the vector region maintain a low average minimum distance (< 4 Å) to the ligand, reflecting stable binding of the ligand's binding motif. This observation is further supported by the robust interactions these residues establish with the ligand. With the exception of the Zn-binding residues Glu424, Glu425, and Asp453, as well as the Arginine patch-stabilizing Glu457, the residues in this region almost entirely contribute favorably to the ΔG_{bind}. Most ligands consistently form hydrogen bonds with residues in the S1′ glutamate pocket (Arg210, Asn257, Lys699, and Tyr700), the active site (Gly518 and Tyr552), and the S1 subsite (Arg534 and Arg536). Specifically, the vectors' Glu residue interacts with Asn257 and Lys699 through its side chain, and with Arg210 and Tyr700 via its terminal carboxy group. The urea moiety adopts a *trans,trans* conformation, here referred to as the conserved urea conformation due to its recurrence in crystal structures of PSMA with Glu-ureido-based ligands[185,186,189,335]. This urea conformation is stabilized by hydrogen bonding through its carbonyl oxygen with Tyr552 on one side, while its two nitrogen (NH) atoms form hydrogen bonds with the amide carbonyl of Gly518 on the other side. The terminal carboxy group of the vectors' Lys residue is stabilized at the S1 subsite by the residues Arg534, Arg536 and Asn519.

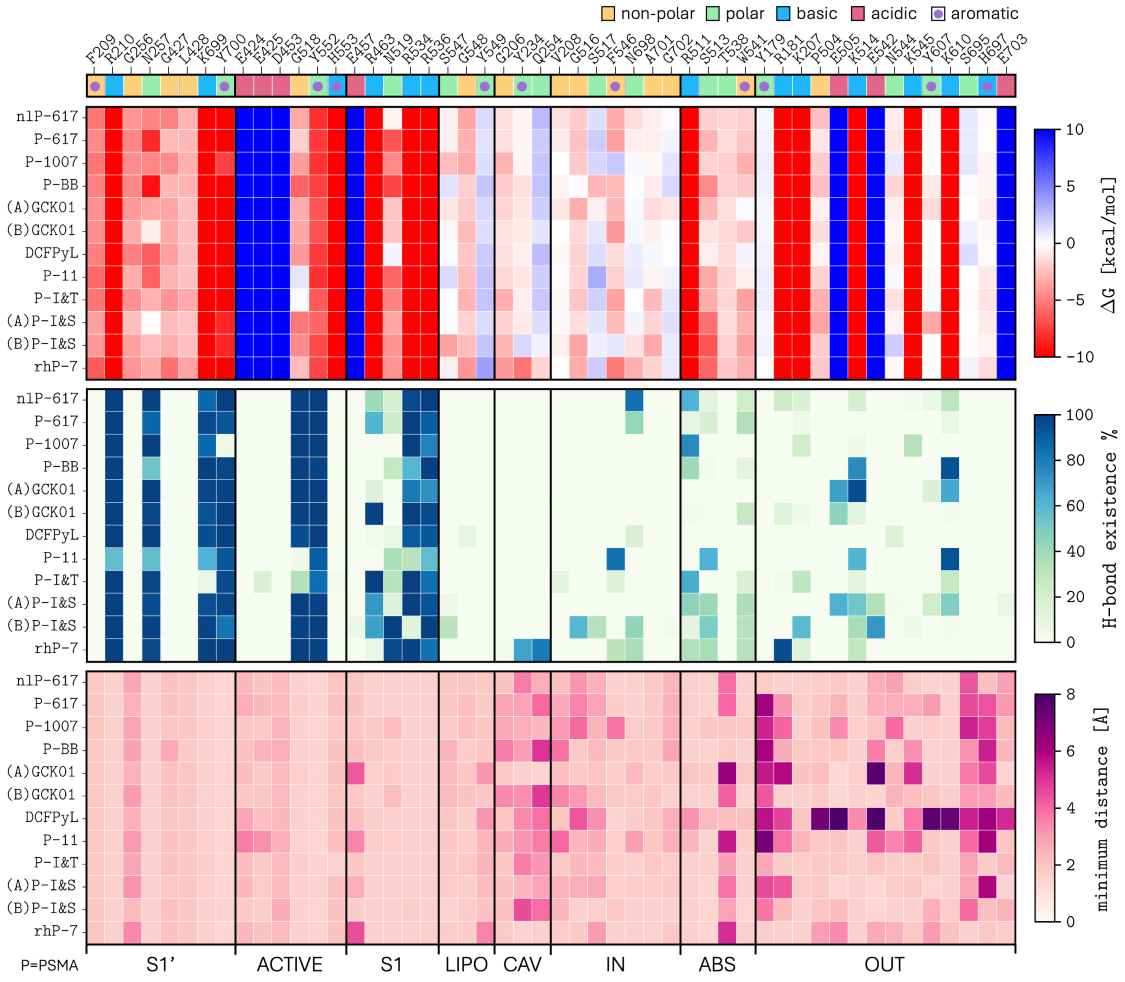


Figure 5.9.: Interaction analysis of PSMA–ligand MD simulations. We performed 100 ns MD simulations of PSMA in complex with several ligands: the co-crystallized structures of PSMA-617 (non-labeled (nl) and Ga-labeled) and PSMA-1007 (PDB: 5O5T), the docking-derived starting poses of PSMA-BB, DCFPyL, PSMA-GCK01, PSMA-11, PSMA-I&T, PSMA-I&S and rhPSMA-7. MD runs of the same ligand started from docking poses with (A) the best affinity and (B) a favorable ligand conformation. Binding-site residues were defined as those within 3.5 Å of the PSMA ligands for at least 80 % of the simulation time. These residues were categorized based on their location within the binding site: the S1′ glutamate pocket (S1′), the active site including residues interacting with the urea moiety (ACTIVE), the S1 subsite including the arginine patch (S1), the S1 subsite including arginine patch (S1), the S1 accessory lipophilic region (LIPO), the hollow cavity between S1′ and LIPO (CAV), the arene-binding site (ABS), and the inner (IN) and outer (OUT) area of the binding funnel entrance. Amino acid side chains are color-coded by property: non-polar (yellow), polar (green), basic (blue), and acidic (red). Aromatic residues are highlighted with a purple dot. For each PSMA–ligand complex, the per-residue decomposition of ΔG_{bind} (kcal/mol) was computed and is shown for the binding-site residues in the top color map. The central plot displays the percentage of simulation time during which at least one hydrogen bond (h-bond) was formed between each binding site residue and PSMA ligand. In the bottom color map the minimum distance (Å) between PSMA ligands and binding-site residues is represented as an average over the entire MD trajectory.

While Arg534 and Arg536 are more persistently involved in interactions with the ligands, the third member of the arginine patch, Arg463, and the adjacent Asn519 participate less frequently in hydrogen bonding. Beyond acidic residues, the most consistent and favorable contribution to ΔG_{bind} across MD simulations arises from the polar aromatic residues Tyr700 and Tyr552.

5.2.5.2. Conformational Analysis of the Binding Mode

Structural analysis of the MD trajectories involved RMSD-based clustering of the ligand to identify its predominant pose, as well as calculation of the root mean square fluctuation (RMSF) to assess its flexibility within the PSMA binding site.

The Catalytic Chamber. Simulations of the co-crystallized ligands PSMA-617 (Fig. A46) and PSMA-1007 (Fig. A47) demonstrate that the binding motif retains the crystal structure conformation with minimal flexibility (< 1.5 Å). This conformational stability was also observed in simulations of other PSMA ligands, including PSMA-GCK01 (Fig. 5.10), DCFPyL (Fig. A49), and PSMA-BB (Fig. A48). Even when the vector was initially suboptimally positioned within the binding site—such as with a twisted urea moiety or a misoriented carboxy group—the simulations typically corrected the vector conformation to resemble that of the co-crystallized ligands. This behavior was observed, for example, in the simulations of PSMA-I&S (Fig. 5.11, runs A and B) and rhPSMA-7 (Fig. 5.12). In some cases, such as the PSMA-11 (Fig. A50) and PSMA-I&T (Fig. A51) simulations, the vector adopted and maintained a conformation distinct from the co-crystallized reference structures, resulting in reduced hydrogen bonding with binding-site residues of the catalytic chamber. Gly518 usually stabilizes the urea moiety in its conserved *trans,trans* conformation through hydrogen bonds between its backbone carbonyl and both nitrogen (NH) atoms. However, this interaction is disrupted by PSMA-11 and PSMA-I&T due to the altered *trans,cis* conformation of the urea moiety. The resulting reduction in hydrogen bonding with Gly518 further leads to less favorable or even ΔG_{bind} contributions. This conformational variation in the urea moiety, especially the opposite orientation of the carbonyl group, may explain the distinct behavior of PSMA-11, where the ligand, including its vector, shifts toward the binding site entrance, thereby increasing its separation from residues of the catalytic chamber.
A notable observation is that PSMA-1007 is the only ligand that does not establish hydrogen bonds with Tyr700, which may be associated with a less favorable binding free energy (ΔG_{bind}) contribution from this residue compared to the other co-crystallized ligand PSMA-617. This absence of hydrogen bonding is linked to a displacement of Tyr700 away from the ligand's binding motif.

This shift can be explained by the rigid, linear conformation of the aromatic linker in PSMA-1007, where the benzene ring and adjacent amide bonds are oriented parallel to the Tyr700 side chain, thereby generating a hydrophobic environment. Consequently, this rearrangement disrupts the hydrogen bond between the benzoyl-adjacent carbonyl and Asn698—a bond that remains stable in simulations of the non-labeled PSMA-617.
Such a distinct conformational change from the protein crystal structure—even involving backbone movement—might suggest instability of the protein–ligand complex. Nevertheless, PSMA-1007 remains stably bound within the binding site, exhibiting minimal flexibility across the entire ligand structure.

The S1 Accessory Lipophilic Pocket. The structural analysis of PSMA-GCK01 and its interactions with binding-site residues starting from two different docking poses—the top-ranked hydration level 0 (run A) and the level 2 pose with the correctly positioned naphthyl group (run B)—are illustrated in Fig. 5.10. The predominant pose of run A exhibits minimal deviation from the starting pose, with its stability confirmed by its persistence for 99 % of the simulation and the low RMSF observed across the entire ligand. In contrast, the ligand structure in run B demonstrates reduced stability, characterized by high flexibility concentrated in the chelator, while the vector and linker remain more rigid. Notably, the naphthyl group in run A shifts slightly toward the S1 lipophilic pocket but never fully occupies it. However, when the naphthyl linker group is initially positioned within this pocket, it maintains a stable conformation. This behavior was also observed in simulations of the co-crystallized ligands PSMA-617 and PSMA-1007. The simulation of PSMA-BB, however, reveals that a starting pose with a naphthyl linker group only partially accommodated by the S1 lipophilic pocket can be unstable, resulting in the naphthyl group vacating this subsite (Fig. A48).
The per-residue decomposition of ΔG_{bind}, visualized by color mapping onto binding-site residues depicted as sticks, highlights key residues involved in ligand interaction. This representation underscores the dominant contribution of positively charged residues—particularly those interacting with the vector (Arg210, Lys699, Arg534, and Arg536)—to the overall binding affinity. The slight tilt observed in the predominant pose of PSMA-GCK01 in run B may be encouraged by Arg463, which forms persistent hydrogen bonds with the carbonyl group in the linker. The chelator in run A exhibits a considerably more stable position, interacting strongly with the residues Lys514 and Lys610 at the binding funnel entrance via hydrogen bonding.

The Arginine-Patch Pocket. MD simulations of an extended DCFPyL starting pose (Fig. A49) resulted in minimal conformational change, preventing the ligand from

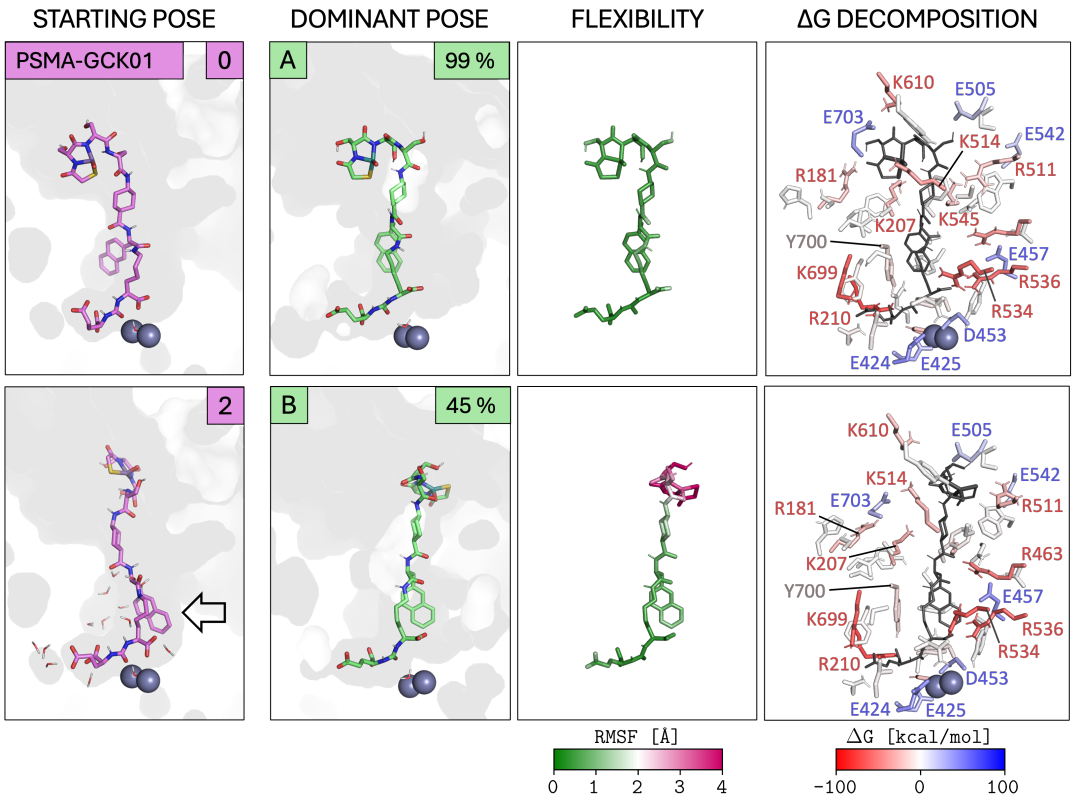


Figure 5.10.: Flexibility and interaction pattern of PSMA-GCK01. The top-ranked docking structure of PSMA-GCK01 from hydration level 2 (A) and the alternative pose from level 2 (B), with the naphthyl group binding to the S1 lipophilic pocket (arrow), were used as starting structures for 100 ns MD simulations in complex with PSMA (PDB: 8BOW). The predominant ligand conformations of the trajectories starting with the level 0 (99 %) or level 2 (45 % sampled) pose were identified through RMSD-based clustering. Ligand flexibility, as assessed by RMSF, is represented by a color shift ranging from rigid (green) to flexible (pink). The per-residue ΔG_{bind} decomposition is displayed for the PSMA binding-site residues defined by their presence of at least 80 % within 3.5 Å of all modeled PSMA ligands.

effectively engaging the R-patch pocket, as found for DCIPBzL (Fig. A44). We assume that DCFPyL is trapped in a local minimum, preventing the structural change required for binding the R-patch pocket. The Lys side chain of the vector would need to bend to allow the FNA moiety to engage this subsite. The energy barrier for a conformational change within the protein, however, is expected to be moderate, as it involves only a minor 2 Å shift of the Arg463 side chain without backbone movement. Moreover, while the other residues of the arginine patch (Arg534 and Arg536) consistently form hydrogen bonds with the carboxy group of the Lys vector residue, Arg463 exhibits notable adaptability, engaging different ligand regions across the PSMA–ligand MD simulations. Frequent hydrogen bonding (occupancy > 50 %) was observed for the Ga-labeled DOTA chelator in PSMA-617, as well as for the outer amide carbonyl of the suberic acid spacer (run A) and the carboxy group of the Lys spacer (run B) in PSMA-I&S simulations. Permanent

hydrogen bonds were also detected for the amide carbonyl adjacent to the cyclohexane linker moiety of PSMA-GCK01 (run B) and, analogous to PSMA-I&S (run B), the carboxy group of the Lys spacer in PSMA-I&T (Fig. 5.9).

The Arene-Binding Site. We performed MD simulations of PSMA-I&S in complex with PSMA starting from two different docking poses, the top-ranked hydration level 1 (run A) and the level 0 pose (run B) with the D-Tyr residue projecting into the ABS (Fig. 5.11). The RMSF analysis reveals pronounced flexibility throughout the linker, suggesting conformational instability of the ligand. This instability likely explains its inability to sustain stable interactions with the ABS. In run A, the ligand never engages with the ABS at all. This lack of interaction with the ABS is consistent across MD

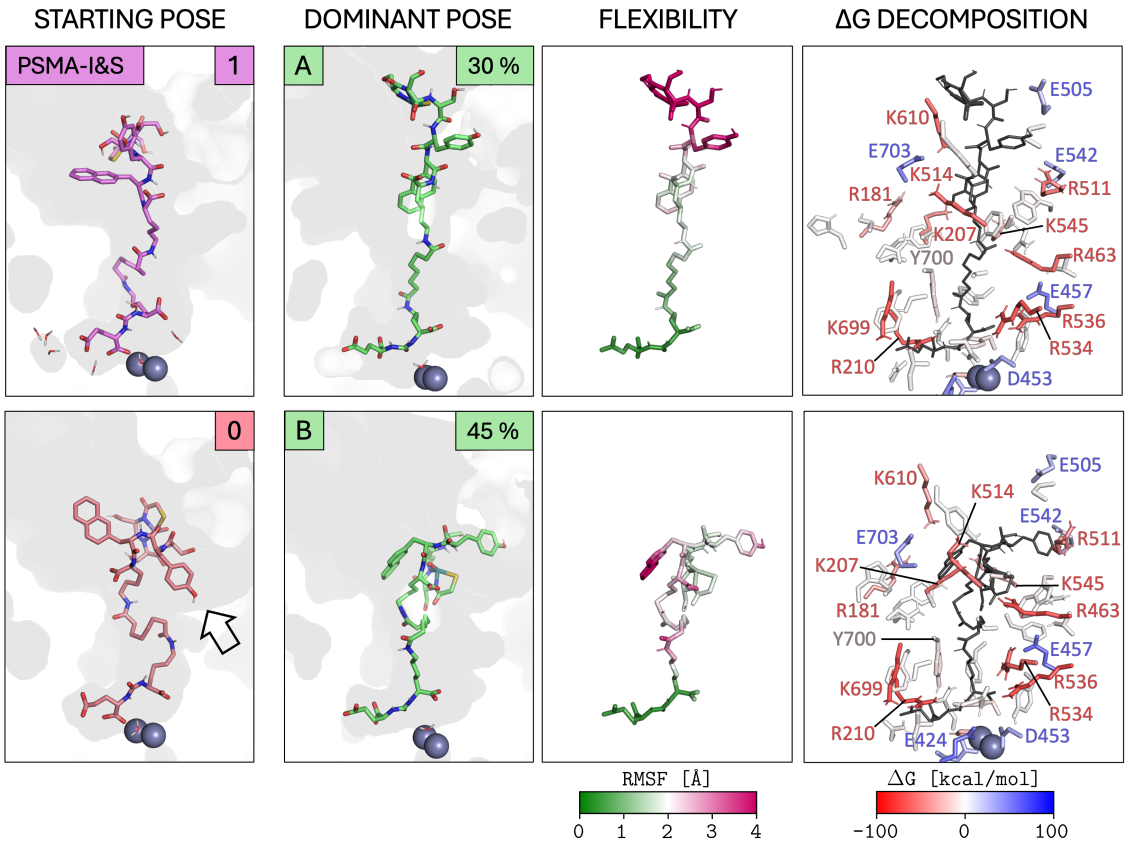


Figure 5.11.: Flexibility and interaction pattern of PSMA-I&S. The top-ranked docking pose of PSMA-I&S in complex with PSMA from hydration level 1 (A; PDB: 8BOW) and the alternative pose from level 0 (B; PDB: 5O5T), with its D-Tyr residue at the arene-binding site (arrow), were used as starting structures for 100 ns MD simulations. The predominant ligand conformation (18 % of the simulation) was identified through RMSD-based clustering. Ligand flexibility, as assessed by RMSF, is represented by a color shift ranging from rigid (green) to flexible (pink). The per-residue ΔG_{bind} decomposition is displayed for the PSMA binding-site residues defined by their presence of at least 80 % within 3.5 Å of all modeled PSMA ligands.

simulations of PSMA-11, PSMA-I&T, and rhPSMA-7—ligands designed to engage this pocket through their aromatic linker or chelator moieties[336–338]—as well as PSMA-BB, which is assumed to interact via its FNA group. However, since they were simulated with the PSMA crystal structure lacking the ABS (PDB: 8BOW), occupying it would require a loop rearrangement, possibly associated with a high energy barrier. In contrast, the co-crystallized ligand PSMA-1007 (PDB: 5O5T) binds the ABS with high stability throughout the simulation. Only the linker glutamate moieties exhibit moderate flexibility, which is in agreement with X-ray crystallographic results, where the second glutamate residue from the FNA moiety remain unresolved[189].

PSMA-BB was included in our set of modeled PSMA ligands due to its known poor affinity for PSMA (IC_{50}=108 nM). It shares many structural features with PSMA-1007 (IC_{50}=6.7 nM[208]), including the Glu-urea-Lys vector, the napthyl group and Glu residues within the linker, and the FNA moiety. Surprisingly, the determined ΔG_{bind} for PSMA-BB (-107.1 kcal/mol) is more favorable than that of the PSMA-1007 (-94.6 kcal/mol), despite the latters engagement with the defined ABS. Breakdown of the ΔG_{bind} into its individual energetic contributions reveals that the more favorable binding affinity of PSMA-BB is driven solely by electrostatic interactions. The most pronounced differences between PSMA-1007 and PSMA-BB are localized in the outer region of the binding site entrance. Here, PSMA-BB exhibits highly favorable interactions with Lys610, whereas PSMA-1007 engages in a strongly unfavorable interaction with Glu542.

The Entrance of the Binding Funnel. The simulation of PSMA in complex with rhPSMA-7 yields the most favorable total binding free energy (-145.3 kcal/mol) of all modeled ligands (Table 5.1). Conformational analysis demonstrates that the entire ligand adopts a highly stable pose (RMSF < 2 Å) throughout the simulation, with the predominant pose sampled 97 % of the simulation (Fig. 5.12).

MD simulations did not enable the SiFA moiety to occupy the ABS. However, instead it adopts a stable binding position within the inner entrance area. Here, Val208 and Ala701 create a hydrophobic environment around the tert-butyl groups, while Asn698 simultaneously interacts with both the fluorine atom and the carbonyl groups of the succinic acid spacer, upon which the SiFA moiety is staggered. The linker carboxy groups facilitate additional hydrogen bonding, further stabilizing the ligand, while the free carboxy group of the DOTAGA chelator forms extensive hydrogen bonds with Arg181. Notably, rhPSMA-7 carries the highest negative net charge (-6) among all ligands, which likely contributes considerably to the electrostatic interaction energy.

With the exception of DCFPyL, all modeled ligands had their chelator or prosthetic group positioned at the entrance of the binding funnel, exhibiting variable flexibility during MD

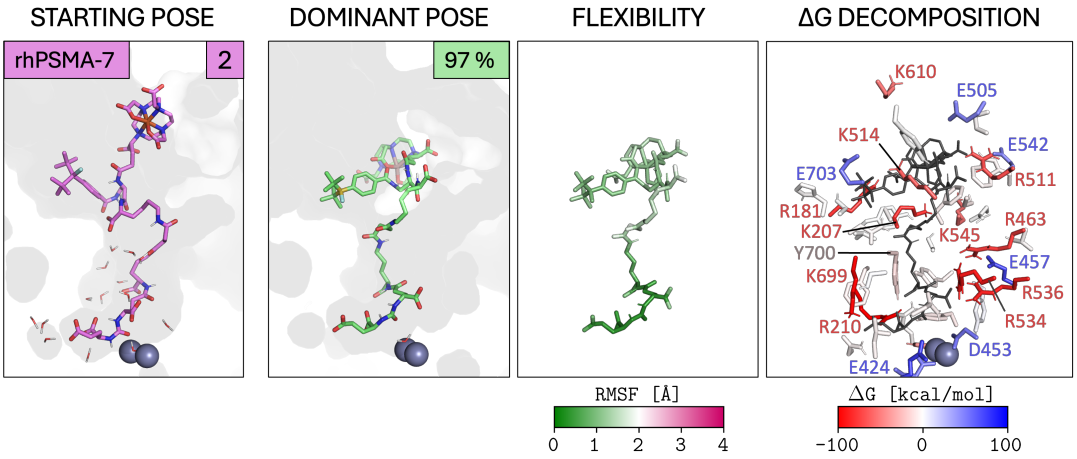


Figure 5.12.: Flexibility and interaction pattern of rhPSMA-7. The top-ranked docking pose of rhPSMA-7 from hydration level 2 served as starting structure for a 100 ns MD simulation in complex with PSMA (PDB: 8BOW). The predominant ligand conformation (97 % of the simulation) was identified through RMSD-based clustering. Ligand flexibility, as assessed by RMSF, is represented by a color shift ranging from rigid (green) to flexible (pink). The per-residue ΔG_{bind} decomposition is displayed for the PSMA binding-site residues defined by their presence of at least 80 % within 3.5 Å of all modeled PSMA ligands.

simulations. For example, the MD simulation of the co-crystallized PSMA-617 structure (PDB: 8BOW) reveals exceptionally high flexibility in the non-labeled DOTA chelator located at the binding funnel entrance, while the rest of the ligand within the binding site retains its starting conformation with high stability. This behavior correlates with the X-ray crystallographic data, in which the chelator is mainly unresolved[335].

The MD simulation of the Ga-labeled PSMA-617 reveals reduced flexibility in the chelator compared to its non-labeled analog. In the predominant pose (54 % sampled), the chelator folds into the internal binding funnel, enabling an interaction with Arg463 through its sole free carboxy group, which is not involved in Ga metal coordination. Since no experimental structure of a metal-labeled PSMA-617 is available, this behavior cannot be verified. However, a comparison of the total ΔG_{bind} values for the non-labeled (−101.7 kcal/mol) and Ga-labeled (−103.8 kcal/mol) PSMA-617 reveals only a minor difference, suggesting limited impact from metal labeling.

While certain ligands, such as PSMA-1007, PSMA-GCK01 (run A), PSMA-I&S (run B), and rhPSMA-7, exhibit greater rigidity in their radiolabel, others, including non-labeled PSMA-617, PSMA-BB, PSMA-11, and PSMA-I&T, display increased flexibility. This variability in flexibility correlates with their positioning: more flexible units tend to localize at the outer entrance, in solvent-exposed regions, whereas rigid groups preferentially interact with the inner entrance area and the ABS, where they are stabilized through hydrophobic interactions or hydrogen bonding.

The Hollow Cavity. A recurring pattern in linker conformation is observed for certain larger ligands containing multiple spacer units. In simulations of PSMA-11, PSMA-I&S (run A), and rhPSMA-7—each starting from random linker conformations—the linker adopts a zig-zag-like conformation. Here, the vectors' Lys or Glu side chain extends into the hollow cavity, while the following spacer unit bends in the opposite direction, stacking upon itself. This conformation is stabilized by hydrogen bonds involving the linkers amide bonds at the bending sites.
For rhPSMA-7, the vectors' Glu moiety projects the deepest into the cavity, allowing the carboxy terminus of the Lys spacer to engage in hydrogen bonding with Arg210, Gln254, and Tyr234. This interaction enhances the ΔG_{bind} contributions of Gln254 and Tyr234 compared to the other ligands.
PSMA-I&T deviates from the pattern, as its vectors' Lys sidechain does not extend into the hollow cavity but instead projects upwards. Nevertheless, its linker retains high rigidity. This stability may originate from the staggered linker formation, which effectively occupies the hollow cavity—a feature similar to the observed pattern. In contrast, PSMA-I&S (run B) does not adopt the observed linker conformation and fails to occupy the hollow cavity, which might be associated with reduced stability, as reflected by increased RMSF values.

To summarize, the MD simulations of PSMA–ligand complexes confirm that our docking procedure successfully generated stably bound ligand poses, with favorable binding interactions predominantly mediated by the vector. The results also underscore the critical importance of the starting structure. While the simulations effectively facilitated smaller conformational adjustments, such as reorienting the vector to align with the crystal structure, they were unable to induce larger structural rearrangements of the ligand or protein. This limitation is exemplified by DCFPyL, which failed to adopt a conformation capable of binding the R-patch pocket at the S1 subsite of PSMA.
An interesting observation is the shared staggered linker conformation among larger PSMA ligands, which effectively occupies the hollow cavity between the S1′ glutamate pocket and the S1 accessory lipophilic region.

5.3. Discussion

In this work, we developed a computational workflow to elucidate the binding modes of a set of clinically relevant PSMA-targeted radioligands. By investigating the relationships between molecular PSMA–ligand interactions, inhibitory activity, and pharmacokinetic properties, we aim to derive valuable insights for future CARD studies.
Our set of PSMA-targeted inhibitors included nine ligands, each exhibiting distinct inhibitory activity and clearance profiles. The computational workflow comprised the following steps: (i) docking the ligands at various hydration levels within the PSMA binding site; (ii) re-scoring the top-10 docking poses to ensure the vector was positioned within the catalytic chamber; (iii) selecting a starting structure for MD simulations based on binding affinity and a rational assessment of the ligand docking conformations; (iv) parameterizing the metal centers in the PSMA protein structure (Zn^{2+} active site) and the ligands chelators (Ga- or Tc-labeled); and (v) performing MD simulations of both co-crystallized and selected ligand docking poses in complex with PSMA.

Pose Generation. Docking at varying hydration levels within the binding site combined with the re-scoring procedure yielded at least one valid pose for each ligand with the vector positioned within the catalytic chamber. Notably, higher hydration levels enabled the naphthyl linker group—present in PSMA-617, PSMA-1007, PSMA-BB, and PSMA-GCK01—to adopt a conformation within the S1 accessory lipophilic region, mirroring its placement in the crystal structure. However, no clear correlation emerged between docking affinity and hydration levels, either generally or for specific ligands.
The docking solutions indicate that smaller ligands, or those sharing key structural features with the co-crystallized ligands, exhibited particularly high affinities. In contrast, larger ligands—specifically PSMA-I&T, PSMA-I&S, and rhPSMA-7—often failed to generate valid poses and generally displayed lower affinities. We assume that the binding funnel of the PSMA crystal structures used as docking receptors (PDB: 8BOW, 5O5T) is optimally adapted to the co-crystallized ligands PSMA-617 and PSMA-1007, both of which feature a similar naphthyl group-containing linker structure. Smaller or structurally analogous ligands may benefit from these preformed adjustments, whereas larger ligands with complex linker architectures might require additional or distinct conformational adaptations of the receptor for optimal accommodation that cannot be accounted for using rigid receptor docking.

Binding Mode Evaluation. To address the limitations encountered during docking, we subjected promising docking poses to 100 ns MD simulations. These simulations aimed to yield insights into protein–ligand stability and affinity, as assessed by RMSF and binding

free energy (ΔG_{bind}) calculations. The binding mode was further analyzed in detail for each ligand and PSMA binding site residue by per-residue ΔG_{bind} decomposition, hydrogen bonding and minimum distance analyses.

Simulations of co-crystallized ligands demonstrated the validity of our derived force-field parameters for both the ligands and the Zn^{2+} ions in the PSMA active site, which remained stable throughout. The simulations preserved overall ligand conformations, with flexibility confined to regions that were also unresolved in the experimental crystal structures. Overall, the majority of the ligands demonstrated stable binding to PSMA. Analysis of ligand interactions with individual binding-site residues indicates the importance of positively charged PSMA residues for ΔG_{bind}, as they contribute most favorably to electrostatic interactions with the ligands, which are all negatively charged. We further found that the vector served as the primary contributor to ΔG_{bind}, characterized by the highest rigidity and formation of the most consistent hydrogen bonds with catalytic chamber residues.

The critical role of the vector in stabilizing the PSMA–ligand complexes is demonstrated by the simulation of PSMA-11. Here, an altered conformation of the urea moiety introduced unfavorable interactions, especially with Gly518, likely leading to its detachment from the active site and subsequent movement toward the binding funnel entrance.

However, in several cases where docked ligands exhibited vector conformations deviating from the reference crystal structures, MD simulations successfully corrected them, for example, by restoring the conserved *trans,trans* conformation of the urea moiety.

While such smaller conformational changes were readily achieved, MD simulations failed to sample larger structural rearrangements in either the ligand or the protein. For example, DCFPyL is likely to bind the S1 R-patch due to its similarities to DCIBzL[185]. However, while docking could not generate such poses because this pocket is absent in both PSMA crystal structures (PDB: 8BOW and 5O5T), MD simulations may have been too short to capture the necessary structural adaptations.

Moreover, occupation of the ABS was proposed for several ligands, including PSMA-I&T and PSMA-I&S through their aromatic linker groups, and PSMA-11 and rhPSMA-7 via their HBED-CC chelator and SiFA moiety, respectively[336–338]. However, both docking and MD simulations suggest that the architecture of these ligands may not be optimal for binding this pocket. For instance, the PSMA-I&S structure, with its D-Tyr residue initially projecting into the ABS, did not maintain this position during MD simulations. Furthermore, it remains uncertain whether larger groups, like the HBED-CC chelator or SiFA moiety, can effectively bind to the ABS, given the limited depth of the pocket due to the β-sheet located at its rear.

The MD simulations revealed a shared staggered linker conformation among the ligands PSMA-11, PSMA-I&T, PSMA-I&S, and rhPSMA-7. Although the long alkyl chains in

the linker suggest intrinsic flexibility, they remain predominantly stable, while occupying the hollow cavity between the S1′ glutamate pocket and the S1 accessory lipophilic region. X-ray crystallography data from Zhang et al. (2010)[186] have shown that the linker of Glu-ureido-based antibody-recruiting molecules targeting prostate cancer engages this cavity. The corresponding structures (e.g., PDB entries 2XEI, 2XEG, and 2XEF) indicate that the linker pose is stabilized by interactions between its triazole ring and the Tyr700 side chain. In contrast, PSMA-11, PSMA-I&T, PSMA-I&S, and rhPSMA-7 feature fully aliphatic linker structures in this region. We assume that their staggered conformation is stabilized by two key factors: the lipophilic environment of the internal PSMA binding funnel, abundant in non-polar residues, as well as intermolecular hydrogen bonds between the linker amide groups at the bending sites and PSMA binding-site residues.

Workflow Improvements. To systematically evaluate the relationship between structural ligand properties and experimental affinity data, knowledge about the PSMA–ligand interactions at the molecular level is required. The computational workflow developed in this study offers a robust basis for analyzing such interactions. Nevertheless, further refinements to the current approach could improve the efficiency and accuracy of the modeling methods. For example, to prevent unfavorable vector interactions in PSMA–ligand complexes, the docking protocol could incorporate constraints to maintain the conserved *trans,trans* conformation of the urea moiety. An adapted re-scoring script could further ensure the correct orientation of the urea moiety relative to critical residues such as Gly518 and Tyr552. This combined approach would preserve a certain flexibility during docking—compared to a fully rigid vector—while ensuring favorable interactions with the urea moiety during MD simulations.

While using the PSMA structure co-crystallized with DCIBzL (PDB: 3D7H) as a receptor may represent the most straightforward approach to generate DCFPyL poses that occupy the R-patch pocket, flexible receptor docking offers an alternative by allowing flexibility of selected binding-site residues. For instance, a shift in Arg463 could facilitate access of the DCFPyL FNA group to the R-patch pocket. However, flexible receptor docking typically restricts flexibility to a limited number of side chains and does not alter the protein backbone[32,225,339,340]. Since rearrangement of the protein backbone may be necessary to accommodate specific PSMA ligands, as demonstrated in the formation of the ABS[186], ensemble docking provides an alternative approach. Here, MD simulations of the target protein are usually performed to sample a range of binding site conformations. However, ensemble docking success depends critically on selecting an appropriate ensemble of target structures, as poor choices can compromise docking accuracy[341]. MD simulations offer a powerful approach for sampling the dynamics of protein–ligand interactions, able to

perform any structural change consistent with the underlying force field. However, in the current conventional MD simulations structures sometimes got stuck in local minima, limiting the exploration of potential binding modes within feasible simulation timescales. Accelerated sampling methods, such as Hamiltonian replica exchange molecular dynamics (HREMD), enhance the crossing of energy barriers, thereby enabling a more thorough and efficient exploration of the conformational landscape[342]. This approach could be applied to the PSMA–ligand complexes modeled in this study, particularly to address challenges observed with ligands such as PSMA-I&S, which exhibited conformational instability throughout run B. By implementing accelerated sampling in both runs A and B, convergence toward a consistent binding mode would not only resolve these instabilities but also serve as a robust validation of the resulting ligand pose. Furthermore, this method could be employed to validate highly stable complexes, such as observed in the simulation of PSMA in complex with rhPSMA-7. If the complex remains stable or dominates the conformational ensemble during accelerated simulations, it would support the reliability of the predicted binding mode.
To elucidate the distinct binding modes of the ligands, we analyzed their interactions with PSMA from the perspective of specific binding-site residues. Understanding how individual ligand components contribute to binding affinity would provide valuable additional insights for future radioligand design. This could be achieved by dividing the ligands into subunits (e.g., vector, linker, and radiolabel) during parameterization, enabling per-residue ΔG_{bind} decomposition of the ligands using gmx_MMPBSA.

Experimental Validation. To bridge the gap between in silico predictions and practical applications, experimental validation is essential. In addition to the clinically relevant PSMA tracers, we included PSMA-BB, a ligand with poor affinity for PSMA (IC_{50} = 108 nM), which shares a similar structure with PSMA-1007 (IC_{50} = 6.7 nM)[208]. Surprisingly, our simulations reveal a more favorable ΔG_{bind} for PSMA-BB, driven by enhanced electrostatic interactions. This observed difference in ΔG_{bind} may diminish upon expanding the computational dataset, for instance, by performing triplicate simulations or applying HREMD, to obtain more robust and statistically reliable results.
Our current approach for binding free energy estimation calculates the effective free energy, which is an approximation of the binding free energy that neglects the entropic term. This approach is considered sufficient for comparative studies of related ligands, while offering computational efficiency[59]. However, with computational predictions deviating from the experimental data, it may be necessary to reconsider or adapt this methodology. For example, by explicitly including entropy estimations to the gmx_MMPBSA calculations, the accuracy of our computational predictions could be further improved. An alternative

approach to MM/PBSA is free energy perturbation (FEP), which explicitly samples the transformation between bound and unbound states, thereby capturing both enthalpic and entropic contributions with higher accuracy. However, both entropy estimations using gmx_MMPBSA and, particularly, FEP are still computationally expensive, demanding extensive sampling and significant computational resources[57,59,343].

While the literature provides experimental data on the clinically relevant PSMA ligands modeled in this study, these data originate from diverse laboratory conditions. To ensure consistency and comparability, affinity measurements for all ligands—using surface plasmon resonance (SPR) within a single laboratory setting—would be more appropriate. By experimentally determining the equilibrium dissociation constant (K_d) and comparing it with our computed binding free energies, we could validate our computational predictions. Such measurements would enable refinement of our models, and ultimately guide the development of more effective PSMA-targeted radiopharmaceuticals.

6. Conclusion

In this thesis, we developed and applied computational workflows for the modeling and design of peptide-based therapeutics. We focused our investigations on two viral proteases from severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the main protease (M^{pro}) and papain-like protease (PL^{pro}), which play essential roles in viral replication, and on the prostate specific membrane antigen (PSMA), a clinically validated target in prostate cancer imaging and therapy. By addressing the intrinsic flexibility of peptides, which poses a key challenge in the prediction of peptide–protein interactions, these workflows establish a robust foundation for structure-based design and the optimization of peptide-based therapeutic candidates. The key findings of this work are summarized below.

In Chapter 3, we modeled the interactions of M^{pro} with previously experimentally validated peptide inhibitors (3CVL-2, 3CVL-4, 3CVL-7) and their retro-inverso D-enantiomeric (Dri) counterparts. Using molecular dynamics (MD) simulations and docking, we identified diverse binding modes, including competitive inhibition at the S-subsite and non-competitive binding at the protomer interface. However, the lack of correlation between predicted binding affinities and experimental IC_{50} values underscored the need for workflow refinements, such as specialized peptide-docking tools like AutoDock CrankPep (ADCP).

In Chapter 4, we explored three strategies of computer-aided peptide inhibitor design (CAPID) against M^{pro} and PL^{pro}, based on modeling both proteases with their natural substrates extracted from the SARS-CoV-2 polypeptide cleavage site sequences. In Strategy I, we tested the retro-inverso method on the substrate peptides, revealing its limitations when peptide binding relies heavily on backbone-mediated hydrogen bonds. In Strategy II, a library of M^{pro}–peptide MD simulations, including validated M^{pro} inhibitors, was created and used for design inspirations. The resulting peptide inhibitor a10 achieved a half-maximal inhibitory concentration (IC_{50}) of $22.3 \pm 4.7\,\mu M$.

Strategy III introduced the computational dynamic saturation mutagenesis (CDSM) protocol, which successfully predicted beneficial mutations, such as Tyr at P1 for M^{pro} inhibitors, leading to the design of peptide c04 with an IC_{50} of $4.8 \pm 0.5\,\mu M$. However, the CDSM protocol failed to predict the beneficial Arg substitution at P3 for M^{pro} peptide inhibitors, suggesting that longer post-mutation MD simulations might be required for substitutions

involving highly flexible side chains. Furthermore, CDSM identified designs z07 and z08, which showed a 20 kcal/mol improvement in binding free energy compared to the substrate peptides. However, *in vitro* inhibition assays revealed negligible activity against PL^{pro}. While our CAPID strategies successfully predicted potent M^{pro} inhibitor designs, further refinements, including ADMET property considerations, are required for accurate experimental validation against PL^{pro}.

In Chapter 5, we developed a workflow combining hydrated docking, re-scoring, metal-site parameterization and MD simulations to investigate the binding of clinically relevant PSMA-targeted radioligands. The workflow generated stable ligand poses and elucidated critical interactions, including the important role of the urea-based vector in stabilizing PSMA–ligand complexes. MD simulations further revealed a predominant zig-zag-like linker conformation in ligands with larger spacers (PSMA-11, PSMA-I&S, and rhPSMA-7), which has not been previously reported. However, limitations in sampling larger conformational rearrangements suggested the need for accelerated sampling methods like Hamiltonian replica exchange molecular dynamics (HREMD).

In conclusion, this thesis established computational workflows that enable accurate analysis of peptide–protein interactions, thereby laying a robust foundation for the design of peptide-based therapeutics. The CDSM protocol, in particular, demonstrates great potential due to its seamless integration into simulation studies using the widely adopted GROMACS platform and its broad applicability across diverse peptide–protein systems. Future work will focus on refining the CDSM protocol to enhance its accuracy through validation with experimental data. To address current limitations in the modeling workflow for PSMA radioligands, we will explore accelerated sampling techniques. The goal is to develop a robust and universally applicable workflow that enables detailed elucidation of PSMA–radioligand interactions. Combined with reliable experimental data, this will allow the interpretation of structure–activity relationships and ultimately inform future computer-aided radiopharmaceutical design (CARD) studies.

Bibliography

[1] Posner, B. I. and Laporte, S. A. (2010) Cellular Signalling: Peptide Hormones and Growth Factors, pages 1–16. Elsevier. ISBN 9780444536174.

[2] Hökfelt, T., Bartfai, T., and Bloom, F. (2003) Neuropeptides: opportunities for drug discovery. *The Lancet Neurology*, 2(8):463–472. ISSN 1474-4422.

[3] Lai, Y. and Gallo, R. L. (2009) AMPed up immunity: how antimicrobial peptides have multiple roles in immune defense. *Trends in Immunology*, 30(3):131–141. ISSN 1471-4906.

[4] Petsalaki, E. and Russell, R. B. (2008) Peptide-mediated interactions in biological systems: new discoveries and applications. *Current Opinion in Biotechnology*, 19 (4):344–350. ISSN 0958-1669.

[5] Wang, L., Wang, N., Zhang, W., Cheng, X., Yan, Z., Shao, G., et al. (2022) Therapeutic peptides: current applications and future directions. *Signal Transduction and Targeted Therapy*, 7(1). ISSN 2059-3635.

[6] Henninot, A., Collins, J. C., and Nuss, J. M. (2018) The Current State of Peptide Drug Discovery: Back to the Future? *Journal of Medicinal Chemistry*, 61(4): 1382–1414.

[7] Global Peptide Therapeutics Market Size (2025) `https://www.grandviewresearch.com/industry-analysis/peptide-therapeutics-market`. Last checked: Thursday, 03. July 2025.

[8] Global Pharmaceutical Market Size (2025) `https://www.grandviewresearch.com/industry-analysis/pharmaceutical-market-report`. Last checked: Friday, 04. July 2025.

[9] Global Conventional Drugs (Small Molecules) Market Size (2025) `https://www.grandviewresearch.com/horizon/statistics/pharmaceutical-market/molecule/conventional-drugs-small-molecules/global`. Last checked: Tuesday, 19. August 2025.

[10] Global Biologics Market Size (2025) https://www.grandviewresearch.com/horizon/outlook/biologics-market-size/global. Last checked: Tuesday, 19. August 2025.

[11] London, N., Raveh, B., and Schueler-Furman, O. (2013) Druggable protein–protein interactions – from hot spots to hot segments. *Current Opinion in Chemical Biology*, 17(6):952–959. ISSN 1367-5931. Synthetic biology * Synthetic biomolecules.

[12] Global Insulin Market Size (2025) https://www.grandviewresearch.com/industry-analysis/insulin-market-report. Last checked: Friday, 04. July 2025.

[13] Muttenthaler, M., King, G. F., Adams, D. J., and Alewood, P. F. (2021) Trends in peptide drug discovery. *Nature Reviews Drug Discovery*, 20(4):309–325. ISSN 1474-1784.

[14] Vecchio, I., Tornali, C., Bragazzi, N. L., and Martini, M. (2018) The Discovery of Insulin: An Important Milestone in the History of Medicine. *Frontiers in Endocrinology*, 9. ISSN 1664-2392.

[15] Smith, G. P. and Petrenko, V. A. (1997) Phage Display. *Chemical Reviews*, 97(2): 391–410.

[16] Akbarian, M. and Chen, S.-H. (2022) Instability Challenges and Stabilization Strategies of Pharmaceutical Proteins. *Pharmaceutics*, 14(11):2533. ISSN 1999-4923.

[17] Pelay-Gimeno, M., Glas, A., Koch, O., and Grossmann, T. N. (2015) Structure-Based Design of Inhibitors of Protein–Protein Interactions: Mimicking Peptide Binding Epitopes. *Angewandte Chemie International Edition*, 54(31):8896–8927. ISSN 1521-3773.

[18] Scharbert, L. and Strodel, B. (2025) Innovative strategies for modeling peptide–protein interactions and rational peptide drug design. *Current Opinion in Structural Biology*, 93:103083.

[19] Sabe, V. T., Ntombela, T., Jhamba, L. A., Maguire, G. E., Govender, T., Naicker, T., et al. (2021) Current trends in computer aided drug design and a highlight of drugs discovered via computational techniques: A review. *European Journal of Medicinal Chemistry*, 224:113705. ISSN 0223-5234.

[20] Bertoline, L. M. F., Lima, A. N., Krieger, J. E., and Teixeira, S. K. (2023) Before and after AlphaFold2: An overview of protein structure prediction. *Frontiers in Bioinformatics*, 3. ISSN 2673-7647.

[21] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, 596 (7873):583–589. ISSN 1476-4687.

[22] Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., et al. (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016):493–500. ISSN 1476-4687.

[23] Mondal, A., Chang, L., and Perez, A. (2022) Modelling peptide–protein complexes: docking, simulations and machine learning. *QRB Discovery*, 3. ISSN 2633-2892.

[24] Macalino, S. J. Y., Gosu, V., Hong, S., and Choi, S. (2015) Role of computer-aided drug design in modern drug discovery. *Archives of Pharmacal Research*, 38(9): 1686–1701. ISSN 1976-3786.

[25] Cherkasov, A., Muratov, E. N., Fourches, D., Varnek, A., Baskin, I. I., Cronin, M., et al. (2014) QSAR Modeling: Where Have You Been? Where Are You Going To? *Journal of Medicinal Chemistry*, 57(12):4977–5010.

[26] Vuorinen, A. and Schuster, D. (2015) Methods for generating and applying pharmacophore models as virtual screening filters and for bioactivity profiling. *Methods*, 71:113–134. ISSN 1046-2023.

[27] Pierce, B. G., Wiehe, K., Hwang, H., Kim, B.-H., Vreven, T., and Weng, Z. (2014) ZDOCK server: interactive docking prediction of protein–protein complexes and symmetric multimers. *Bioinformatics*, 30(12):1771–1773.

[28] Ramírez-Aportela, E., López-Blanco, J. R., and Chacón, P. (2016) FRODOCK 2.0: fast protein–protein docking server. *Bioinformatics*, 32(15):2386–2388. ISSN 1367-4803.

[29] Weng, G., Wang, E., Wang, Z., Liu, H., Zhu, F., Li, D., et al. (2019) HawkDock: a web server to predict and analyze the protein–protein complex based on computational docking and MM/GBSA. *Nucleic acids research*, 47(W1):W322–W330.

[30] Jones, G., Willett, P., Glen, R. C., Leach, A. R., and Taylor, R. (1997) Development and validation of a genetic algorithm for flexible docking. *Journal of molecular biology*, 267(3):727–748.

[31] Jain, A. N. (2007) Surflex-Dock 2.1: robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. *Journal of computer-aided molecular design*, 21:281–306.

[32] Trott, O. and Olson, A. J. (2009) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2):455–461. ISSN 1096-987X.

[33] Lee, H., Heo, L., Lee, M. S., and Seok, C. (2015) GalaxyPepDock: a protein–peptide docking tool based on interaction similarity and energy optimization. *Nucleic acids research*, 43(W1):W431–W435.

[34] Xu, X., Yan, C., and Zou, X. (2018) MDockPeP: An ab-initio protein–peptide docking server. *Journal of computational chemistry*, 39(28):2409–2413.

[35] Zhou, P., Jin, B., Li, H., and Huang, S.-Y. (2018) HPEPDOCK: a web server for blind peptide–protein docking based on a hierarchical algorithm. *Nucleic acids research*, 46(W1):W443–W450.

[36] Kurcinski, M., Jamroz, M., Blaszczyk, M., Kolinski, A., and Kmiecik, S. (2015) CABS-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site. *Nucleic acids research*, 43(W1):W419–W424.

[37] de Vries, S. J., Rey, J., Schindler, C. E., Zacharias, M., and Tuffery, P. (2017) The pepATTRACT web server for blind, large-scale peptide–protein docking. *Nucleic Acids Research*, 45(W1):W361–W364.

[38] Antunes, D. A., Moll, M., Devaurs, D., Jackson, K. R., Lizée, G., and Kavraki, L. E. (2017) DINC 2.0: a new protein–peptide docking webserver using an incremental approach. *Cancer research*, 77(21):e55–e57.

[39] Zhang, Y. and Sanner, M. F. (2019) AutoDock CrankPep: combining folding and docking to predict protein–peptide complexes. *Bioinformatics*, 35(24):5121–5127. ISSN 1367-4811.

[40] Trellet, M., Melquiond, A. S., and Bonvin, A. M. (2013) A unified conformational selection and induced fit approach to protein-peptide docking. *PloS one*, 8(3): e58769.

[41] Weng, G., Gao, J., Wang, Z., Wang, E., Hu, X., Yao, X., et al. (2020) Comprehensive Evaluation of Fourteen Docking Programs on Protein–Peptide Complexes. *Journal of Chemical Theory and Computation*, 16(6):3959–3969. ISSN 1549-9626.

[42] Podtelezhnikov, A. A. and Wild, D. L. (2005) Exhaustive Metropolis Monte Carlo sampling and analysis of polyalanine conformations adopted under the influence of hydrogen bonds. *Proteins: Structure, Function, and Bioinformatics*, 61(1):94–104. ISSN 1097-0134.

[43] Podtelezhnikov, A. A. and Wild, D. L. (2008) CRANKITE: A fast polypeptide backbone conformation sampler. *Source Code for Biology and Medicine*, 3(1). ISSN 1751-0473.

[44] Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., et al. (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry*, 30(16):2785–2791.

[45] Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., et al. (2021) Protein complex prediction with AlphaFold-Multimer.

[46] Johansson-Åkhe, I. and Wallner, B. (2022) Improving peptide-protein docking with AlphaFold-Multimer using forced sampling. *Frontiers in Bioinformatics*, 2. ISSN 2673-7647.

[47] Shanker, S. and Sanner, M. F. (2023) Predicting Protein–Peptide Interactions: Benchmarking Deep Learning Techniques and a Comparison with Focused Docking. *Journal of Chemical Information and Modeling*, 63(10):3158–3170.

[48] Manshour, N., Ren, J. Z., Esmaili, F., Bergstrom, E., and Xu, D. (2024) Comprehensive Evaluation of AlphaFold-Multimer, AlphaFold3 and ColabFold, and Scoring Functions in Predicting Protein-Peptide Complex Structures.

[49] Sinha, S., Tam, B., and Wang, S. M. (2022) Applications of Molecular Dynamics Simulation in Protein Study. *Membranes*, 12(9):844. ISSN 2077-0375.

[50] Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., et al. (2005) Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26(16):1781–1802. ISSN 1096-987X.

[51] Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., et al. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19 – 25. ISSN 2352-7110.

[52] Weiner, P. K. and Kollman, P. A. (1981) AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions. *Journal of Computational Chemistry*, 2(3):287–303.

[53] Rico, M. I., Lebedenko, C. G., Mitchell, S. M., and Banerjee, I. A. (2022) Molecular dynamics simulations, docking and MMGBSA studies of newly designed peptide-conjugated glucosyloxy stilbene derivatives with tumor cell receptors. *Molecular Diversity*, 26(5):2717–2743. ISSN 1573-501X.

[54] Jin, J., Wang, N., Wang, S., Alsaikhan, F., and Ahmad, I. (2024) Targeting the main SARS-Cov-2 pathways with peptide inhibitors by molecular docking and molecular simulation approaches. *Journal of Molecular Liquids*, 398:124193. ISSN 0167-7322.

[55] Miao, Q., Zhang, W., Zhang, K., Li, H., Zhu, J., and Jiang, S. (2021) Rational design of a potent macrocyclic peptide inhibitor targeting the PD-1/PD-L1 protein–protein interaction. *RSC Advances*, 11(38):23270–23279. ISSN 2046-2069.

[56] Malmstrom, R. D., Lee, C. T., Van Wart, A. T., and Amaro, R. E. (2014) Application of Molecular-Dynamics Based Markov State Models to Functional Proteins. *Journal of Chemical Theory and Computation*, 10(7):2648–2657.

[57] Wang, S., Sun, X., Cui, W., and Yuan, S. (2022) MM/PB(GB)SA benchmarks on soluble proteins and membrane proteins. *Frontiers in Pharmacology*, 13. ISSN 1663-9812.

[58] Miller, B. R., McGee, T. D., Swails, J. M., Homeyer, N., Gohlke, H., and Roitberg, A. E. (2012) MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *Journal of Chemical Theory and Computation*, 8(9):3314–3321. ISSN 1549-9626.

[59] Valdés-Tresanco, M. S., Valdés-Tresanco, M. E., Valiente, P. A., and Moreno, E. (2021) gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS. *Journal of Chemical Theory and Computation*, 17(10):6281–6291. ISSN 1549-9626.

[60] Kumari, R., Kumar, R., and Lynn, A. (2014) g_mmpbsa—A GROMACS Tool for High-Throughput MM-PBSA Calculations. *Journal of Chemical Information and Modeling*, 54(7):1951–1962.

[61] Paissoni, C., Spiliotopoulos, D., Musco, G., and Spitaleri, A. (2015) GMXPBSA 2.1: A GROMACS tool to perform MM/PBSA and computational alanine scanning. *Comput. Phys. Commun.*, 186(Supplement C):105–107.

[62] Chen, S., Lin, T., Basu, R., Ritchey, J., Wang, S., Luo, Y., et al. (2024) Design of target specific peptide inhibitors using generative deep learning and molecular dynamics simulations. *Nature Communications*, 15(1):1611.

[63] Wu, Y., Deng, S., Wang, X., Thunders, M., Qiu, J., and Li, Y. (2023) Discovery and Mechanism of Action of a Novel Antimicrobial Peptide from an Earthworm. *Microbiology Spectrum*, 11(1):e03206–22.

[64] Niu, P., Tao, Y., Meng, Q., Huang, Y., Li, S., Ding, K., et al. (2024) Discovery of novel macrocyclic derivatives as potent and selective cyclin-dependent kinase 2 inhibitors. *Bioorganic & Medicinal Chemistry*, 104:117711. ISSN 0968-0896.

[65] Eberle, R. J., Sevenich, M., Gering, I., Scharbert, L., Strodel, B., Lakomek, N. A., et al. (2023) Discovery of all-d-peptide inhibitors of SARS-CoV-2 3C-like protease. *ACS chemical biology*, 18(2):315–330.

[66] Valiente, P. A., Wen, H., Nim, S., Lee, J., Kim, H. J., Kim, J., et al. (2021) Computational Design of Potent D-Peptide Inhibitors of SARS-CoV-2. *Journal of Medicinal Chemistry*, 64(20):14955–14967.

[67] Amarasinghe, K. N., De Maria, L., Tyrchan, C., Eriksson, L. A., Sadowski, J., and Petrović, D. (2022) Virtual Screening Expands the Non-Natural Amino Acid Palette for Peptide Optimization. *Journal of Chemical Information and Modeling*, 62(12): 2999–3007.

[68] Shahab, M., Khan, S. S., Zulfat, M., Bin Jardan, Y. A., Mekonnen, A. B., Bourhia, M., et al. (2023) In silico mutagenesis-based designing of oncogenic SHP2 peptide to inhibit cancer progression. *Scientific Reports*, 13(1). ISSN 2045-2322.

[69] Khan, A., Umbreen, S., Hameed, A., Fatima, R., Zahoor, U., Babar, Z., et al. (2021) In Silico Mutagenesis-Based Remodelling of SARS-CoV-1 Peptide (ATLQAIAS) to Inhibit SARS-CoV-2: Structural-Dynamics and Free Energy Calculations. *Interdisciplinary Sciences: Computational Life Sciences*, 13(3):521–534. ISSN 1867-1462.

[70] Ghobadi, Z., Mahnam, K., and Shakhsi-Niaei, M. (2022) In-silico design of peptides for inhibition of HLA-A* 03-KLIETYFSK complex as a new drug design for treatment of multiples sclerosis disease. *Journal of Molecular Graphics and Modelling*, 111:108079.

[71] Wood, C. W., Ibarra, A. A., Bartlett, G. J., Wilson, A. J., Woolfson, D. N., and Sessions, R. B. (2020) BAlaS: fast, interactive and accessible computational alanine-scanning using BudeAlaScan. *Bioinformatics*, 36(9):2917–2919. ISSN 1367-4803.

[72] Ribeiro, J. V., Cerqueira, N. M., Moreira, I. S., Fernandes, P. A., and Ramos, M. J. (2014) CompASM: an Amber-VMD alanine scanning mutagenesis plug-in. *Marco Antonio Chaer Nascimento: A Festschrift from Theoretical Chemistry Accounts*, pages 81–87.

[73] Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. *Journal of molecular graphics*, 14(1):33–38.

[74] Chan, H. T. H., Moesser, M. A., Walters, R. K., Malla, T. R., Twidale, R. M., John, T., et al. (2021) Discovery of SARS-CoV-2 Mpro peptide inhibitors from modelling substrate and ligand binding. *Chemical Science*, 12(41):13686–13703. ISSN 2041-6539.

[75] Hetherington, K., Dutt, S., Ibarra, A. A., Cawood, E. E., Hobor, F., Woolfson, D. N., et al. (2021) Towards optimizing peptide-based inhibitors of protein–protein interactions: predictive saturation variation scanning (PreSaVS). *RSC Chemical Biology*, 2(5):1474–1478. ISSN 2633-0679.

[76] McIntosh-Smith, S., Price, J., Sessions, R. B., and Ibarra, A. A. (2014) High performance in silico virtual drug screening on many-core processors. *The International Journal of High Performance Computing Applications*, 29(2):119–134. ISSN 1741-2846.

[77] Vu, Q. N., Young, R., Sudhakar, H. K., Gao, T., Huang, T., Tan, Y. S., et al. (2021) Cyclisation strategies for stabilising peptides with irregular conformations. *RSC Medicinal Chemistry*, 12(6):887–901.

[78] Marsault, E. and Peterson, M. L. (2011) Macrocycles Are Great Cycles: Applications, Opportunities, and Challenges of Synthetic Macrocycles in Drug Discovery. *Journal of Medicinal Chemistry*, 54(7):1961–2004.

[79] Bluntzer, M. T. J., O'Connell, J., Baker, T. S., Michel, J., and Hulme, A. N. (2021) Designing stapled peptides to inhibit protein-protein interactions: An analysis of successes in a rapidly changing field. *Peptide Science*, 113(1):e24191.

[80] Li, Y., Wu, M., Fu, Y., Xue, J., Yuan, F., Qu, T., et al. (2024) Therapeutic stapled peptides: Efficacy and molecular targets. *Pharmacological Research*, 203:107137. ISSN 1043-6618.

[81] Barman, P., Joshi, S., Sharma, S., Preet, S., Sharma, S., and Saini, A. (2023) Strategic Approaches to Improvise Peptide Drugs as Next Generation Therapeutics.

International Journal of Peptide Research and Therapeutics, 29(4). ISSN 1573-3904.

[82] Doti, N., Mardirossian, M., Sandomenico, A., Ruvo, M., and Caporale, A. (2021) Recent applications of retro-inverso peptides. *International journal of molecular sciences*, 22(16):8677.

[83] Lombardi, A., Concepcion, E., Hou, H., Arib, H., Mezei, M., Osman, R., et al. (2020) Retro-inverso D-peptides as a novel targeted immunotherapy for Type 1 diabetes. *Journal of autoimmunity*, 115:102543.

[84] Kumar Sharma, A., Sharma, R., Dev Sarma, H., Mukherjee, A., Das, T., and Satpati, D. (2023) Design, synthesis and evaluation of 177Lu-labeled inverso and retro-inverso A9 peptide variants targeting HER2-overexpression. *Bioorganic Chemistry*, 140:106761. ISSN 0045-2068.

[85] Maack, T., Johnson, V., Kau, S. T., Figueiredo, J., and Sigulem, D. (1979) Renal filtration, transport, and metabolism of low-molecular-weight proteins: a review. *Kidney international*, 16(3):251–270.

[86] Knudsen, L. B. and Lau, J. (2019) The Discovery and Development of Liraglutide and Semaglutide. *Frontiers in Endocrinology*, 10. ISSN 1664-2392.

[87] Ferreira, L. L. and Andricopulo, A. D. (2019) ADMET modeling approaches in drug discovery. *Drug Discovery Today*, 24(5):1157–1165. ISSN 1359-6446.

[88] Fu, L., Shi, S., Yi, J., Wang, N., He, Y., Wu, Z., et al. (2024) ADMETlab 3.0: an updated comprehensive online ADMET prediction platform enhanced with broader coverage, improved performance, API functionality and decision support. *Nucleic Acids Research*, 52(W1):W422–W431. ISSN 1362-4962.

[89] Swanson, K., Walther, P., Leitz, J., Mukherjee, S., Wu, J. C., Shivnaraine, R. V., et al. (2024) ADMET-AI: a machine learning ADMET platform for evaluation of large-scale chemical libraries. *Bioinformatics*, 40(7). ISSN 1367-4811.

[90] Bento, A. P., Gaulton, A., Hersey, A., Bellis, L. J., Chambers, J., Davies, M., et al. (2013) The ChEMBL bioactivity database: an update. *Nucleic Acids Research*, 42 (D1):D1083–D1090. ISSN 1362-4962.

[91] Gashaw, I., Ellinghaus, P., Sommer, A., and Asadullah, K. (2011) What makes a good drug target? *Drug Discovery Today*, 16(23–24):1037–1043. ISSN 1359-6446.

[92] Payne, S. (2017) Family Coronaviridae, pages 149–158. Elsevier. ISBN 9780128031094.

[93] Liu, D. X., Liang, J. Q., and Fung, T. S. (2021) Human Coronavirus-229E, -OC43, -NL63, and -HKU1 (Coronaviridae), pages 428–440. Elsevier. ISBN 9780128145166.

[94] V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., and Thiel, V. (2020) Coronavirus biology and replication: implications for SARS-CoV-2. *Nature Reviews Microbiology*, 19(3):155–170. ISSN 1740-1534.

[95] Statement on the fifteenth meeting of the IHR (2005) Emergency Committee on the COVID-19 pandemic (2025) `https://www.who.int/news`. Last checked: Sunday, 06. July 2025.

[96] Weekly epidemiological update on COVID-19 - 4 May 2023 (2023) `https://www.who.int/publications`. Last checked: Sunday, 06. July 2025.

[97] Pak, A., Adegboye, O. A., Adekunle, A. I., Rahman, K. M., McBryde, E. S., and Eisen, D. P. (2020) Economic Consequences of the COVID-19 Outbreak: the Need for Epidemic Preparedness. *Frontiers in Public Health*, 8. ISSN 2296-2565.

[98] Harper, L., Kalfa, N., Beckers, G., Kaefer, M., Nieuwhof-Leppink, A., Fossum, M., et al. (2020) The impact of COVID-19 on research. *Journal of Pediatric Urology*, 16(5):715–716. ISSN 1477-5131.

[99] Zhang, J., Xiao, T., Cai, Y., and Chen, B. (2021) Structure of SARS-CoV-2 spike protein. *Current Opinion in Virology*, 50:173–182. ISSN 1879-6257.

[100] Casadevall, A. and Focosi, D. (2025) Lessons from the Use of Monoclonal Antibodies to SARS-CoV-2 Spike Protein During the COVID-19 Pandemic. *Annual Review of Medicine*, 76(1):1–12. ISSN 1545-326X.

[101] Echaide, M., Chocarro de Erauso, L., Bocanegra, A., Blanco, E., Kochan, G., and Escors, D. (2023) mRNA Vaccines against SARS-CoV-2: Advantages and Caveats. *International Journal of Molecular Sciences*, 24(6):5944. ISSN 1422-0067.

[102] Mascellino, M. T., Di Timoteo, F., De Angelis, M., and Oliva, A. (2021) Overview of the Main Anti-SARS-CoV-2 Vaccines: Mechanism of Action, Efficacy and Safety. *Infection and Drug Resistance*, Volume 14:3459–3476. ISSN 1178-6973.

[103] Hattab, D., Amer, M. F. A., Al-Alami, Z. M., and Bakhtiar, A. (2024) SARS-CoV-2 journey: from alpha variant to omicron and its sub-variants. *Infection*, 52(3): 767–786. ISSN 1439-0973.

[104] Robinson, B. W. S., Tai, A., and Springer, K. (2021) Why we still need drugs for COVID-19 and can't just rely on vaccines. *Respirology*, 27(2):109–111. ISSN 1440-1843.

[105] Li, G., Hilgenfeld, R., Whitley, R., and De Clercq, E. (2023) Therapeutic strategies for COVID-19: progress and lessons learned. *Nature Reviews Drug Discovery*, 22 (6):449–475. ISSN 1474-1784.

[106] Ullrich, S. and Nitsche, C. (2020) The SARS-CoV-2 main protease as drug target. *Bioorganic & Medicinal Chemistry Letters*, 30(17):127377. ISSN 0960-894X.

[107] Ullrich, S. and Nitsche, C. (2022) SARS-CoV-2 Papain-Like Protease: Structure, Function and Inhibition. *ChemBioChem*, 23(19). ISSN 1439-7633.

[108] Steiner, S., Kratzel, A., Barut, G. T., Lang, R. M., Aguiar Moreira, E., Thomann, L., et al. (2024) SARS-CoV-2 biology and host interactions. *Nature Reviews Microbiology*, 22(4):206–225. ISSN 1740-1534.

[109] Crooks, G. E., Hon, G., Chandonia, J.-M., and Brenner, S. E. (2004) WebLogo: A Sequence Logo Generator: Figure 1. *Genome Research*, 14(6):1188–1190. ISSN 1088-9051.

[110] Gordon, C. J., Tchesnokov, E. P., Woolner, E., Perry, J. K., Feng, J. Y., Porter, D. P., et al. (2020) Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from severe acute respiratory syndrome coronavirus 2 with high potency. *Journal of Biological Chemistry*, 295(20):6785–6797. ISSN 0021-9258.

[111] Owen, D. R., Allerton, C. M. N., Anderson, A. S., Aschenbrenner, L., Avery, M., Berritt, S., et al. (2021) An oral SARS-CoV-2 M pro inhibitor clinical candidate for the treatment of COVID-19. *Science*, 374(6575):1586–1593. ISSN 1095-9203.

[112] Marzi, M., Vakil, M. K., Bahmanyar, M., and Zarenezhad, E. (2022) Paxlovid: Mechanism of Action, Synthesis, and In Silico Study. *BioMed Research International*, 2022(1). ISSN 2314-6141.

[113] Jayk Bernal, A., Gomes da Silva, M. M., Musungaie, D. B., Kovalchuk, E., Gonzalez, A., Delos Reyes, V., et al. (2022) Molnupiravir for Oral Treatment of Covid-19 in Nonhospitalized Patients. *New England Journal of Medicine*, 386(6):509–520. ISSN 1533-4406.

[114] Singh, A. K., Singh, A., Singh, R., and Misra, A. (2021) Molnupiravir in COVID-19: A systematic review of literature. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 15(6):102329. ISSN 1871-4021.

[115] Wen, W., Chen, C., Tang, J., Wang, C., Zhou, M., Cheng, Y., et al. (2022) Efficacy and safety of three new oral antiviral treatment (molnupiravir, fluvoxamine and Paxlovid) for COVID-19: a meta-analysis. *Annals of Medicine*, 54(1):516–523. ISSN 1365-2060.

[116] Zur, M., Peselev, T., Yanko, S., Rotshild, V., and Matok, I. (2024) Efficacy and safety of antiviral treatments for symptomatic COVID-19 outpatients: Systematic review and network meta-analysis. *Antiviral Research*, 221:105768. ISSN 0166-3542.

[117] Schechter, I. and Berger, A. (1967) On the size of the active site in proteases. I. Papain. *Biochemical and Biophysical Research Communications*, 27(2):157–162. ISSN 0006-291X.

[118] Prescott, L. (2022) SARS-CoV-2 3CLpro whole human proteome cleavage prediction and enrichment/depletion analysis. *Computational Biology and Chemistry*, 98: 107671. ISSN 1476-9271.

[119] Roe, M. K., Junod, N. A., Young, A. R., Beachboard, D. C., and Stobart, C. C. (2021) Targeting novel structural and functional features of coronavirus protease nsp5 (3CLpro, Mpro) in the age of COVID-19. *Journal of General Virology*, 102(3). ISSN 1465-2099.

[120] Al Adem, K., Ferreira, J. C., Villanueva, A. J., Fadl, S., El-Sadaany, F., Masmoudi, I., et al. (2024) 3-chymotrypsin-like protease in SARS-CoV-2. *Bioscience Reports*, 44(8). ISSN 1573-4935.

[121] Lee, J. T., Yang, Q., Gribenko, A., Perrin, B. S., Zhu, Y., Cardin, R., et al. (2022) Genetic Surveillance of SARS-CoV-2 M pro Reveals High Sequence and Structural Conservation Prior to the Introduction of Protease Inhibitor Paxlovid. *mBio*, 13(4). ISSN 2150-7511.

[122] Ferreira, J. C., Fadl, S., and Rabeh, W. M. (2022) Key dimer interface residues impact the catalytic activity of 3CLpro, the main protease of SARS-CoV-2. *Journal of Biological Chemistry*, 298(6):102023. ISSN 0021-9258.

[123] Günther, S., Reinke, P. Y. A., Fernández-García, Y., Lieske, J., Lane, T. J., Ginn, H. M., et al. (2021) X-ray screening identifies active site and allosteric inhibitors of SARS-CoV-2 main protease. *Science*, 372(6542):642–646. ISSN 1095-9203.

[124] Zhao, Y., Zhu, Y., Liu, X., Jin, Z., Duan, Y., Zhang, Q., et al. (2022) Structural basis for replicase polyprotein cleavage and substrate specificity of main protease from SARS-CoV-2. *Proceedings of the National Academy of Sciences*, 119(16). ISSN 1091-6490.

[125] Hu, Q., Xiong, Y., Zhu, G., Zhang, Y., Zhang, Y., Huang, P., et al. (2022) The SARS-CoV-2 main protease (Mpro): Structure, function, and emerging therapies for COVID-19. *MedComm*, 3(3). ISSN 2688-2663.

[126] Olubiyi, O. O., Olagunju, M., Keutmann, M., Loschwitz, J., and Strodel, B. (2020) High Throughput Virtual Screening to Discover Inhibitors of the Main Protease of the Coronavirus SARS-CoV-2. *Molecules*, 25(14):3193. ISSN 1420-3049.

[127] Luttens, A., Gullberg, H., Abdurakhmanov, E., Vo, D. D., Akaberi, D., Talibov, V. O., et al. (2022) Ultralarge Virtual Screening Identifies SARS-CoV-2 Main Protease Inhibitors with Broad-Spectrum Activity against Coronaviruses. *Journal of the American Chemical Society*, 144(7):2905–2920.

[128] Clyde, A., Galanie, S., Kneller, D. W., Ma, H., Babuji, Y., Blaiszik, B., et al. (2022) High-Throughput Virtual Screening and Validation of a SARS-CoV-2 Main Protease Noncovalent Inhibitor. *Journal of Chemical Information and Modeling*, 62 (1):116–128.

[129] Ton, A., Gentile, F., Hsing, M., Ban, F., and Cherkasov, A. (2020) Rapid Identification of Potential Inhibitors of SARS-CoV-2 Main Protease by Deep Docking of 1.3 Billion Compounds. *Molecular Informatics*, 39(8). ISSN 1868-1751.

[130] Fischer, A., Sellner, M., Neranjan, S., Smieško, M., and Lill, M. A. (2020) Potential Inhibitors for Novel Coronavirus Protease Identified by Virtual Screening of 606 Million Compounds. *International Journal of Molecular Sciences*, 21(10):3626. ISSN 1422-0067.

[131] Bhowmick, S., Saha, A., Osman, S. M., Alasmary, F. A., Almutairi, T. M., and Islam, M. A. (2021) Structure-based identification of SARS-CoV-2 main protease inhibitors from anti-viral specific chemical libraries: an exhaustive computational screening approach. *Molecular Diversity*, 25(3):1979–1997. ISSN 1573-501X.

[132] Gimeno, A., Mestres-Truyol, J., Ojeda-Montes, M. J., Macip, G., Saldivar-Espinoza, B., Cereto-Massagué, A., et al. (2020) Prediction of Novel Inhibitors of the Main Protease (M-pro) of SARS-CoV-2 through Consensus Docking and Drug Reposition. *International Journal of Molecular Sciences*, 21(11):3793. ISSN 1422-0067.

[133] Zang, Y., Su, M., Wang, Q., Cheng, X., Zhang, W., Zhao, Y., et al. (2022) High-throughput screening of SARS-CoV-2 main and papain-like protease inhibitors. *Protein & Cell*. ISSN 1674-8018.

[134] Dotolo, S., Marabotti, A., Facchiano, A., and Tagliaferri, R. (2020) A review on drug repurposing applicable to COVID-19. *Briefings in Bioinformatics*, 22(2):726–741. ISSN 1477-4054.

[135] Boby, M. L., Fearon, D., Ferla, M., Filep, M., Koekemoer, L., Robinson, M. C., et al. (2023) Open science discovery of potent noncovalent SARS-CoV-2 main protease inhibitors. *Science*, 382(6671). ISSN 1095-9203.

[136] PostEra AI (2023) Mpro Activity Data. `https://covid.postera.ai/covid/activity_data`. Accessed: 2025-08-12.

[137] Du, Z., Wang, L., Bai, Y., Liu, Y., Lau, E. H., Galvani, A. P., et al. (2024) A retrospective cohort study of Paxlovid efficacy depending on treatment time in hospitalized COVID-19 patients. *eLife*, 13. ISSN 2050-084X.

[138] Osipiuk, J., Azizi, S.-A., Dvorkin, S., Endres, M., Jedrzejczak, R., Jones, K. A., et al. (2021) Structure of papain-like protease from SARS-CoV-2 and its complexes with non-covalent inhibitors. *Nature Communications*, 12(1). ISSN 2041-1723.

[139] Yang, M., Mariano, J., Su, R., Smith, C. E., Das, S., Gill, C., et al. (2023) SARS-CoV-2 papain-like protease plays multiple roles in regulating cellular proteins in the endoplasmic reticulum. *Journal of Biological Chemistry*, 299(12):105346. ISSN 0021-9258.

[140] Patchett, S., Lv, Z., Rut, W., Békés, M., Drag, M., Olsen, S. K., et al. (2021) A molecular sensor determines the ubiquitin substrate specificity of SARS-CoV-2 papain-like protease. *Cell Reports*, 36(13):109754. ISSN 2211-1247.

[141] Rhamadianti, A. F., Abe, T., Tanaka, T., Ono, C., Katayama, H., Makino, Y., et al. (2024) SARS-CoV-2 papain-like protease inhibits ISGylation of the viral nucleocapsid protein to evade host anti-viral immunity. *Journal of Virology*, 98(9). ISSN 1098-5514.

[142] Wu, X., Go, M., Nguyen, J. V., Kuchel, N. W., Lu, B. G. C., Zeglinski, K., et al. (2024) Mutational profiling of SARS-CoV-2 papain-like protease reveals requirements for function, structure, and drug escape. *Nature Communications*, 15(1). ISSN 2041-1723.

[143] Gao, X., Qin, B., Chen, P., Zhu, K., Hou, P., Wojdyla, J. A., et al. (2021) Crystal structure of SARS-CoV-2 papain-like protease. *Acta Pharmaceutica Sinica B*, 11 (1):237–245. ISSN 2211-3835.

[144] Barretto, N., Jukneliene, D., Ratia, K., Chen, Z., Mesecar, A. D., and Baker, S. C. (2005) The Papain-Like Protease of Severe Acute Respiratory Syndrome Coronavirus Has Deubiquitinating Activity. *Journal of Virology*, 79(24):15189–15198. ISSN 1098-5514.

[145] Garland, O., Ton, A.-T., Moradi, S., Smith, J. R., Kovacic, S., Ng, K., et al. (2023) Large-Scale Virtual Screening for the Discovery of SARS-CoV-2 Papain-like Protease (PLpro) Non-covalent Inhibitors. *Journal of Chemical Information and Modeling*, 63(7):2158–2169.

[146] Kulandaisamy, R., Kushwaha, T., Dalal, A., Kumar, V., Singh, D., Baswal, K., et al. (2022) Repurposing of FDA Approved Drugs Against SARS-CoV-2 Papain-Like Protease: Computational, Biochemical, and in vitro Studies. *Frontiers in Microbiology*, 13. ISSN 1664-302X.

[147] Cho, C., Li, S. G., Lalonde, T. J., Yang, K. S., Yu, G., Qiao, Y., et al. (2021) Drug Repurposing for the SARS-CoV-2 Papain-Like Protease. *ChemMedChem*, 17(1). ISSN 1860-7187.

[148] Delre, P., Caporuscio, F., Saviano, M., and Mangiatordi, G. F. (2020) Repurposing Known Drugs as Covalent and Non-covalent Inhibitors of the SARS-CoV-2 Papain-Like Protease. *Frontiers in Chemistry*, 8. ISSN 2296-2646.

[149] Tan, H., Hu, Y., Jadhav, P., Tan, B., and Wang, J. (2022) Progress and Challenges in Targeting the SARS-CoV-2 Papain-like Protease. *Journal of Medicinal Chemistry*, 65(11):7561–7580.

[150] Kerti, L. and Frecer, V. (2024) Design of inhibitors of SARS-CoV-2 papain-like protease deriving from GRL0617: Structure–activity relationships. *Bioorganic & Medicinal Chemistry*, 113:117909. ISSN 0968-0896.

[151] Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., et al. (2024) Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 74(3):229–263. ISSN 1542-4863.

[152] Schafer, E. J., Laversanne, M., Sung, H., Soerjomataram, I., Briganti, A., Dahut, W., et al. (2025) Recent Patterns and Trends in Global Prostate Cancer Incidence and Mortality: An Update. *European Urology*, 87(3):302–313. ISSN 0302-2838.

[153] Zhang, W., Cao, G., Wu, F., Wang, Y., Liu, Z., Hu, H., et al. (2023) Global Burden of Prostate Cancer and Association with Socioeconomic Status, 1990–2019: A Systematic Analysis from the Global Burden of Disease Study. *Journal of Epidemiology and Global Health*, 13(3):407–421. ISSN 2210-6014.

[154] Hamdy, F. C., Donovan, J. L., Lane, J. A., Mason, M., Metcalfe, C., Holding, P., et al. (2016) 10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer. *New England Journal of Medicine*, 375(15):1415–1424. ISSN 1533-4406.

[155] Al Hussein Al Awamlh, B., Wallis, C. J. D., Penson, D. F., Huang, L.-C., Zhao, Z., Conwill, R., et al. (2024) Functional Outcomes After Localized Prostate Cancer Treatment. *JAMA*, 331(4):302. ISSN 0098-7484.

[156] Wade, C. and Kyprianou, N. (2018) Profiling Prostate Cancer Therapeutic Resistance. *International Journal of Molecular Sciences*, 19(3):904. ISSN 1422-0067.

[157] Karantanos, T., Corn, P. G., and Thompson, T. C. (2013) Prostate cancer progression after androgen deprivation therapy: mechanisms of castrate resistance and novel therapeutic approaches. *Oncogene*, 32(49):5501–5511. ISSN 1476-5594.

[158] Mahon, K. L., Henshall, S. M., Sutherland, R. L., and Horvath, L. G. (2011) Pathways of chemotherapy resistance in castration-resistant prostate cancer. *Endocrine-Related Cancer*, 18(4):R103–R123. ISSN 1479-6821.

[159] Moreira, D. M., Howard, L. E., Sourbeer, K. N., Amarasekara, H. S., Chow, L. C., Cockrell, D. C., et al. (2017) Predicting Time From Metastasis to Overall Survival in Castration-Resistant Prostate Cancer: Results From SEARCH. *Clinical Genitourinary Cancer*, 15(1):60–66.e2. ISSN 1558-7673.

[160] Kang, J., La Manna, F., Bonollo, F., Sampson, N., Alberts, I. L., Mingels, C., et al. (2022) Tumor microenvironment mechanisms and bone metastatic disease progression of prostate cancer. *Cancer Letters*, 530:156–169. ISSN 0304-3835.

[161] Williams, I. S., McVey, A., Perera, S., O'Brien, J. S., Kostos, L., Chen, K., et al. (2022) Modern paradigms for prostate cancer detection and management. *Medical Journal of Australia*, 217(8):424–433. ISSN 1326-5377.

[162] Descotes, J.-L. (2019) Diagnosis of prostate cancer. *Asian Journal of Urology*, 6(2): 129–136. ISSN 2214-3882.

[163] Sekhoacha, M., Riet, K., Motloung, P., Gumenku, L., Adegoke, A., and Mashele, S. (2022) Prostate Cancer Review: Genetics, Diagnosis, Treatment Options, and Alternative Approaches. *Molecules*, 27(17):5730. ISSN 1420-3049.

[164] Luining, W. I., Cysouw, M. C. F., Meijer, D., Hendrikse, N. H., Boellaard, R., Vis, A. N., et al. (2022) Targeting PSMA Revolutionizes the Role of Nuclear Medicine in Diagnosis and Treatment of Prostate Cancer. *Cancers*, 14(5):1169. ISSN 2072-6694.

[165] Jochumsen, M. R. and Bouchelouche, K. (2024) PSMA PET/CT for Primary Staging of Prostate Cancer - An Updated Overview. *Seminars in Nuclear Medicine*, 54(1): 39–45. ISSN 0001-2998.

[166] Schwenck, J., Sonanini, D., Cotton, J. M., Rammensee, H.-G., la Fougère, C., Zender, L., et al. (2023) Advances in PET imaging of cancer. *Nature Reviews Cancer*, 23(7):474–490. ISSN 1474-1768.

[167] Zhou, J., Gou, Z., Wu, R., Yuan, Y., Yu, G., and Zhao, Y. (2019) Comparison of PSMA-PET/CT, choline-PET/CT, NaF-PET/CT, MRI, and bone scintigraphy in the diagnosis of bone metastases in patients with prostate cancer: a systematic review and meta-analysis. *Skeletal Radiology*, 48(12):1915–1924. ISSN 1432-2161.

[168] Dondi, F., Albano, D., Bertagna, F., and Treglia, G. (2022) Bone Scintigraphy versus PSMA-Targeted PET/CT or PET/MRI in Prostate Cancer: Lessons Learned from Recent Systematic Reviews and Meta-Analyses. *Cancers*, 14(18):4470. ISSN 2072-6694.

[169] Hofman, M. S., Lawrentschuk, N., Francis, R. J., Tang, C., Vela, I., Thomas, P., et al. (2020) Prostate-specific membrane antigen PET-CT in patients with high-risk prostate cancer before curative-intent surgery or radiotherapy (proPSMA): a prospective, randomised, multicentre study. *The Lancet*, 395(10231):1208–1216. ISSN 0140-6736.

[170] Khalil, M. M., Tremoleda, J. L., Bayomy, T. B., and Gsell, W. (2011) Molecular SPECT Imaging: An Overview. *International Journal of Molecular Imaging*, 2011: 1–15. ISSN 2090-1720.

[171] Jansen, F. P. and Vanderheyden, J.-L. (2007) The future of SPECT in a time of PET. *Nuclear Medicine and Biology*, 34(7):733–735. ISSN 0969-8051.

[172] Rahmim, A. and Zaidi, H. (2008) PET versus SPECT: strengths, limitations and challenges. *Nuclear Medicine Communications*, 29(3):193–207. ISSN 0143-3636.

[173] Jones, W., Griffiths, K., Barata, P. C., and Paller, C. J. (2020) PSMA Theranostics: Review of the Current Status of PSMA-Targeted Imaging and Radioligand Therapy. *Cancers*, 12(6):1367. ISSN 2072-6694.

[174] Miszczyk, M., Rajwa, P., Yanagisawa, T., Nowicka, Z., Shim, S. R., Laukhtina, E., et al. (2024) The Efficacy and Safety of Metastasis-directed Therapy in Patients with Prostate Cancer: A Systematic Review and Meta-analysis of Prospective Studies. *European Urology*, 85(2):125–138. ISSN 0302-2838.

[175] Marvaso, G., Corrao, G., Zaffaroni, M., Vincini, M. G., Lorubbio, C., Gandini, S., et al. (2025) ADT with SBRT versus SBRT alone for hormone-sensitive oligorecurrent prostate cancer (RADIOSA): a randomised, open-label, phase 2 clinical trial. *The Lancet Oncology*, 26(3):300–311. ISSN 1470-2045.

[176] Jang, A., Kendi, A. T., and Sartor, O. (2023) Status of PSMA-targeted radioligand therapy in prostate cancer: current data and future trials. *Therapeutic Advances in Medical Oncology*, 15. ISSN 1758-8359.

[177] Sgouros, G., Bodei, L., McDevitt, M. R., and Nedrow, J. R. (2020) Radiopharmaceutical therapy in cancer: clinical advances and challenges. *Nature Reviews Drug Discovery*, 19(9):589–608. ISSN 1474-1784.

[178] Evans, J. C., Malhotra, M., Cryan, J. F., and O'Driscoll, C. M. (2016) The therapeutic and diagnostic potential of the prostate specific membrane antigen/glutamate carboxypeptidase II (PSMA/GCPII) in cancer and neurological disease. *British Journal of Pharmacology*, 173(21):3041–3079. ISSN 1476-5381.

[179] Bakht, M. K. and Beltran, H. (2024) Biological determinants of PSMA expression, regulation and heterogeneity in prostate cancer. *Nature Reviews Urology*, 22(1): 26–45. ISSN 1759-4820.

[180] O'Keefe, D. S., Bacich, D. J., Huang, S. S., and Heston, W. D. (2018) A Perspective on the Evolving Story of PSMA Biology, PSMA-Based Imaging, and Endoradiotherapeutic Strategies. *Journal of Nuclear Medicine*, 59(7):1007–1013. ISSN 2159-662X.

[181] Schülke, N., Varlamova, O. A., Donovan, G. P., Ma, D., Gardner, J. P., Morrissey, D. M., et al. (2003) The homodimer of prostate-specific membrane antigen is a

functional target for cancer therapy. *Proceedings of the National Academy of Sciences*, 100(22):12590–12595. ISSN 1091-6490.

[182] Chang, S. S. (2004) Overview of prostate-specific membrane antigen. *Reviews in urology*, 6(Suppl 10):S13.

[183] Davis, M. I., Bennett, M. J., Thomas, L. M., and Bjorkman, P. J. (2005) Crystal structure of prostate-specific membrane antigen, a tumor marker and peptidase. *Proceedings of the National Academy of Sciences*, 102(17):5981–5986. ISSN 1091-6490.

[184] Piron, S., Verhoeven, J., Vanhove, C., and De Vos, F. (2022) Recent advancements in 18F-labeled PSMA targeting PET radiopharmaceuticals. *Nuclear Medicine and Biology*, 106:29–51.

[185] Barinka, C., Byun, Y., Dusich, C. L., Banerjee, S. R., Chen, Y., Castanares, M., et al. (2008) Interactions between Human Glutamate Carboxypeptidase II and Urea-Based Inhibitors: Structural Characterization. *Journal of Medicinal Chemistry*, 51(24): 7737–7743.

[186] Zhang, A. X., Murelli, R. P., Barinka, C., Michel, J., Cocleaza, A., Jorgensen, W. L., et al. (2010) A Remote Arene-Binding Site on Prostate Specific Membrane Antigen Revealed by Antibody-Recruiting Small Molecules. *Journal of the American Chemical Society*, 132(36):12711–12716.

[187] Kopka, K., Benešová, M., Bařinka, C., Haberkorn, U., and Babich, J. (2017) Glu-Ureido–Based Inhibitors of Prostate-Specific Membrane Antigen: Lessons Learned During the Development of a Novel Class of Low-Molecular-Weight Theranostic Radiotracers. *Journal of Nuclear Medicine*, 58(Supplement 2):17S–26S. ISSN 2159-662X.

[188] El Fakiri, M., Geis, N. M., Ayada, N., Eder, M., and Eder, A.-C. (2021) PSMA-Targeting Radiopharmaceuticals for Prostate Cancer Therapy: Recent Developments and Future Perspectives. *Cancers*, 13(16):3967. ISSN 2072-6694.

[189] Cardinale, J., Roscher, M., Schäfer, M., Geerlings, M., Benešová, M., Bauder-Wüst, U., et al. (2020) Development of PSMA-1007-Related Series of 18F-Labeled Glu-Ureido-Type PSMA Inhibitors. *Journal of Medicinal Chemistry*, 63(19):10897–10907.

[190] Kozikowski, A. P., Nan, F., Conti, P., Zhang, J., Ramadan, E., Bzdega, T., et al. (2001) Design of Remarkably Simple, Yet Potent Urea-Based Inhibitors of Glutamate Carboxypeptidase II (NAALADase). *Journal of Medicinal Chemistry*, 44(3): 298–301. ISSN 1520-4804.

[191] Kozikowski, A. P., Zhang, J., Nan, F., Petukhov, P. A., Grajkowska, E., Wroblewski, J. T., et al. (2004) Synthesis of Urea-Based Inhibitors as Active Site Probes of Glutamate Carboxypeptidase II: Efficacy as Analgesic Agents. *Journal of Medicinal Chemistry*, 47(7):1729–1738.

[192] Vermeulen, K., Vandamme, M., Bormans, G., and Cleeren, F. (2019) Design and Challenges of Radiopharmaceuticals. *Seminars in Nuclear Medicine*, 49(5):339–356. ISSN 0001-2998.

[193] Delbeke, D. (1999) Oncological applications of FDG PET imaging. *Journal of Nuclear Medicine*, 40(10):1706–1715.

[194] Abusalem, M., Martiniova, L., Soebianto, S., DePalatis, L., and Ravizzini, G. (2023) Current Status of Radiolabeled Monoclonal Antibodies Targeting PSMA for Imaging and Therapy. *Cancers*, 15(18):4537. ISSN 2072-6694.

[195] Maresca, K. P., Hillier, S. M., Femia, F. J., Keith, D., Barone, C., Joyal, J. L., et al. (2009) A Series of Halogenated Heterodimeric Inhibitors of Prostate Specific Membrane Antigen (PSMA) as Radiolabeled Probes for Targeting Prostate Cancer. *Journal of Medicinal Chemistry*, 52(2):347–357.

[196] Hillier, S. M., Maresca, K. P., Femia, F. J., Marquis, J. C., Foss, C. A., Nguyen, N., et al. (2009) Preclinical Evaluation of Novel Glutamate-Urea-Lysine Analogues That Target Prostate-Specific Membrane Antigen as Molecular Imaging Pharmaceuticals for Prostate Cancer. *Cancer Research*, 69(17):6932–6940. ISSN 1538-7445.

[197] Barrett, J. A., Coleman, R. E., Goldsmith, S. J., Vallabhajosula, S., Petry, N. A., Cho, S., et al. (2013) First-in-Man Evaluation of 2 High-Affinity PSMA-Avid Small Molecules for Imaging Prostate Cancer. *Journal of Nuclear Medicine*, 54(3): 380–387. ISSN 2159-662X.

[198] Eder, M., Schäfer, M., Bauder-Wüst, U., Hull, W.-E., Wängler, C., Mier, W., et al. (2012) 68Ga-Complex Lipophilicity and the Targeting Property of a Urea-Based PSMA Inhibitor for PET Imaging. *Bioconjugate Chemistry*, 23(4):688–697.

[199] Afshar-Oromieh, A., Haberkorn, U., Eder, M., Eisenhut, M., and Zechmann, C. (2012) [68Ga]Gallium-labelled PSMA ligand as superior PET tracer for the diagnosis of prostate cancer: comparison with 18F-FECH. *European Journal of Nuclear Medicine and Molecular Imaging*, 39(6):1085–1086. ISSN 1619-7089.

[200] Benešová, M., Schäfer, M., Bauder-Wüst, U., Afshar-Oromieh, A., Kratochwil, C., Mier, W., et al. (2015) Preclinical Evaluation of a Tailor-Made DOTA-Conjugated PSMA Inhibitor with Optimized Linker Moiety for Imaging and Endoradiotherapy of Prostate Cancer. *Journal of Nuclear Medicine*, 56(6):914–920. ISSN 2159-662X.

[201] Hofman, M. S., Violet, J., Hicks, R. J., Ferdinandus, J., Thang, S. P., Akhurst, T., et al. (2018) [177 Lu]-PSMA-617 radionuclide treatment in patients with metastatic castration-resistant prostate cancer (LuPSMA trial): a single-centre, single-arm, phase 2 study. *The Lancet Oncology*, 19(6):825–833. ISSN 1470-2045.

[202] Sun, M., Niaz, M. O., Nelson, A., Skafida, M., Niaz, M. J., and Niaz, M. O. (2020) Review of 177Lu-PSMA-617 in patients with metastatic castration-resistant prostate cancer. *Cureus*, 12(6).

[203] Hofman, M. S., Emmett, L., Sandhu, S., Iravani, A., Joshua, A. M., Goh, J. C., et al. (2021) [177Lu] Lu-PSMA-617 versus cabazitaxel in patients with metastatic castration-resistant prostate cancer (TheraP): a randomised, open-label, phase 2 trial. *The Lancet*, 397(10276):797–804.

[204] Sartor, O., De Bono, J., Chi, K. N., Fizazi, K., Herrmann, K., Rahbar, K., et al. (2021) Lutetium-177–PSMA-617 for metastatic castration-resistant prostate cancer. *New England Journal of Medicine*, 385(12):1091–1103.

[205] Dhiantravan, N., Emmett, L., Joshua, A. M., Pattison, D. A., Francis, R. J., Williams, S., et al. (2021) UpFrontPSMA: a randomized phase 2 study of sequential 177Lu-PSMA-617 and docetaxel vs docetaxel in metastatic hormone-naïve prostate cancer (clinical trial protocol). *BJU international*, 128(3):331–342.

[206] Morris, M. J., De Bono, J. S., Chi, K. N., Fizazi, K., Herrmann, K., Rahbar, K., et al. (2021) Phase III study of lutetium-177-PSMA-617 in patients with metastatic castration-resistant prostate cancer (VISION).

[207] Fallah, J., Agrawal, S., Gittleman, H., Fiero, M. H., Subramaniam, S., John, C., et al. (2022) FDA Approval Summary: Lutetium Lu 177 Vipivotide Tetraxetan for Patients with Metastatic Castration-Resistant Prostate Cancer. *Clinical Cancer Research*, 29(9):1651–1657. ISSN 1557-3265.

[208] Cardinale, J., Schäfer, M., Benešová, M., Bauder-Wüst, U., Leotta, K., Eder, M., et al. (2016) Preclinical Evaluation of 18F-PSMA-1007, a New Prostate-Specific Membrane Antigen Ligand for Prostate Cancer Imaging. *Journal of Nuclear Medicine*, 58(3):425–431. ISSN 2159-662X.

[209] Chen, Y., Pullambhatla, M., Foss, C. A., Byun, Y., Nimmagadda, S., Senthamizhchelvan, S., et al. (2011) 2-(3-1-Carboxy-5-[(6-[18F]Fluoro-Pyridine-3-Carbonyl)-Amino]-Pentyl-Ureido)-Pentanedioic Acid, [18F]DCFPyL, a PSMA-Based PET Imaging Agent for Prostate Cancer. *Clinical Cancer Research*, 17(24): 7645–7653. ISSN 1557-3265.

[210] Rowe, S. P., Macura, K. J., Mena, E., Blackford, A. L., Nadal, R., Antonarakis, E. S., et al. (2016) PSMA-Based [18F]DCFPyL PET/CT Is Superior to Conventional Imaging for Lesion Detection in Patients with Metastatic Prostate Cancer. *Molecular Imaging and Biology*, 18(3):411–419. ISSN 1860-2002.

[211] Giesel, F. L., Hadaschik, B., Cardinale, J., Radtke, J., Vinsensia, M., Lehnert, W., et al. (2016) F-18 labelled PSMA-1007: biodistribution, radiation dosimetry and histopathological validation of tumor lesions in prostate cancer patients. *European Journal of Nuclear Medicine and Molecular Imaging*, 44(4):678–688. ISSN 1619-7089.

[212] Giesel, F. L., Kesch, C., Yun, M., Cardinale, J., Haberkorn, U., Kopka, K., et al. (2017) 18F-PSMA-1007 PET/CT Detects Micrometastases in a Patient With Biochemically Recurrent Prostate Cancer. *Clinical Genitourinary Cancer*, 15(3):e497–e499. ISSN 1558-7673.

[213] Weineisen, M., Simecek, J., Schottelius, M., Schwaiger, M., and Wester, H.-J. (2014) Synthesis and preclinical evaluation of DOTAGA-conjugated PSMA ligands for functional imaging and endoradiotherapy of prostate cancer. *EJNMMI Research*, 4 (1). ISSN 2191-219X.

[214] Weineisen, M., Schottelius, M., Simecek, J., Baum, R. P., Yildiz, A., Beykan, S., et al. (2015) 68Ga- and 177Lu-Labeled PSMA I&T: Optimization of a PSMA-Targeted Theranostic Concept and First Proof-of-Concept Human Studies. *Journal of Nuclear Medicine*, 56(8):1169–1176. ISSN 2159-662X.

[215] Heck, M. M., Retz, M., D'Alessandria, C., Rauscher, I., Scheidhauer, K., Maurer, T., et al. (2016) Systemic Radioligand Therapy with 177 Lu Labeled Prostate Specific Membrane Antigen Ligand for Imaging and Therapy in Patients with Metastatic

Castration Resistant Prostate Cancer. *Journal of Urology*, 196(2):382–391. ISSN 1527-3792.

[216] Schottelius, M., Wirtz, M., Eiber, M., Maurer, T., and Wester, H.-J. (2015) [111In]PSMA-I&T: expanding the spectrum of PSMA-I&T applications towards SPECT and radioguided surgery. *EJNMMI Research*, 5(1). ISSN 2191-219X.

[217] Robu, S., Schottelius, M., Eiber, M., Maurer, T., Gschwend, J., Schwaiger, M., et al. (2016) Preclinical Evaluation and First Patient Application of 99mTc-PSMA-I&S for SPECT Imaging and Radioguided Surgery in Prostate Cancer. *Journal of Nuclear Medicine*, 58(2):235–242. ISSN 2159-662X.

[218] García-Pérez, F. O., Davanzo, J., López-Buenrostro, S., Santos-Cuevas, C., Ferro-Flores, G., Jímenez-Ríos, M. A., et al. (2018) Head to head comparison performance of 99mTc-EDDA/HYNIC-iPSMA SPECT/CT and 68Ga-PSMA-11 PET/CT a prospective study in biochemical recurrence prostate cancer patients. *American journal of nuclear medicine and molecular imaging*, 8(5):332.

[219] Ferro-Flores, G., Luna-Gutiérrez, M., Ocampo-García, B., Santos-Cuevas, C., Azorín-Vega, E., Jiménez-Mancilla, N., et al. (2017) Clinical translation of a PSMA inhibitor for 99m Tc-based SPECT. *Nuclear Medicine and Biology*, 48:36–44. ISSN 0969-8051.

[220] Cardinale, J., Giesel, F. L., Wensky, C., Rathke, H. G., Haberkorn, U., and Kratochwil, C. (2023) PSMA-GCK01: A Generator-Based 99mTc/188Re Theranostic Ligand for the Prostate-Specific Membrane Antigen. *Journal of Nuclear Medicine*, 64(7):1069–1075. ISSN 2159-662X.

[221] Mamlins, E., Scharbert, L., Cardinale, J., Krotov, M., Winter, E., Rathke, H., et al. (2023) The Theranostic Optimization of PSMA-GCK01 Does Not Compromise the Imaging Characteristics of [99mTc]Tc-PSMA-GCK01 Compared to Dedicated Diagnostic [99mTc]Tc-EDDA/HYNIC-iPSMA in Prostate Cancer. *Molecular Imaging and Biology*, 26(1):81–89. ISSN 1860-2002.

[222] Wurzer, A., Di Carlo, D., Schmidt, A., Beck, R., Eiber, M., Schwaiger, M., et al. (2019) Radiohybrid Ligands: A Novel Tracer Concept Exemplified by 18F- or 68Ga-Labeled rhPSMA Inhibitors. *Journal of Nuclear Medicine*, 61(5):735–742. ISSN 2159-662X.

[223] Eiber, M., Kroenke, M., Wurzer, A., Ulbrich, L., Jooß, L., Maurer, T., et al. (2019) 18F-rhPSMA-7 PET for the Detection of Biochemical Recurrence of Prostate Cancer

After Radical Prostatectomy. *Journal of Nuclear Medicine*, 61(5):696–701. ISSN 2159-662X.

[224] Schrödinger, LLC (2022) Maestro, Release 2022-3.

[225] Ravindranath, P. A., Forli, S., Goodsell, D. S., Olson, A. J., and Sanner, M. F. (2015) AutoDockFR: Advances in Protein-Ligand Docking with Explicitly Specified Binding Site Flexibility. *PLOS Computational Biology*, 11(12):e1004586. ISSN 1553-7358.

[226] Huey, R., Morris, G. M., Olson, A. J., and Goodsell, D. S. (2007) A semiempirical free energy force field with charge-based desolvation. *Journal of Computational Chemistry*, 28(6):1145–1152. ISSN 1096-987X.

[227] Liu, J. and Wang, R. (2015) Classification of Current Scoring Functions. *Journal of Chemical Information and Modeling*, 55(3):475–482. ISSN 1549-960X.

[228] Meli, R., Morris, G. M., and Biggin, P. C. (2022) Scoring Functions for Protein-Ligand Binding Affinity Prediction Using Structure-based Deep Learning: A Review. *Frontiers in Bioinformatics*, 2. ISSN 2673-7647.

[229] Durant, G., Boyles, F., Birchall, K., Marsden, B., and Deane, C. M. (2025) Robustly interrogating machine learning-based scoring functions: what are they learning? *Bioinformatics*, 41(2). ISSN 1367-4811.

[230] Mariani, V., Biasini, M., Barbato, A., and Schwede, T. (2013) lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics*, 29(21):2722–2728. ISSN 1367-4811.

[231] Zhang, Y. and Skolnick, J. (2004) Scoring function for automated assessment of protein structure template quality. *Proteins: Structure, Function, and Bioinformatics*, 57(4):702–710. ISSN 1097-0134.

[232] Zhang, Y., Forli, S., Omelchenko, A., and Sanner, M. F. (2019) AutoGridFR: Improvements on AutoDock Affinity Maps and Associated Software Tools. *Journal of Computational Chemistry*, 40(32):2882–2886. ISSN 1096-987X.

[233] Morris, G. M., Goodsell, D. S., Pique, M., Lindstrom, W., Huey, R., Forli, S., et al. (2010) Automated docking of flexible ligands to flexible receptors. *user guide AutoDock*.

[234] Ravindranath, P. A. and Sanner, M. F. (2016) AutoSite: an automated approach for pseudo-ligands prediction—from ligand-binding sites identification to predicting key ligand atoms. *Bioinformatics*, 32(20):3142–3149. ISSN 1367-4803.

[235] Ferreira, L., Dos Santos, R., Oliva, G., and Andricopulo, A. (2015) Molecular Docking and Structure-Based Drug Design Strategies. *Molecules*, 20(7):13384–13421. ISSN 1420-3049.

[236] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953) Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092. ISSN 1089-7690.

[237] Fletcher, R. (1987) Practical Methods of Optimization. John Wiley & Sons, 2nd edition.

[238] Shanno, D. F. (1970) Conditioning of quasi-Newton methods for function minimization. *Mathematics of computation*, 24(111):647–656.

[239] Dunbrack, R. L. and Cohen, F. E. (1997) Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Science*, 6(8):1661–1681. ISSN 1469-896X.

[240] Schrödinger, LLC (2015) The PyMOL Molecular Graphics System, Version 2.5.

[241] Allen, M. P. and others (2004) Introduction to molecular dynamics simulation. *Computational soft matter: from synthetic polymers to proteins*, 23(1):1–28.

[242] Lindahl, E. R. (2008) Molecular Modeling of Proteins.

[243] Hospital, A., Goni, J. R., Orozco, M., and Gelpi, J. L. (2015) Molecular dynamics simulations: advances and applications. *Adv. Appl. Bioinform. Chem.*, 8:37–47.

[244] González, M. (2011) Force fields and molecular dynamics simulations. *École thématique de la Société Française de la Neutronique*, 12:169–200.

[245] Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A smooth particle mesh Ewald method. *The Journal of chemical physics*, 103(19):8577–8593.

[246] Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, 98(12):10089–10092.

[247] Leimkuhler, B. and Matthews, C. (2016) Molecular Dynamics. Springer.

[248] Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., et al. (2009) CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry*, 30(10):1545–1614. ISSN 1096-987X.

[249] MacKerell, A. D. J., Bashford, D., Bellott, M., Dunbrack, R. L. J., Evanseck, J. D., Field, M. J., et al. (1998) All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *The Journal of Physical Chemistry B*, 102(18): 3586–3616.

[250] Christen, M., Hünenberger, P. H., Bakowies, D., Baron, R., Bürgi, R., Geerke, D. P., et al. (2005) The GROMOS software for biomolecular simulation: GROMOS05. *Journal of Computational Chemistry*, 26(16):1719–1751. ISSN 1096-987X.

[251] Oostenbrink, C., Villa, A., Mark, A. E., and Van Gunsteren, W. F. (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *Journal of computational chemistry*, 25 (13):1656–1676.

[252] Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*, 11(8):3696–3713. ISSN 1549-9626.

[253] Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., et al. (1995) A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society*, 117(19):5179–5197.

[254] Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 65(3):712–725. ISSN 1097-0134.

[255] Allen, M. P. and Tildesley, D. J. (2017) Computer simulation of liquids. Oxford university press.

[256] Frenkel, D. and Smit, B. (2001) Understanding molecular simulation: from algorithms to applications, volume 1. Elsevier.

[257] Verlet, L. (1967) Computer" experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical review*, 159(1):98.

[258] Cuendet, M. A. and van Gunsteren, W. F. (2007) On the calculation of velocity-dependent properties in molecular dynamics simulations using the leapfrog integration algorithm. *The Journal of chemical physics*, 127(18):184102.

[259] Swope, W. C., Andersen, H. C., Berens, P. H., and Wilson, K. R. (1982) A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *The Journal of chemical physics*, 76(1):637–649.

[260] Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of computational physics*, 23(3):327–341.

[261] Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997) LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.*, 18(12): 1463–1472.

[262] Kutzner, C., Páll, S., Fechner, M., Esztermann, A., de Groot, B. L., and Grubmüller, H. (2019) More bang for your buck: Improved use of GPU nodes for GROMACS 2018. *Journal of computational chemistry*, 40(27):2418–2431.

[263] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2):926–935.

[264] Cauchy, M. A. (1847) Méthode généerale pour la résolution des systèmes d' éequations simultanéees. *CR Hebd. Acad. Sci.*, 25:536–538.

[265] Bussi, G., Donadio, D., and Parrinello, M. (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126(1):014101.

[266] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81(8):3684–3690.

[267] Nosé, S. (1984) A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, 81(1):511–519.

[268] Hoover, W. (1985) Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev.*, 31:1695–1697.

[269] Parrinello, M. and Rahman, A. (1980) Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. *Phys. Rev. Lett.*, 45:1196–1199.

[270] Parrinello, M. and Rahman, A. (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.*, 52(12):7182–7190.

[271] Parrinello, M. and Rahman, A. (1982) Strain fluctuations and elastic constants. *J. Chem. Phys.*, 76(5):2662–2666.

[272] Wang, E., Sun, H., Wang, J., Wang, Z., Liu, H., Zhang, J. Z. H., et al. (2019) End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chemical Reviews*, 119(16):9478–9508. ISSN 1520-6890.

[273] Homeyer, N. and Gohlke, H. (2012) Free Energy Calculations by the Molecular Mechanics Poisson–Boltzmann Surface Area Method. *Molecular Informatics*, 31 (2):114–122. ISSN 1868-1751.

[274] Zhao, C., Rong, Y., Shi, S., Gao, W.-c., and Zhang, C. (2024) A novel method for synthesizing authentic SARS-CoV-2 main protease. *Protein Expression and Purification*, 222:106531. ISSN 1046-5928.

[275] Hanahan, D. (1983) Studies on transformation of Escherichia coli with plasmids. *Journal of molecular biology*, 166(4):557–580.

[276] Studier, F. W. (2005) Protein production by auto-induction in high-density shaking cultures. *Protein expression and purification*, 41(1):207–234.

[277] Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., et al. (2005) Protein Identification and Analysis Tools on the ExPASy Server, pages 571–607. Humana Press, Totowa, NJ. ISBN 978-1-59259-890-8.

[278] Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *nature*, 227(5259):680–685.

[279] Tien, M. Z., Sydykova, D. K., Meyer, A. G., and Wilke, C. O. (2013) PeptideBuilder: A simple Python library to generate model peptides. *PeerJ*, 1:e80. ISSN 2167-8359.

[280] Garton, M., Nim, S., Stone, T. A., Wang, K. E., Deber, C. M., and Kim, P. M. (2018) Method to generate highly stable D-amino acid analogs of bioactive helical peptides using a mirror image of the entire PDB. *Proceedings of the National Academy of Sciences*, 115(7):1505–1510.

[281] Touw, W. G., Baakman, C., Black, J., te Beek, T. A. H., Krieger, E., Joosten, R. P., et al. (2014) A series of PDB-related databanks for everyday needs. *Nucleic Acids Research*, 43(D1):D364–D368. ISSN 0305-1048.

[282] Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12): 2577–2637. ISSN 1097-0282.

[283] Daura, X., Gademann, K., Jaun, B., Seebach, D., van Gunsteren, W. F., and Mark, A. E. (1999) Peptide Folding: When Simulation Meets Experiment. *Angewandte Chemie International Edition*, 38(1–2):236–240. ISSN 1521-3773.

[284] Laurent, B., Chavent, M., Cragnolini, T., Dahl, A. C. E., Pasquali, S., Derreumaux, P., et al. (2015) Epock: rapid analysis of protein pocket dynamics. *Bioinformatics*, 31(9):1478–1480.

[285] Dallakyan, S. (2010) MGLTools. `http://mgltools.scripps.edu/`.

[286] Eberle, R. J., Olivier, D. S., Amaral, M. S., Gering, I., Willbold, D., Arni, R. K., et al. (2021) The Repurposed Drugs Suramin and Quinacrine Cooperatively Inhibit SARS-CoV-2 3CLpro In Vitro. *Viruses*, 13(5):873. ISSN 1999-4915.

[287] Chen, H., Wei, P., Huang, C., Tan, L., Liu, Y., and Lai, L. (2006) Only One Protomer Is Active in the Dimer of SARS 3C-like Proteinase. *Journal of Biological Chemistry*, 281(20):13894–13898. ISSN 0021-9258.

[288] Holst, M., Baker, N., and Wang, F. (2000) Adaptive multilevel finite element solution of the Poisson–Boltzmann equation I. Algorithms and examples. *J. Comput. Chem.*, 21(15):1319–1342.

[289] Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272. ISSN 1548-7105.

[290] Shaqra, A. M., Zvornicanin, S. N., Huang, Q. Y. J., Lockbaum, G. J., Knapp, M., Tandeske, L., et al. (2022) Defining the substrate envelope of SARS-CoV-2 main protease to predict and avoid drug resistance. *Nature Communications*, 13(1). ISSN 2041-1723.

[291] Báez-Santos, Y. M., St. John, S. E., and Mesecar, A. D. (2015) The SARS-coronavirus papain-like protease: Structure, function and inhibition by designed antiviral compounds. *Antiviral Research*, 115:21–38. ISSN 0166-3542.

[292] Xue, X., Yang, H., Shen, W., Zhao, Q., Li, J., Yang, K., et al. (2007) Production of Authentic SARS-CoV Mpro with Enhanced Activity: Application as a Novel Tag-cleavage Endopeptidase for Protein Overproduction. *Journal of Molecular Biology*, 366(3):965–975. ISSN 0022-2836.

[293] Rong, Y., Zhang, C., Gao, W.-C., and Zhao, C. (2023) Optimization of the expression of the main protease from SARS-CoV-2. *Protein Expression and Purification*, 203: 106208. ISSN 1046-5928.

[294] Cesaratto, F., Burrone, O. R., and Petris, G. (2016) Tobacco Etch Virus protease: A shortcut across biotechnologies. *Journal of Biotechnology*, 231:239–249. ISSN 0168-1656.

[295] Ma, C., Sacco, M. D., Xia, Z., Lambrinidis, G., Townsend, J. A., Hu, Y., et al. (2021) Discovery of SARS-CoV-2 Papain-like Protease Inhibitors through a Combination of High-Throughput Screening and a FlipGFP-Based Reporter Assay. *ACS Central Science*, 7(7):1245–1260. ISSN 2374-7951.

[296] Kappel, K., Wereszczynski, J., Clubb, R. T., and McCammon, J. A. (2012) The binding mechanism, multiple binding modes, and allosteric regulation of Staphylococcus aureus Sortase A probed by molecular dynamics simulations. *Protein Science*, 21(12):1858–1871. ISSN 1469-896X.

[297] Siebenmorgen, T. and Zacharias, M. (2020) Efficient Refinement and Free Energy Scoring of Predicted Protein–Protein Complexes Using Replica Exchange with Repulsive Scaling. *Journal of Chemical Information and Modeling*, 60(11):5552–5562.

[298] Wang, J., Alekseenko, A., Kozakov, D., and Miao, Y. (2019) Improved Modeling of Peptide-Protein Binding Through Global Docking and Accelerated Molecular Dynamics Simulations. *Frontiers in Molecular Biosciences*, 6. ISSN 2296-889X.

[299] Wang, J. and Miao, Y. (2020) Peptide Gaussian accelerated molecular dynamics (Pep-GaMD): Enhanced sampling and free energy and kinetics calculations of peptide binding. *The Journal of Chemical Physics*, 153(15). ISSN 1089-7690.

[300] Wang, J., Koirala, K., Do, H. N., and Miao, Y. (2024) PepBinding: A Workflow for Predicting Peptide Binding Structures by Combining Peptide Docking and Peptide Gaussian Accelerated Molecular Dynamics Simulations. *The Journal of Physical Chemistry B*, 128(30):7332–7340.

[301] Tiberti, M., Terkelsen, T., Degn, K., Beltrame, L., Cremers, T. C., da Piedade, I., et al. (2022) MutateX: an automated pipeline for in silico saturation mutagenesis of protein structures and structural ensembles. *Briefings in Bioinformatics*, 23(3). ISSN 1477-4054.

[302] Guerois, R., Nielsen, J. E., and Serrano, L. (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *Journal of molecular biology*, 320(2):369–387.

[303] Thieker, D. F., Maguire, J. B., Kudlacek, S. T., Leaver-Fay, A., Lyskov, S., and Kuhlman, B. (2022) Stabilizing proteins, simplified: A Rosetta-based webtool for predicting favorable mutations. *Protein Science*, 31(10). ISSN 1469-896X.

[304] Barlow, K. A., Ó Conchúir, S., Thompson, S., Suresh, P., Lucas, J. E., Heinonen, M., et al. (2018) Flex ddG: Rosetta ensemble-based estimation of changes in protein–protein binding affinity upon mutation. *The Journal of Physical Chemistry B*, 122 (21):5389–5399.

[305] Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J. (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, 118(45):11225–11236.

[306] Funk, L.-M., Poschmann, G., Rabe von Pappenheim, F., Chari, A., Stegmann, K. M., Dickmanns, A., et al. (2024) Multiple redox switches of the SARS-CoV-2 main protease in vitro provide opportunities for drug design. *Nature Communications*, 15 (1). ISSN 2041-1723.

[307] Kuan, Y., Chu, H.-F., Hsu, P.-H., Hsu, K.-C., Lin, T.-H., Huang, C.-H., et al. (2024) Disulfiram inhibits coronaviral main protease by conjugating to its substrate entry site. *International Journal of Biological Macromolecules*, 276:133955. ISSN 0141-8130.

[308] Fetse, J., Kandel, S., Mamani, U.-F., and Cheng, K. (2023) Recent advances in the development of therapeutic peptides. *Trends in Pharmacological Sciences*, 44(7): 425–441.

[309] Fernandez-Escamilla, A.-M., Rousseau, F., Schymkowitz, J., and Serrano, L. (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nature Biotechnology*, 22(10):1302–1306. ISSN 1546-1696.

[310] Eschbach, E., Deibler, K., Korani, D., and Swanson, S. (2026) Predicting Peptide Aggregation with Protein Language Model Embeddings. *Journal of Chemical Information and Modeling*, 66(9):5206–5216.

[311] Levine, H. (1993) Thioflavine T interaction with synthetic Alzheimer's disease β-amyloid peptides: Detection of amyloid aggregation in solution. *Protein Science*, 2(3):404–410. ISSN 1469-896X.

[312] Pignataro, M. F., Herrera, M. G., and Dodero, V. I. (2020) Evaluation of Peptide/Protein Self-Assembly and Aggregation by Spectroscopic Methods. *Molecules*, 25(20):4854. ISSN 1420-3049.

[313] Rut, W., Lv, Z., Zmudzinski, M., Patchett, S., Nayak, D., Snipas, S. J., et al. (2020) Activity profiling and crystal structures of inhibitor-bound SARS-CoV-2 papain-like protease: A framework for anti–COVID-19 drug design. *Science Advances*, 6(42). ISSN 2375-2548.

[314] Kreutzer, A. G., Krumberger, M., Diessner, E. M., Parrocha, C. M. T., Morris, M. A., Guaglianone, G., et al. (2021) A cyclic peptide inhibitor of the SARS-CoV-2 main protease. *European Journal of Medicinal Chemistry*, 221:113530. ISSN 0223-5234.

[315] Johansen-Leete, J., Ullrich, S., Fry, S. E., Frkic, R., Bedding, M. J., Aggarwal, A., et al. (2022) Antiviral cyclic peptides targeting the main protease of SARS-CoV-2. *Chemical Science*, 13(13):3826–3836. ISSN 2041-6539.

[316] Zhao, Y., Fang, C., Zhang, Q., Zhang, R., Zhao, X., Duan, Y., et al. (2021) Crystal structure of SARS-CoV-2 main protease in complex with protease inhibitor PF-07321332. *Protein & Cell*, 13(9):689–693. ISSN 1674-8018.

[317] Liu, X., Ren, X., Hua, M., Liu, F., Ren, X., Sui, C., et al. (2023) Progress of SARS-CoV-2 Main protease peptide-like inhibitors. *Chemical Biology & Drug Design*, 103(1). ISSN 1747-0285.

[318] Noske, G. D., de Souza Silva, E., de Godoy, M. O., Dolci, I., Fernandes, R. S., Guido, R. V. C., et al. (2023) Structural basis of nirmatrelvir and ensitrelvir activity against naturally occurring polymorphisms of the SARS-CoV-2 main protease. *Journal of Biological Chemistry*, 299(3):103004. ISSN 0021-9258.

[319] Lockbaum, G. J., Reyes, A. C., Lee, J. M., Tilvawala, R., Nalivaika, E. A., Ali, A., et al. (2021) Crystal Structure of SARS-CoV-2 Main Protease in Complex with the Non-Covalent Inhibitor ML188. *Viruses*, 13(2):174. ISSN 1999-4915.

[320] Kitamura, N., Sacco, M. D., Ma, C., Hu, Y., Townsend, J. A., Meng, X., et al. (2022) Expedited Approach toward the Rational Design of Noncovalent SARS-CoV-2 Main Protease Inhibitors. *Journal of Medicinal Chemistry*, 65(4):2848–2865.

[321] Holden, J. K., Pavlovicz, R., Gobbi, A., Song, Y., and Cunningham, C. N. (2022) Computational Site Saturation Mutagenesis of Canonical and Non-Canonical Amino Acids to Probe Protein-Peptide Interactions. *Frontiers in Molecular Biosciences*, 9. ISSN 2296-889X.

[322] Traboulsi, H., Larkin, H., Bonin, M.-A., Volkov, L., Lavoie, C. L., and Marsault, É. (2015) Macrocyclic Cell Penetrating Peptides: A Study of Structure-Penetration Properties. *Bioconjugate Chemistry*, 26(3):405–411.

[323] Madani, F., Lindberg, S., Langel, U., Futaki, S., and Gräslund, A. (2011) Mechanisms of Cellular Uptake of Cell-Penetrating Peptides. *Journal of Biophysics*, 2011: 1–10. ISSN 1687-8019.

[324] Leung, D., Abbenante, G., and Fairlie, D. P. (2000) Protease Inhibitors: Current Status and Future Prospects. *Journal of Medicinal Chemistry*, 43(3):305–341.

[325] Gourni, E. and Henriksen, G. (2017) Metal-Based PSMA Radioligands. *Molecules*, 22(4):523. ISSN 1420-3049.

[326] Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., et al. (2024) PubChem 2025 update. *Nucleic Acids Research*, 53(D1):D1516–D1525. ISSN 1362-4962.

[327] Viola, N. A., Rarig, R. S., Ouellette, W., and Doyle, R. P. (2006) Synthesis, structure and thermal analysis of the gallium complex of 1, 4, 7, 10-tetraazacyclo-dodecane-N, N', N", N"'-tetraacetic acid (DOTA). *Polyhedron*, 25(18):3457–3462. ISSN 0277-5387.

[328] Grummon, G., Rajagopalan, R., Palenik, G. J., Koziol, A. E., and Nosco, D. L. (1995) Synthesis, Characterization and Crystal Structures of Technetium(V)-Oxo Complexes Useful in Nuclear Medicine. 1. Complexes of Mercaptoacetylglycylglycylglycine (MAG3) and Its Methyl Ester Derivative (MAG3OMe). *Inorganic Chemistry*, 34(7):1764–1772. ISSN 1520-510X.

[329] Larsen, S. K., Jenkins, B. G., Memon, N. G., and Lauffer, R. B. (1990) Structure-affinity relationships in the binding of unsubstituted iron phenolate complexes to human serum albumin. Molecular structure of iron(III) N, N'-bis(2-hydroxybenzyl)ethylenediamine-N, N'-diacetate. *Inorganic Chemistry*, 29(6):1147–1152. ISSN 1520-510X.

[330] Groom, C. R., Bruno, I. J., Lightfoot, M. P., and Ward, S. C. (2016) The Cambridge Structural Database. *Acta Crystallographica Section B Structural Science, Crystal Engineering and Materials*, 72(2):171–179. ISSN 2052-5206.

[331] Frisch, M., Trucks, G., Schlegel, H., Scuseria, G., Robb, M., Cheeseman, J., et al. (2009) Gaussian 09, Revision A.02. *Gaussian, Inc., Wallingford CT*.

[332] Li, P. and Merz, K. M. (2016) MCPB.py: A Python Based Metal Center Parameter Builder. *Journal of Chemical Information and Modeling*, 56(4):599–604. ISSN 1549-960X.

[333] Case, D. A., Aktulga, H. M., Belfon, K., Cerutti, D. S., Cisneros, G. A., Cruzeiro, V. W. D., et al. (2023) AmberTools. *Journal of Chemical Information and Modeling*, 63(20):6183–6191.

[334] Sousa da Silva, A. W. and Vranken, W. F. (2012) ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Research Notes*, 5(1). ISSN 1756-0500.

[335] Schäfer, M., Bauder-Wüst, U., Roscher, M., Motlová, L., Kutilová, Z., Remde, Y., et al. (2025) Structure–Activity Relationships and Biological Insights into PSMA-617 and Its Derivatives with Modified Lipophilic Linker Regions. *ACS Omega*, 10 (7):7077–7090. ISSN 2470-1343.

[336] Wester, H.-J. and Schottelius, M. (2019) PSMA-Targeted Radiopharmaceuticals for Imaging and Therapy. *Seminars in Nuclear Medicine*, 49(4):302–312. ISSN 0001-2998.

[337] Wirtz, M., Schmidt, A., Schottelius, M., Robu, S., Günther, T., Schwaiger, M., et al. (2018) Synthesis and in vitro and in vivo evaluation of urea-based PSMA inhibitors with increased lipophilicity. *EJNMMI Research*, 8(1). ISSN 2191-219X.

[338] Kunert, J.-P., Müller, M., Günther, T., Stopper, L., Urtz-Urban, N., Beck, R., et al. (2023) Synthesis and preclinical evaluation of novel 99mTc-labeled PSMA ligands for radioguided surgery of prostate cancer. *EJNMMI Research*, 13(1). ISSN 2191-219X.

[339] Gaudreault, F. and Najmanovich, R. J. (2015) FlexAID: Revisiting Docking on Non-Native-Complex Structures. *Journal of Chemical Information and Modeling*, 55(7):1323–1336.

[340] Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., and Taylor, R. D. (2003) Improved protein–ligand docking using GOLD. *Proteins: Structure, Function, and Bioinformatics*, 52(4):609–623.

[341] Korb, O., Olsson, T. S. G., Bowden, S. J., Hall, R. J., Verdonk, M. L., Liebeschuetz, J. W., et al. (2012) Potential and Limitations of Ensemble Docking. *Journal of Chemical Information and Modeling*, 52(5):1262–1274.

[342] Wang, K., Chodera, J. D., Yang, Y., and Shirts, M. R. (2013) Identifying ligand binding sites and poses using GPU-accelerated Hamiltonian replica exchange molecular dynamics. *Journal of Computer-Aided Molecular Design*, 27(12):989–1007. ISSN 1573-4951.

[343] Wang, L., Wu, Y., Deng, Y., Kim, B., Pierce, L., Krilov, G., et al. (2015) Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *Journal of the American Chemical Society*, 137(7):2695–2703.

A. Appendix

A.1. Supplementary Figures

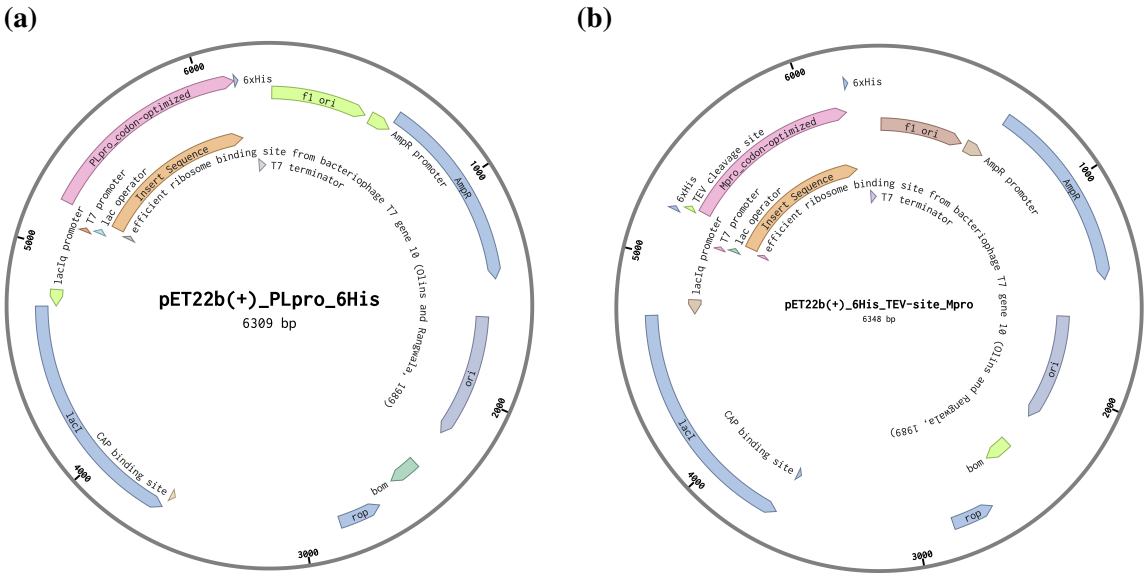
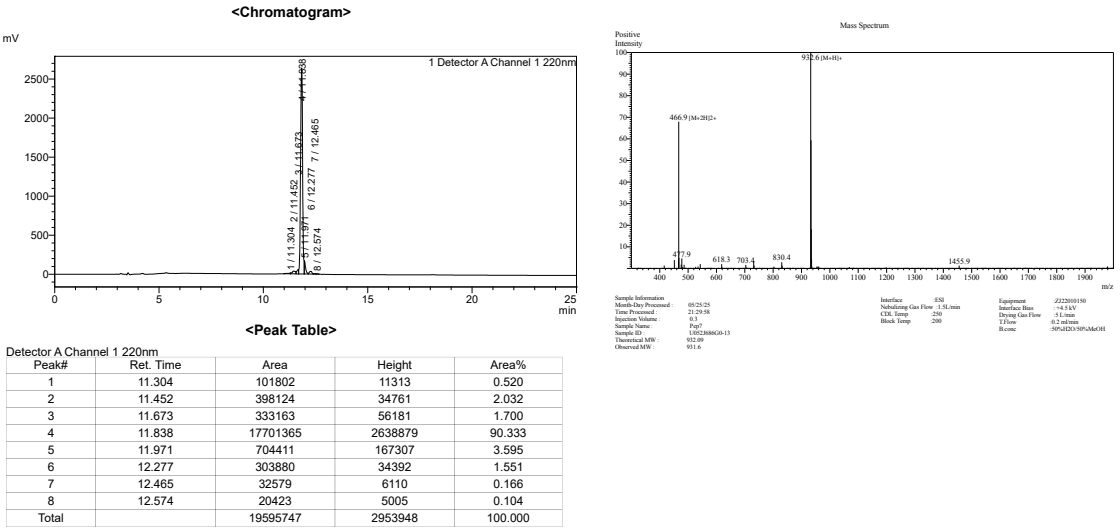


Figure A1.: Plasmid maps for recombinant protein expression in *E. coli*. (a) Plasmid map of pET22b_PLpro_6His. (b) Plasmid map of pET22b_6His_TEV-site_Mpro.

(a) **(b)**



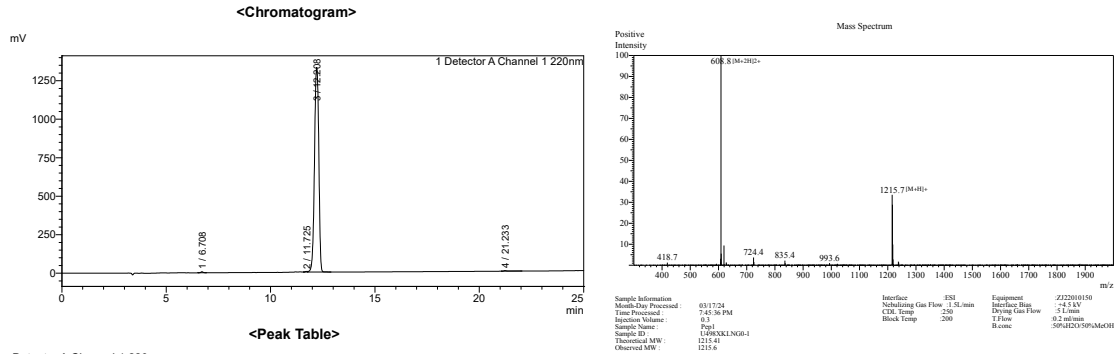
<Peak Table>

Detector A Channel 1 220nm

Peak#	Ret. Time	Area	Height	Area%
1	11.304	101802	11313	0.520
2	11.452	398124	34761	2.032
3	11.673	333163	56181	1.700
4	11.838	17701365	2638879	90.333
5	11.971	704411	167307	3.595
6	12.277	303880	34392	1.551
7	12.465	32579	6110	0.166
8	12.574	20423	5005	0.104
Total		19595747	2953948	100.000

Figure A2.: Evaluation of the purity of peptide Dri-m01. The purity of Dri-m01 (rfGsqlva) was evaluated by GenScript using (a) reverse-phase high-performance liquid chromatography (RP-HPLC) and (b) electrospray ionization mass spectrometry (ESI-MS) analyses.

(a) **(b)**



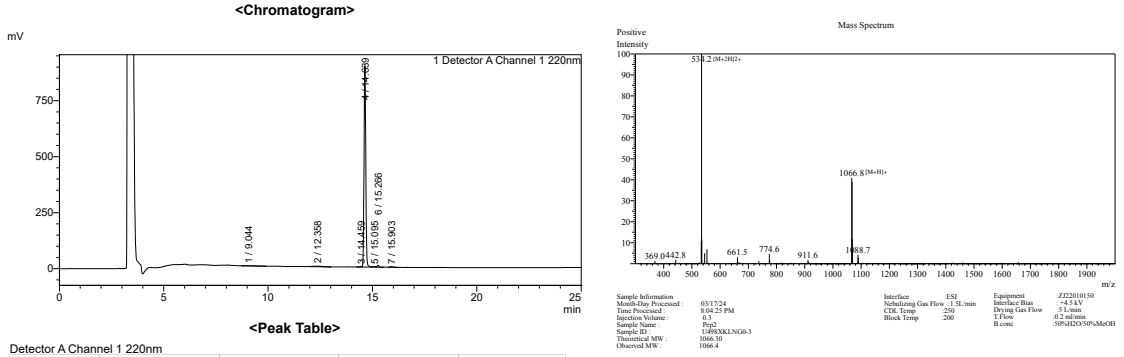
<Peak Table>

Detector A Channel 1 220nm

Peak#	Ret. Time	Area	Height	Area%
1	6.708	46546	6768	0.243
2	11.725	32524	4411	0.170
3	12.208	19024841	1328908	99.360
4	21.233	43378	3960	0.227
Total		19147289	1344047	100.000

Figure A3.: Evaluation of the purity of peptide a05. The purity of a05 (PRWQPKWY) was evaluated by GenScript using (a) RP-HPLC and (b) ESI-MS analyses.

(a) **(b)**



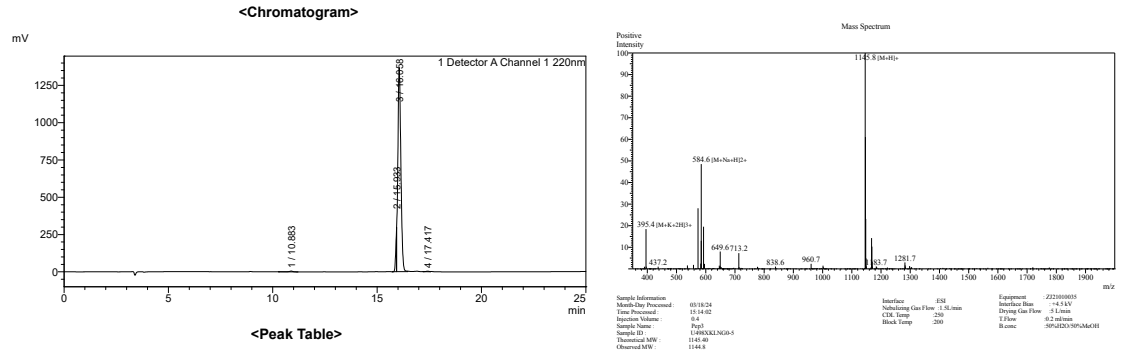
<Peak Table>

Detector A Channel 1 220nm

Peak#	Ret. Time	Area	Height	Area%
1	9.044	29811	776	0.544
2	12.358	83283	2655	1.521
3	14.459	19785	3154	0.361
4	14.639	5239148	896055	95.669
5	15.095	27809	3063	0.508
6	15.266	49282	8219	0.900
7	15.903	27231	2912	0.497
Total		5476349	916834	100.000

Figure A4.: Evaluation of the purity of peptide a06. The purity of a06 (IHLQVRVF) was evaluated by GenScript using (a) RP-HPLC and (b) ESI-MS analyses.

(a) **(b)**



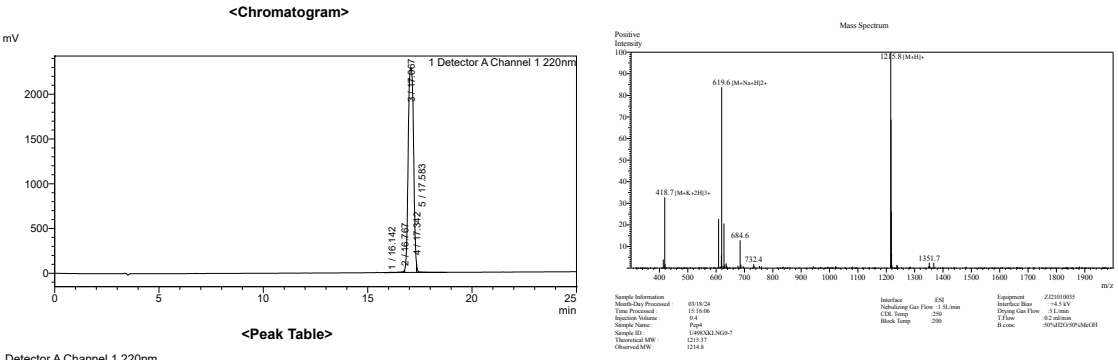
<Peak Table>

Detector A Channel 1 220nm

Peak#	Ret. Time	Area	Height	Area%
1	10.883	62963	5140	0.426
2	15.933	779888	297140	5.281
3	16.058	13886712	1367209	94.030
4	17.417	38847	3991	0.263
Total		14768411	1673479	100.000

Figure A5.: Evaluation of the purity of peptide a10. The purity of a10 (LRWQIVYI) was evaluated by GenScript using (a) RP-HPLC and (b) ESI-MS analyses.

(a) **(b)**



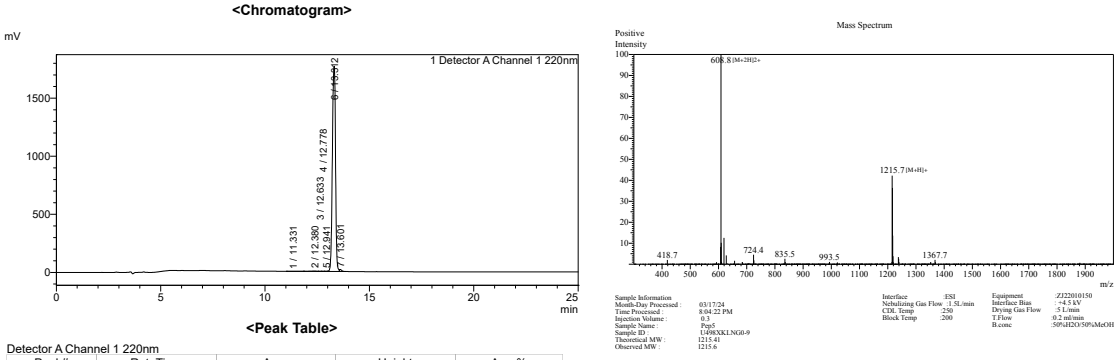
<Peak Table>

Detector A Channel 1 220nm

Peak#	Ret. Time	Area	Height	Area%
1	16.142	82165	4419	0.200
2	16.767	113112	25023	0.276
3	17.067	40549136	2287432	98.806
4	17.342	230110	91310	0.561
5	17.583	64605	2859	0.157
Total		41039128	2411042	100.000

Figure A6.: Evaluation of the purity of peptide a21. The purity of a21 (PRWQPQWY) was evaluated by GenScript using (a) RP-HPLC and (b) ESI-MS analyses.

(a) **(b)**



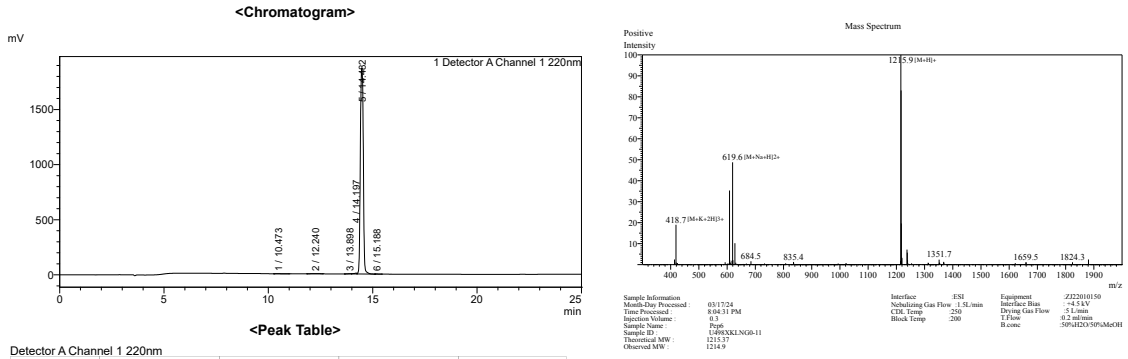
<Peak Table>

Detector A Channel 1 220nm

Peak#	Ret. Time	Area	Height	Area%
1	11.331	18565	664	0.099
2	12.380	71312	2619	0.379
3	12.633	21546	2389	0.115
4	12.778	17601	2327	0.094
5	12.941	17227	2419	0.092
6	13.312	18551477	1766292	98.592
7	13.601	118716	16680	0.631
Total		18816444	1793389	100.000

Figure A7.: Evaluation of the purity of peptide b06. The purity of b06 (PRWQpKWY) was evaluated by GenScript using (a) RP-HPLC and (b) ESI-MS analyses.

(a) **(b)**



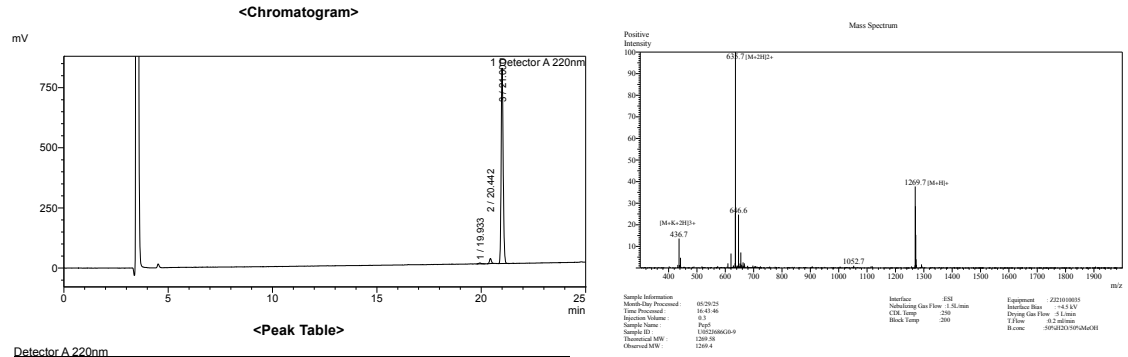
<Peak Table>

Detector A Channel 1 220nm

Peak#	Ret. Time	Area	Height	Area%
1	10.473	12884	485	0.069
2	12.240	50135	1861	0.269
3	13.898	43030	3404	0.231
4	14.197	80763	8031	0.434
5	14.482	18421409	1867930	98.953
6	15.188	8039	1389	0.043
Total		18616260	1883100	100.000

Figure A8.: Evaluation of the purity of peptide b07. The purity of b07 (PRWQpQWY) was evaluated by GenScript using (a) RP-HPLC and (b) ESI-MS analyses.

(a) **(b)**



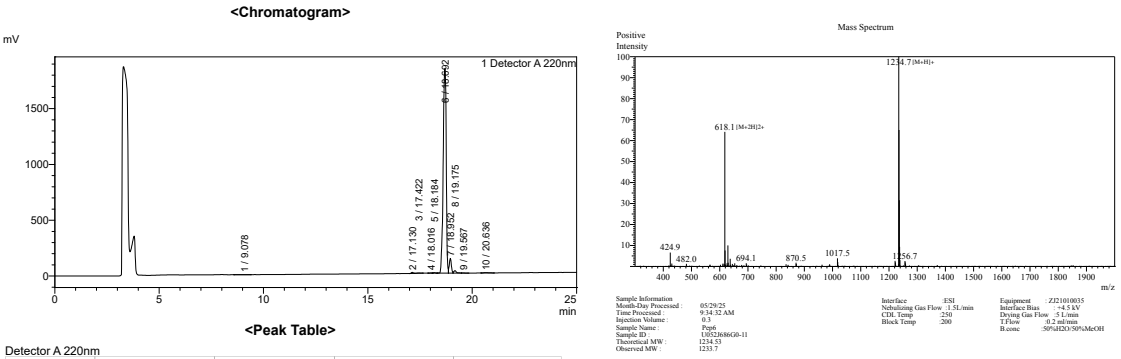
<Peak Table>

Detector A 220nm

Peak#	Ret. Time	Area	Height	Area%
1	19.933	29352	3776	0.521
2	20.442	134789	21421	2.391
3	21.000	5473494	813394	97.088
Total		5637634	838591	100.000

Figure A9.: Evaluation of the purity of peptide c04. The purity of c04 (MRLYIWFW) was evaluated by GenScript using (a) RP-HPLC and (b) ESI-MS analyses.

(a) **(b)**



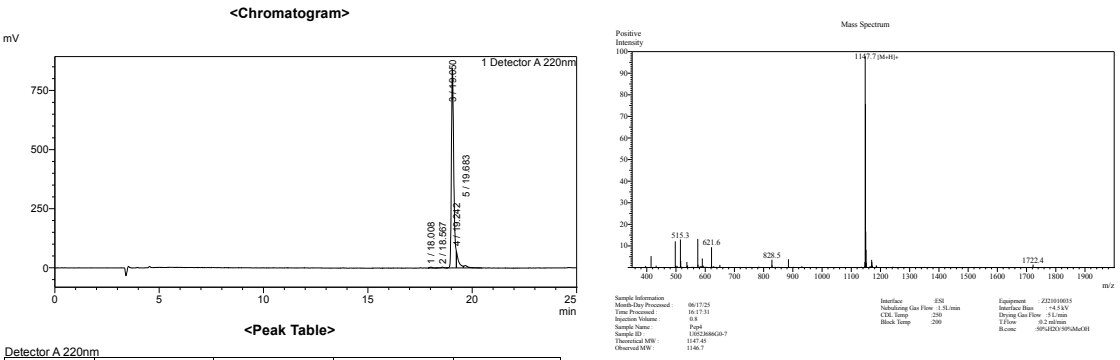
<Peak Table>

Detector A 220nm

Peak#	Ret. Time	Area	Height	Area%
1	9.078	23217	674	0.117
2	17.130	66006	9724	0.332
3	17.422	31880	2129	0.160
4	18.016	8591	1318	0.043
5	18.184	28718	4600	0.144
6	18.692	18553125	1833859	93.324
7	18.952	921899	131345	4.637
8	19.175	193065	22489	0.971
9	19.567	8868	940	0.045
10	20.636	44931	2761	0.226
Total		19880300	2009839	100.000

Figure A10.: Evaluation of the purity of peptide c05. The purity of c05 (MRLQIWFW) was evaluated by GenScript using (a) RP-HPLC and (b) ESI-MS analyses.

(a) **(b)**



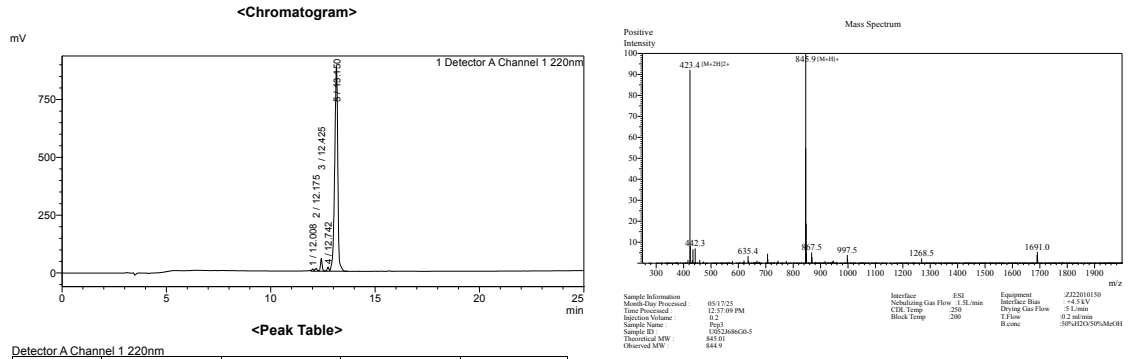
<Peak Table>

Detector A 220nm

Peak#	Ret. Time	Area	Height	Area%
1	18.008	32708	3641	0.444
2	18.567	32374	3771	0.439
3	19.050	6653572	844947	90.286
4	19.242	498511	70592	6.765
5	19.683	152304	10008	2.067
Total		7369468	932959	100.000

Figure A11.: Evaluation of the purity of peptide c07. The purity of c07 (MRLQIVFW) was evaluated by GenScript using (a) RP-HPLC and (b) ESI-MS analyses.

(a) **(b)**



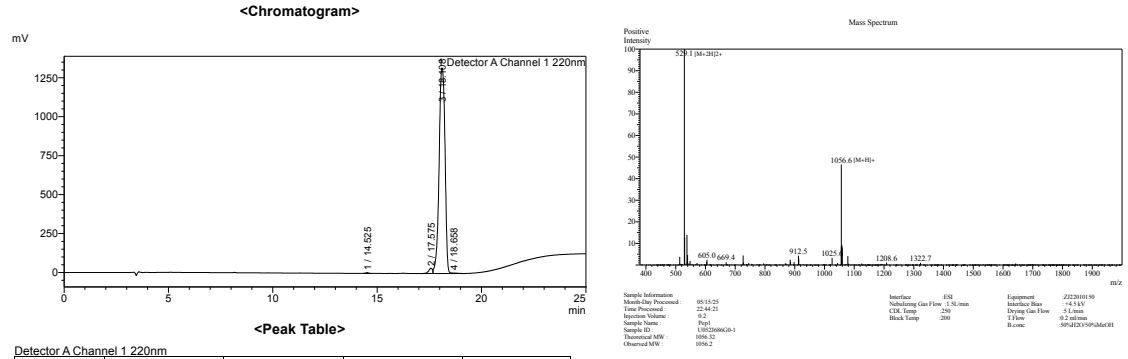
<Peak Table>

Detector A Channel 1 220nm

Peak#	Ret. Time	Area	Height	Area%
1	12.008	61047	8789	0.630
2	12.175	87138	11333	0.900
3	12.425	312387	54415	3.225
4	12.742	100629	17612	1.039
5	13.150	9124036	880387	94.206
Total		9685238	972538	100.000

Figure A12.: Evaluation of the purity of peptide Dri-p02. The purity of Dri-p02 (paGGkltf) was evaluated by GenScript using (a) RP-HPLC and (b) ESI-MS analyses.

(a) **(b)**



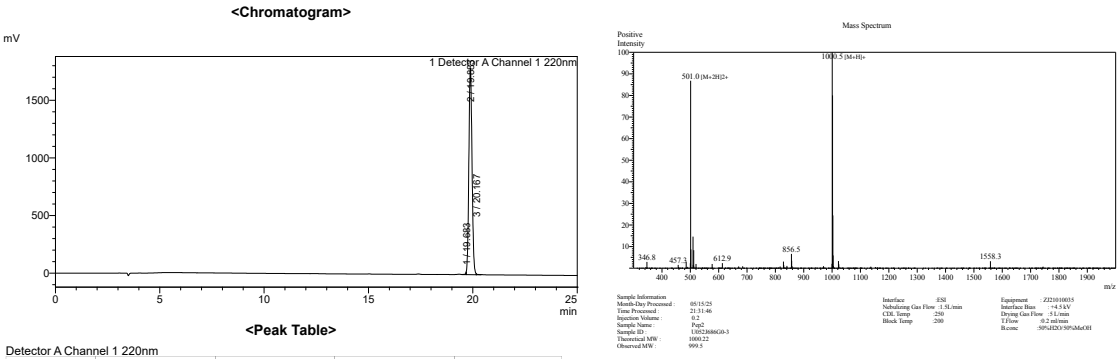
<Peak Table>

Detector A Channel 1 220nm

Peak#	Ret. Time	Area	Height	Area%
1	14.525	34271	4925	0.130
2	17.575	425526	32806	1.610
3	18.108	25938475	1319809	98.121
4	18.658	37028	2794	0.140
Total		26435300	1360333	100.000

Figure A13.: Evaluation of the purity of peptide z07. The purity of z07 (PFRLIGWL) was evaluated by GenScript using (a) RP-HPLC and (b) ESI-MS analyses.

(a) **(b)**



Detector A Channel 1 220nm

Peak#	Ret. Time	Area	Height	Area%
1	19.683	54401	22676	0.284
2	19.883	19046807	1792597	99.535
3	20.167	34613	8180	0.181
Total		19135820	1823453	100.000

Figure A14.: Evaluation of the purity of peptide z08. The purity of z08 (PFRLGGWL) was evaluated by GenScript using (a) RP-HPLC and (b) ESI-MS analyses.

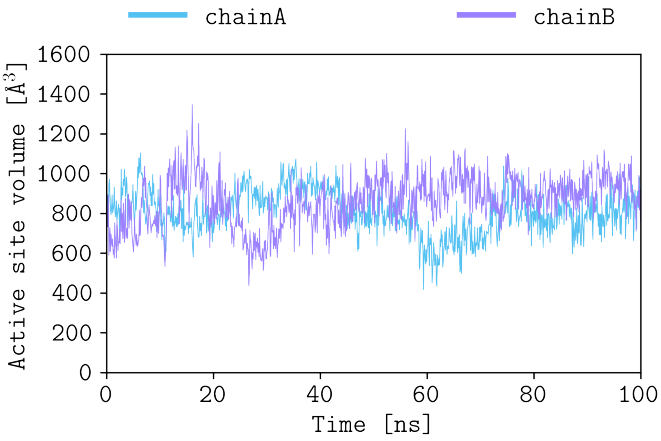


Figure A15.: Volume of the active sites in the unbound M^{pro} dimer MD simulation. The active site volume of both protomers—chain A and chain B—of the M^{pro} dimer (PDB: 6M2N) was measured throughout the 200 ns simulation using the VMD plugin Epock [284].

	run1: binding affinity (kcal/mol)						run2: distance (Å)						run3: binding affinity/distance (norm)					
	L-2	Dri-2	L-7	Dri-7	L-4	Dri-4	L-2	Dri-2	L-7	Dri-7	L-4	Dri-4	L-2	Dri-2	L-7	Dri-7	L-4	Dri-4
cl01	-7.3	-6.3	-6.4	-6.0	-8.3	-8.4	2.3	2.4	3.1	2.3	5.5	7.0	0.46	0.76	0.86	0.87	0.26	● 0.38
cl02	-6.2	-6.7	-6.7	-7.1	-9.2	-7.7	2.6	2.5	2.2	● 1.4	9.5	7.6	0.87	0.73	0.48	0.21	0.59	0.61
cl03	● -7.9	-6.7	-7.6	● -7.9	-7.1	-7.5	2.3	2.2	2.3	2.0	● 5.7	6.7	0.33	0.63	0.26	0.26	0.50	0.58
cl04	-5.9	-8.2	-7.5	-6.6	-7.8	● -9.2	2.7	2.5	2.1	1.8	7.7	8.0	0.97	0.47	● 0.20	0.52	0.63	0.27
cl05	-7.0	-6.7	-7.4	-6.7	-7.6	-7.8	2.6	● 1.5	2.6	2.6	9.4	10.8	0.67	0.30	0.40	0.80	0.87	0.85
cl06	-6.1	-7.7	-6.9	-7.5	-8.0	-8.5	2.3	1.5	2.2	2.2	7.9	7.6	0.75	● 0.13	0.43	0.43	0.61	0.41
cl07	-6.5	-7.1	-5.9	-6.5	-7.0	-8.4	2.1	2.6	2.5	2.1	8.4	8.1	0.60	0.72	0.81	0.68	0.86	0.47
cl08	-6.7	-7.4	● -7.7	-6.1	● -9.7	-7.2	2.7	2.5	2.0	2.3	8.5	10.1	0.80	0.60	0.14	0.85	0.37	0.94
cl09	-7.9	● -8.3	-7.1	-7.6	-8.0	-8.1	1.5	1.7	● 1.6	1.9	6.9	10.3	● 0.00	0.09	0.17	● 0.29	0.48	0.73
cl10	-7.4	-5.6	-7.4	-6.2	-9.3	-7.8	● 1.8	2.5	2.3	2.6	8.5	● 4.8	0.25	0.93	0.30	0.95	● 0.44	0.35

Figure A16.: Docking solutions of the peptide structures (cl01–10) and selection criteria for starting structures of MD runs 1–3 in complex with M^{pro}. The 3CVL (L) and 3CVLri (Dri) inhibitors were docked to M^{pro} using peptide conformations derived from cluster analysis of unbound peptide MD trajectories (clusters 01–10). In the selection process of starting structures for subsequent MD simulations (marked by a dot), the docking solutions for each peptide were ranked based on binding affinity (run 1), minimum distance to the binding site (run 2), and the normalized average of both criteria (run 3).

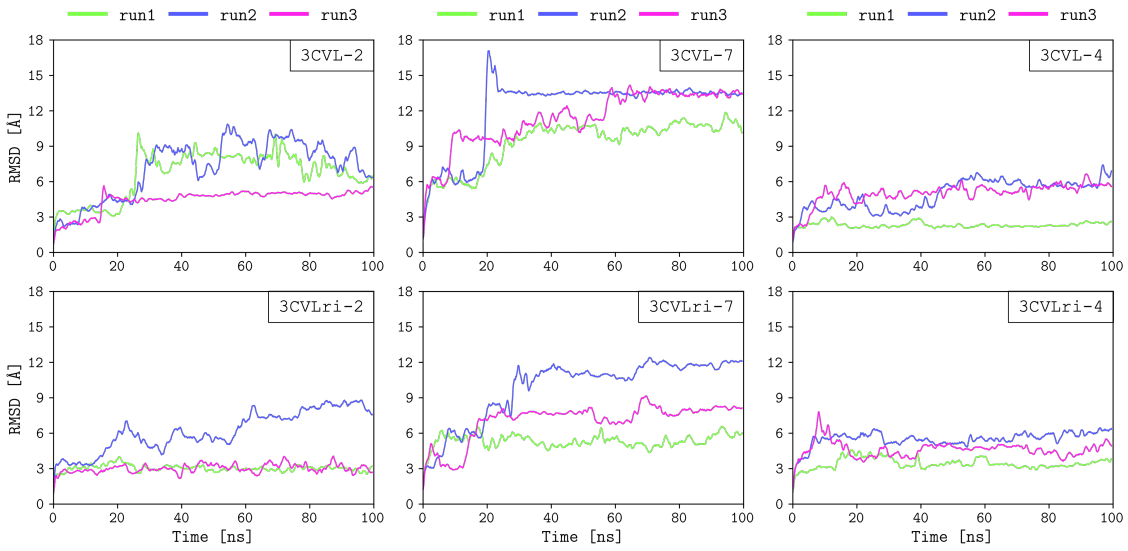


Figure A17.: Conformational changes of 3CVL(ri) peptides in complex with M^{pro} during different MD runs. The root mean square deviation (RMSD) of the 3CVL(ri) peptides (backbone) in complex with M^{pro} with reference to the starting structure was measured throughout the 100 ns simulation for run1-3.

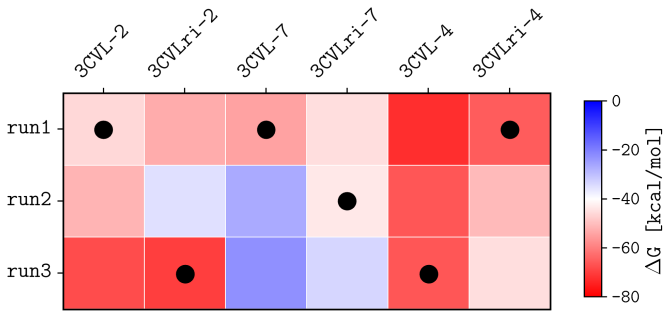


Figure A18.: Binding free energy estimation of the M^{pro}-3CVL(ri) complex MD simulations The binding free energy (ΔG_{bind}) was determined between the 3CVL(ri) peptides and M^{pro} for each run (1-3) using gmx_mmpbsa[59]. Dots indicate runs that were selected for extended triplicate simulations.

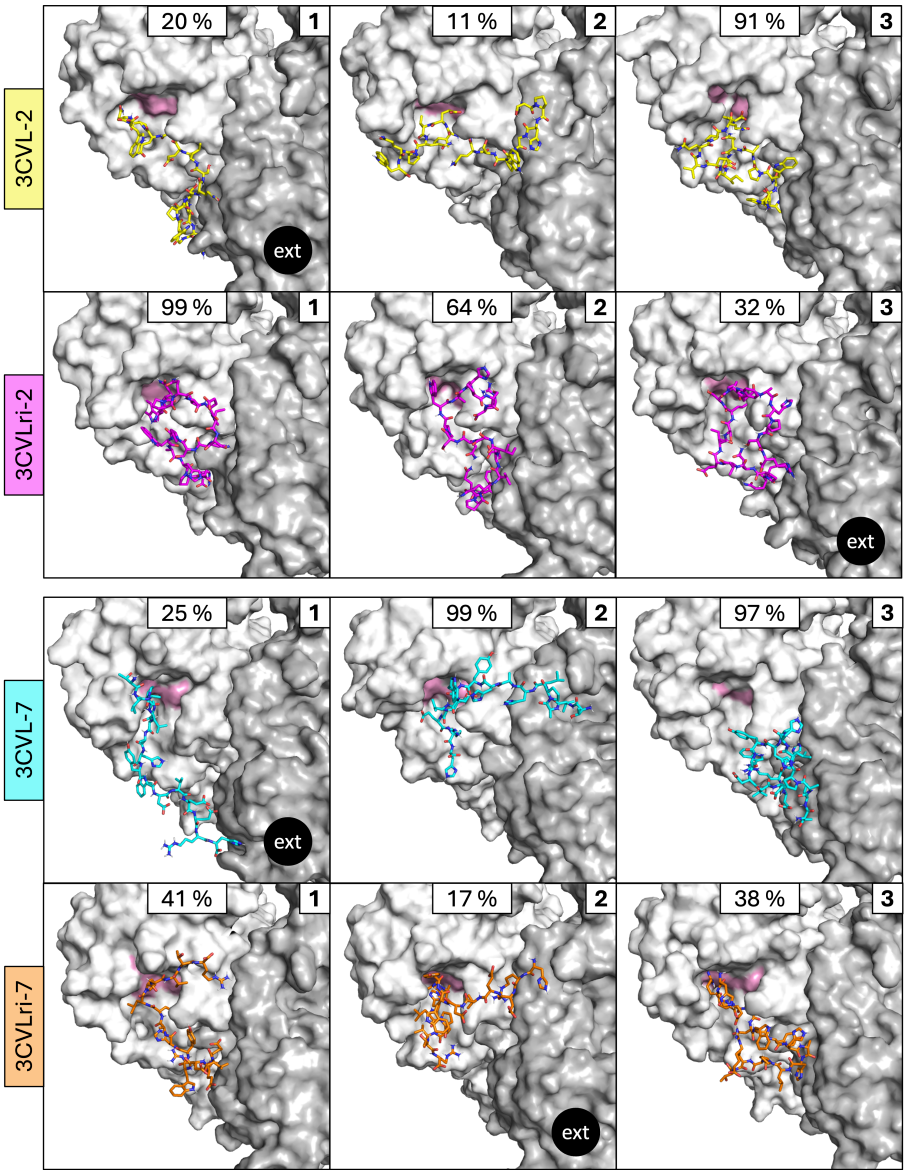


Figure A19.: Binding modes from different MD runs of M^{pro} in complex with competitive peptide inhibitors The trajectories of M^{pro} in complex with 3CVL(ri)-2 and 3CVL(ri)-7 were subjected to cluster analysis. The most populated peptide conformations (sampling in %) within the M^{pro} binding site are shown for each run (1-3), starting from different docking poses. The surface of chain A (white) and chain B (grey) of M^{pro} is shown, with the peptide structures represented as sticks. The catalytic dyad (His41, Cys145) is marked in pink. The black circles (ext) indicate runs that were selected for extended triplicate simulations. Adapted from Eberle et al.[65]

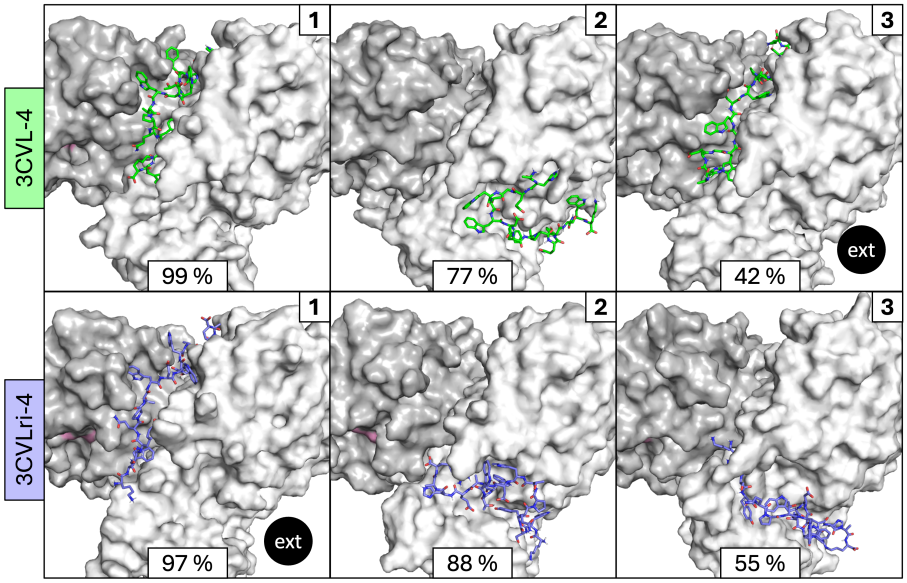


Figure A20.: Binding modes from different MD runs of M^{pro} in complex with non-competitive peptide inhibitors The trajectories of M^{pro} in complex with 3CVL-4 and 3CVLri-4 were subjected to cluster analysis. The most populated peptide conformations (sampling in %) within the M^{pro} binding site are shown for each run (1-3), starting from different docking poses. The surface of chain A (white) and chain B (grey) of M^{pro} is shown, with the peptide structures represented as sticks. The black circles (ext) indicate runs that were selected for extended triplicate simulations. Adapted from Eberle et al.[65]

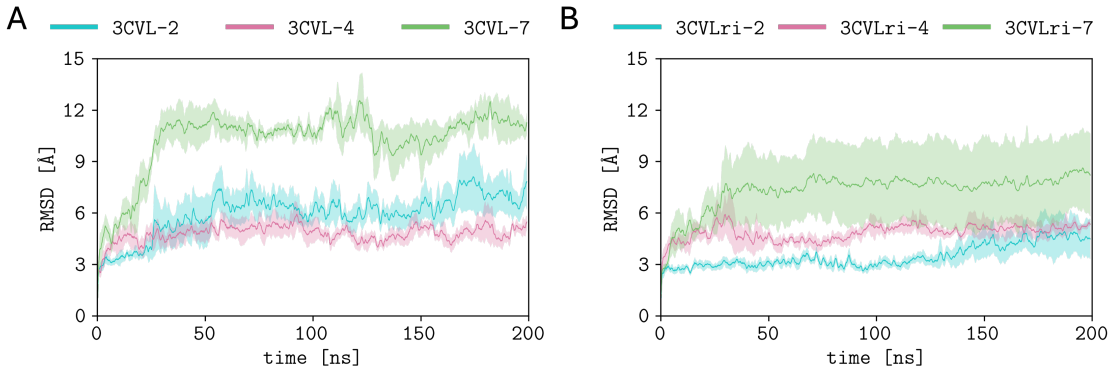


Figure A21.: Conformational changes of 3CVL(ri) peptides in complex with M^{pro} during extensive MD simulations. The RMSD of the 3CVL(ri) peptides (backbone) in complex with M^{pro} with reference to the starting structure is shown as average (±SEM) of the triplicate 200 ns simulations.

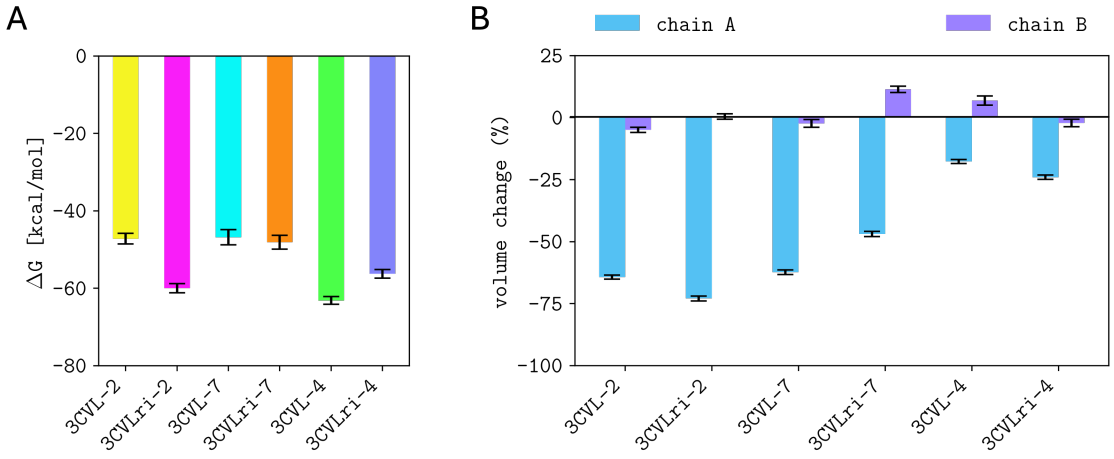


Figure A22.: Total binding free energy and change in active site volume of M^{pro}-peptide complex simulations. The triplicate 200 ns MD simulations of M^{pro} in complex with 3CVL(ri)-2, 3CVL(ri)-4, and 3CVL(ri)-7 were analyzed as follows: (A) The total binding free energy was calculated for each M^{pro}-peptide complex and is shown as the average (±SEM) of all replicas. (B) The barplot shows the change of the active site volume of both M^{pro} protomers (chain A and chain B) upon inhibitor binding as the deviation of the averaged volume of each replica run from that of the unbound (apo) M^{pro} simulation in percent ((Vol$_{complex}$–Vol$_{apo}$)/Vol$_{apo}$). For the competitive inhibitors (3CVL(ri)-2 and 3CVL(ri)-7), the chain A bar includes the volume displaced by the peptide to compare their occupation potential.

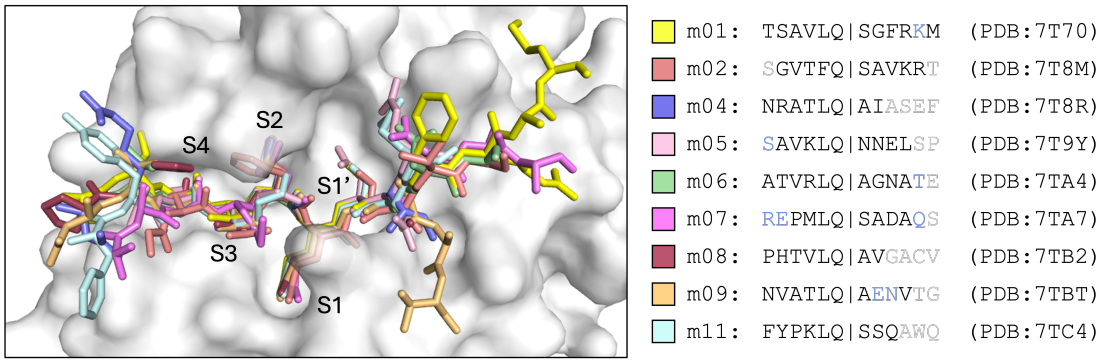


Figure A23.: Co-crystallized M^{pro} substrate peptides. Co-crystal structures of peptides corresponding to the non-structural cleavage sites nsp4/5–nsp15/16 (m01–m11) bound to the active site of M^{pro}, as reported by Shaqra et al.[290]. Superimposed peptide structures are displayed as sticks in distinct colors. The M^{pro} structure is shown as a surface representation, with binding subsites S4–S1′ annotated. The one-letter amino acid sequence and PDB ID of each substrate are listed on the right. Blue and grey letters indicate unresolved side chains or fully unresolved residues, respectively. The missing m03 (nsp6/7) and m10 (nsp14/15) structures could not be resolved.

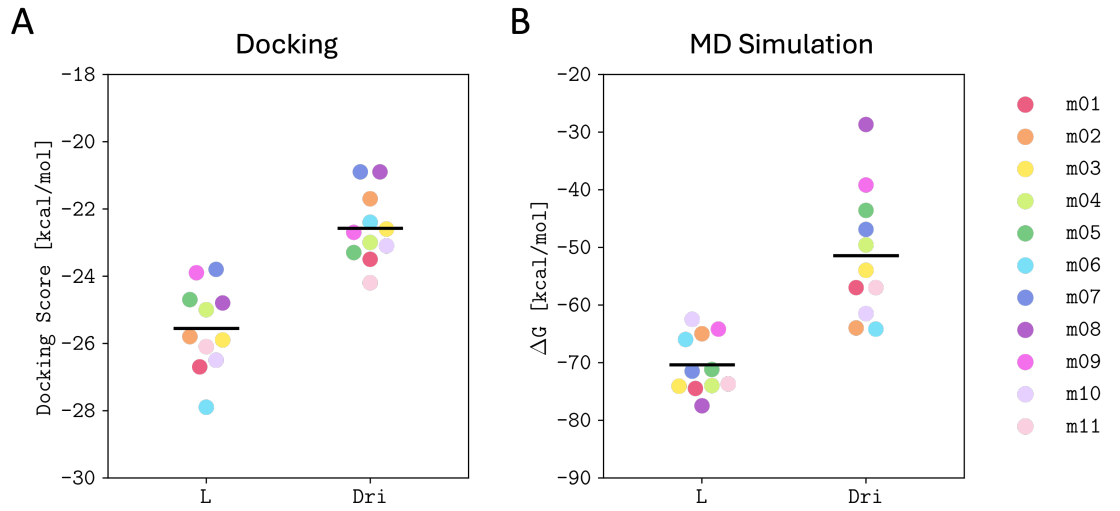


Figure A24.: Binding affinities of L- and Dri-peptides m01–m11 with M^{pro} estimated by Docking and MD simulations. (A) The L- and Dri-peptides (m01–m11) were docked against the active site of M^{pro} using ADCP and the re-scored docking scores are shown for each individual peptide. (B) The M^{pro}-peptide models derived from docking were simulated for 100 ns and ΔG_{bind} was computed using the molecular mechanics Poisson–Boltzmann surface area (MM/PBSA) method. The horizontal black lines indicate the mean of all 11 substrate.

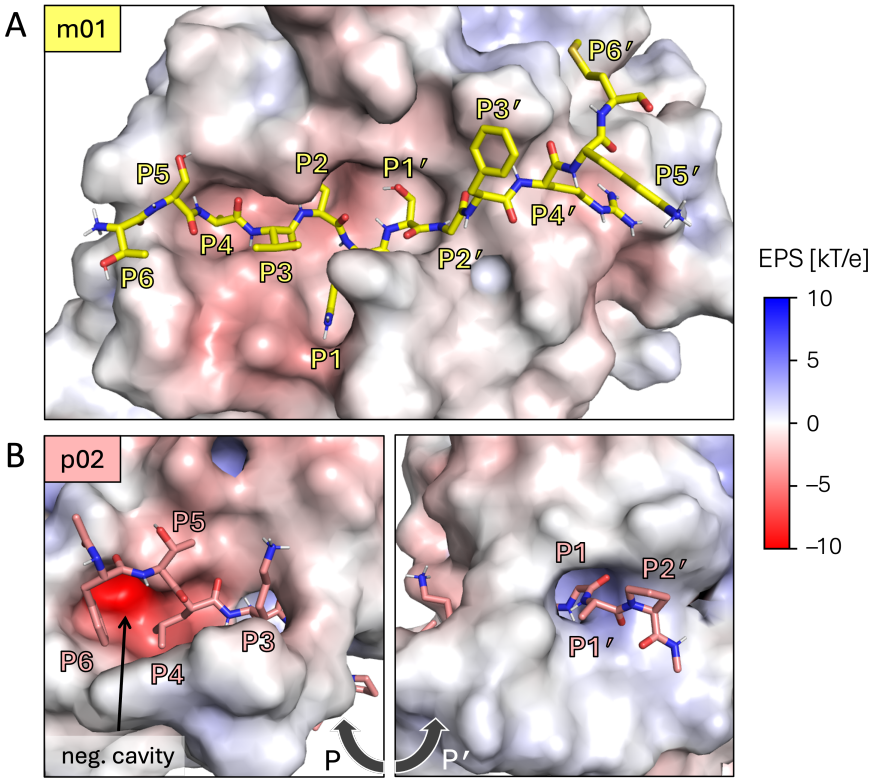


Figure A25.: Electrostatic potential surface of M^{pro} and PL^{pro}. The electrostatic potential surface (EPS) was computed for (A) the crystal structure of M^{pro} bound to nsp4/5 (m01) and (B) the AlphaFold 3 (AF3)-predicted PL^{pro} structure in complex with p02 (8 amino acids). Red and blue indicate negatively and positively charged areas on the protease surface, respectively. The peptide structures are displayed as sticks, with peptide positions (P and P′) marked. The arrow indicates the negatively charged cavity within the PL^{pro} binding site.

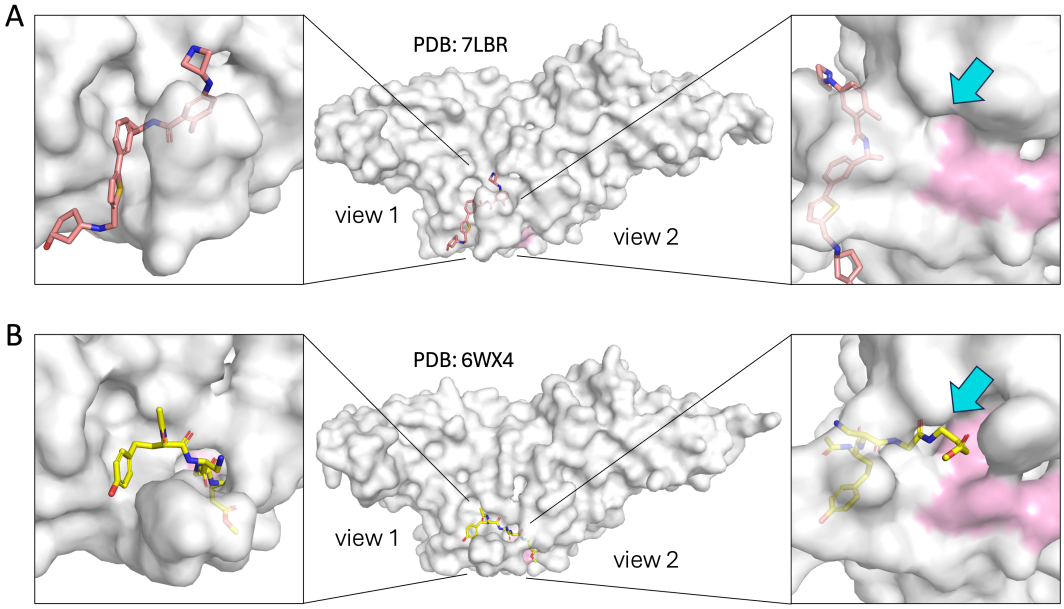


Figure A26.: Comparison of two PL^{pro} crystal structures and their suitability as receptors for peptide docking. PL^{pro} in complex with inhibitors (A) XR8-89 (PDB: 7LBR) and (B) VIR251 (PDB: 6WX4). PL^{pro} is shown as a white surface, with the catalytic site (Cys111, His272, and Asp286) highlighted in pink. Inhibitors are shown as sticks. Cyan arrows indicate the tunnel subsite that usually accommodates the GG part of the LXGG substrate binding motif.

A **ADCP Docking:** Affinity [kcal/mol]

	6aa	8aa	10aa	12aa		6aa	8aa	10aa	12aa
L–p01	–16.3	–20.3	–21.9	–25.3	**Dri–p01**	–14.6	**–19.7**	–22.1	-22.9
L–p02	–16.8	–18.8	–21.3	–21.5	**Dri–p02**	–15.5	**–19.1**	–19.8	-21.7
L–p03	–17.6	–19.6	–23.3	–25.7	**Dri–p03**	–16.4	**–18.7**	–21.2	-23.3

B **AlphaFold3:** ipTM

	6aa	8aa	10aa	12aa
L–p01	0.71	0.62	0.47	0.42
L–p02	0.95	**0.95**	0.91	0.86
L–p03	0.95	0.89	0.59	0.35

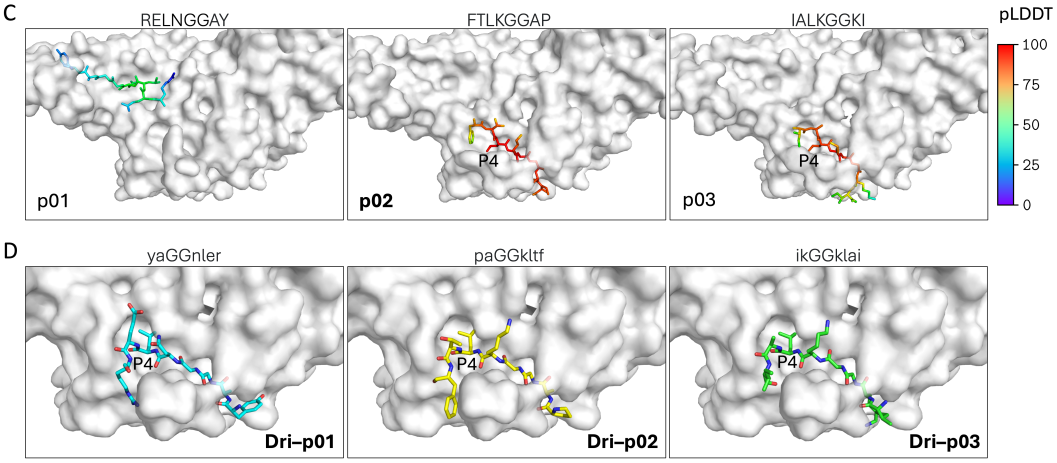


Figure A27.: Solutions for PL^{pro}–peptide complex generation with L- and Dri-peptides via a consensus docking approach. (A) Affinities (kcal/mol) of re-scored docking solutions for L- and Dri-peptides of different lengths docked against PL^{pro} using ADCP. (B) interface predicted Template Modeling score (ipTM) values indicating the confidence of AF3 complex predictions for PL^{pro} with L-substrates p01–p03 at different peptide lengths. (C) AF3-predicted PL^{pro}–peptide complexes (8-residue L-peptides p01–p03) with the protease surface in white and peptides as sticks, color-coded by predicted Local Distance Difference Test (pLDDT) from blue (low confidence) to red (high confidence). (D) Re-scored docking poses of Dri-peptides with the protease surface in white and peptides as sticks. Amino acid sequences of the peptides are shown above each complex representation.

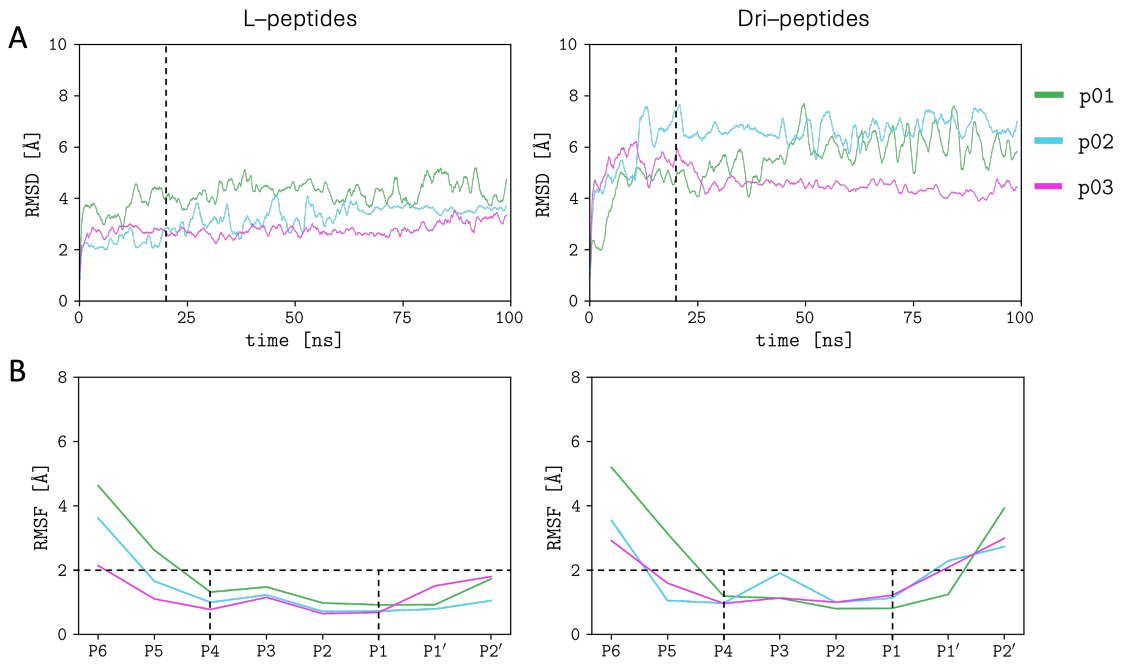


Figure A28.: Stability and flexibility of L- and Dri-peptides in complex with PL^{pro} during MD simulations. The (A) RMSD and (B) root mean square fluctuation (RMSF) was computed for the heavy atoms of the L- and Dri- substrate peptides (p01–p03), simulated for 100 ns in complex with PL^{pro}. (A) The vertical dashed line indicates the time (20 ns), after which the peptide conformations exhibit mostly stable RMSD values. (B) The RMSF is displayed for the peptide positions P6–P2′ with the horizontal dashed line at 2 Å and the vertical dashed lines marking the binding motif LXGG (P4–P1).

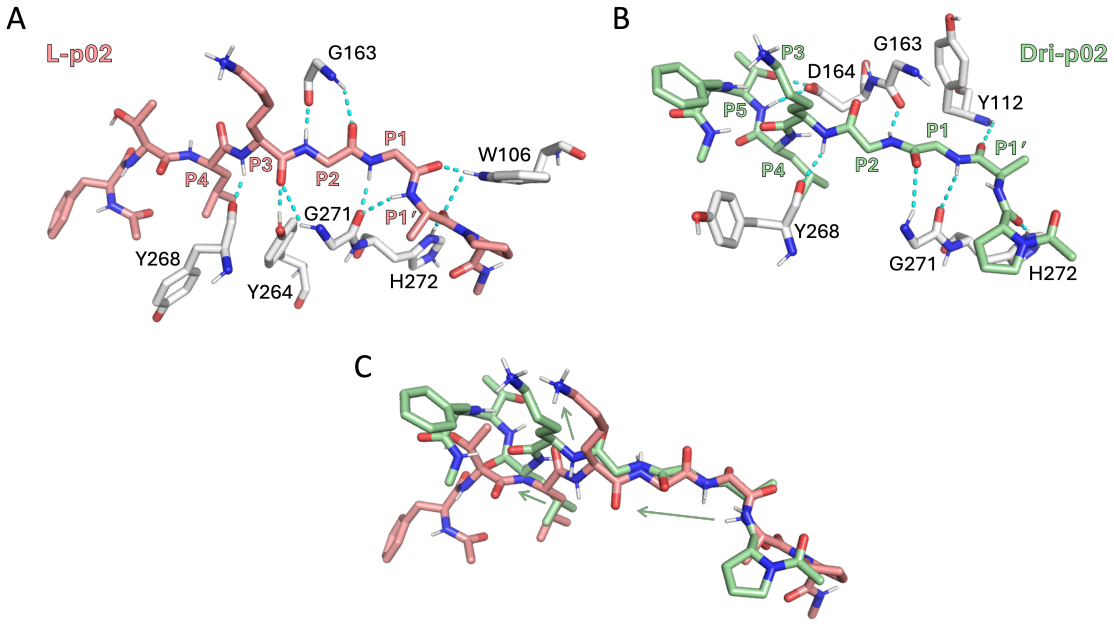


Figure A29.: Substrate peptides p01 and p03 forming hydrogen bonds with PL^{pro}. The structure of the most populated conformation of the simulations with (A) L-p01 (46 % sampled) and (B) L-p02 (99 % sampled) are shown in complex with PL^{pro}. The peptide structures are displayed as sticks in (A) purple and (B) pink. PL^{pro} residues forming hydrogen bonds with the peptides are shown as white sticks, with hydrogen bonds displayed as cyan dotted lines.

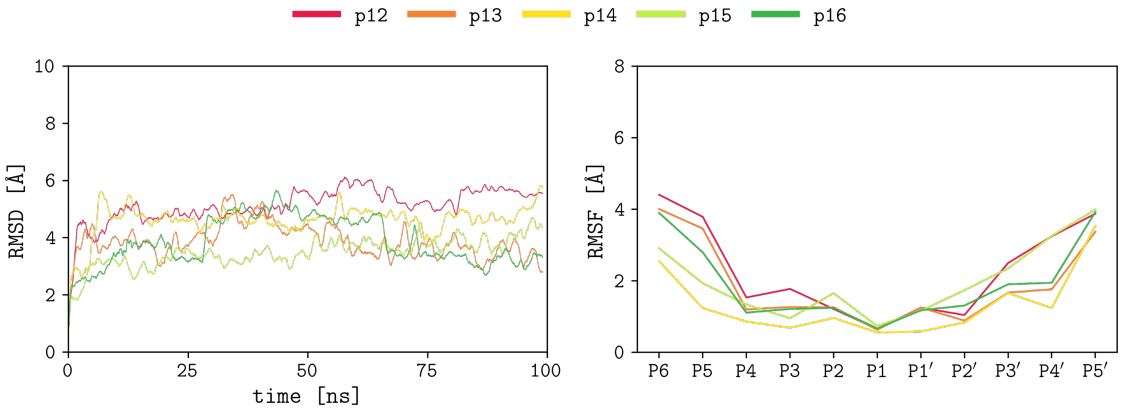


Figure A30.: Stability and flexibility of peptide inhibitors p12–p16 in complex with M^{pro} during MD simulations. The RMSD and RMSF was computed for the heavy atoms of the peptides p12–p16, simulated for 100 ns in complex with M^{pro}. The per-residue RMSF is displayed for the peptide positions P6–P6′.

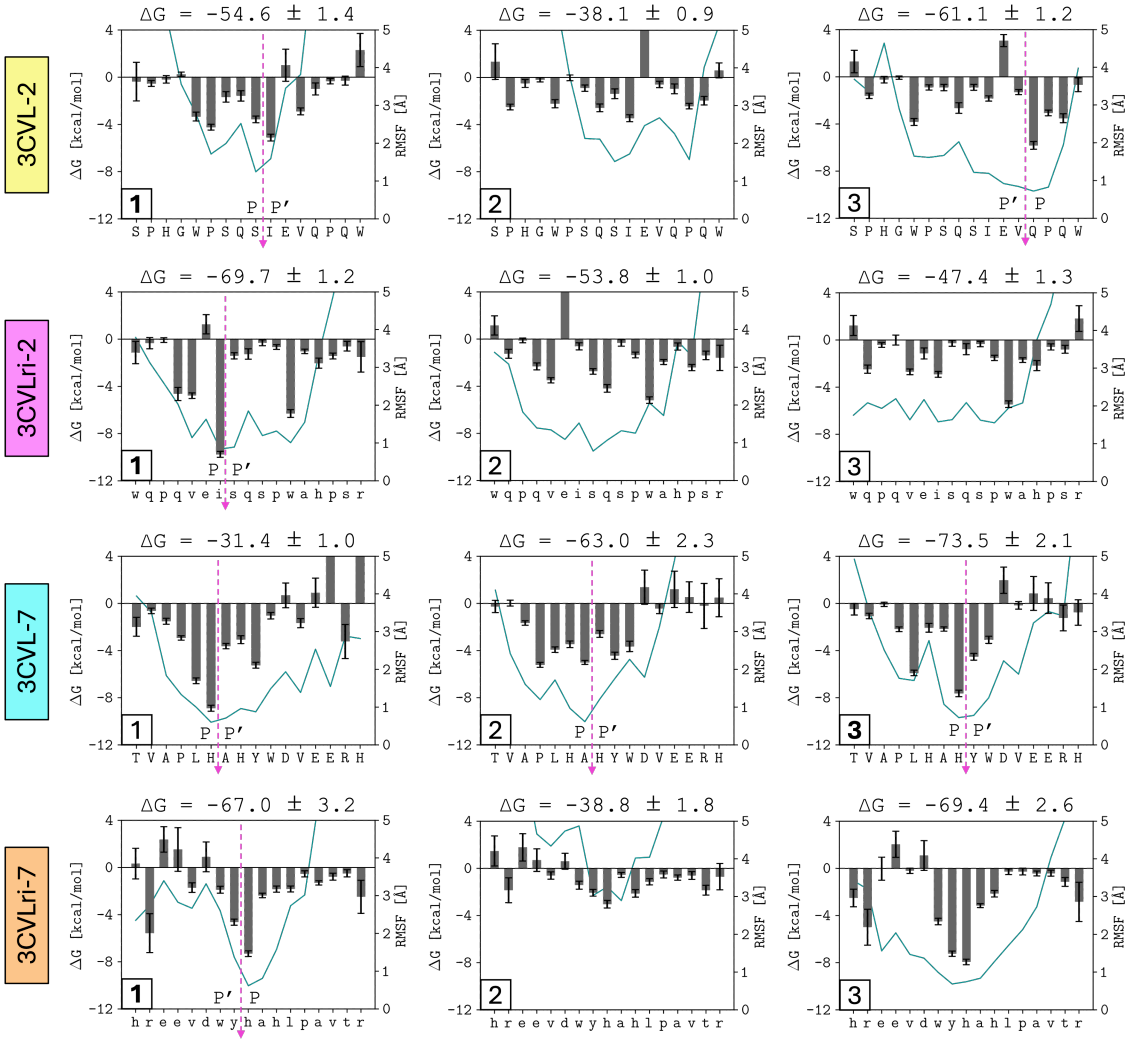


Figure A31.: Binding free energy decomposition and structural flexibility of the M^{pro} inhibitors 3CVL(ri). The per-residue decomposition of the binding free energy (ΔG) is shown for run1–3 (top to bottom) of the 3CVL(ri) inhibitors in complex with Mpro (grey bars). The RMSF of each residue is depicted as the blue line on the secondary axis. From left to right, the sequences of 3CVL and 3CVLri peptides are displayed from N- to C-terminus. The pink vertical array marks the peptide bond between P1/P1′ or P1′/P1 depending on the orientation of the peptide in the binding site. Above the plots, the total ΔG of the peptide is indicated.

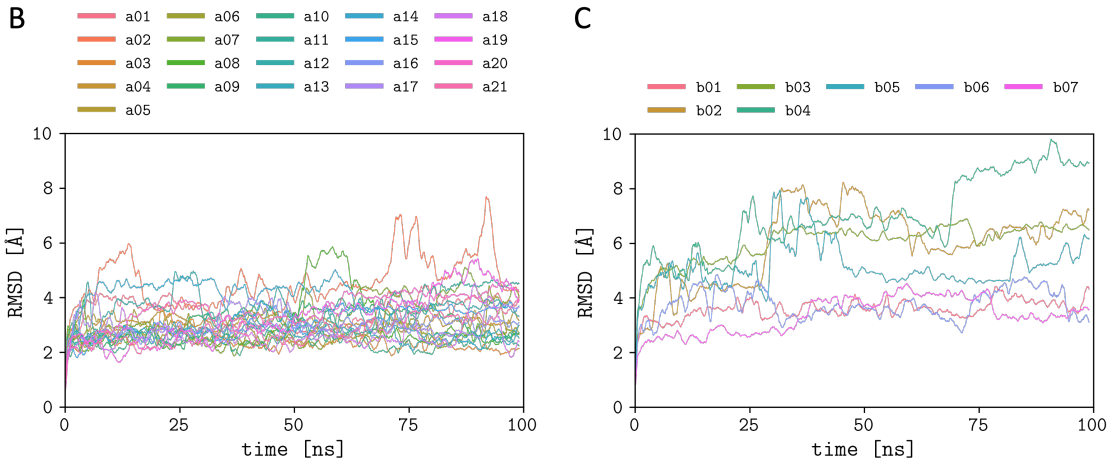


Figure A32.: Stability of the peptide designs a01–a21 and b01–b07 from strategy II in complex with M^{pro} during MD simulations. The RMSD was computed for the heavy atoms of (A) the a-designs and (B) b-designs that were simulated 100 ns in complex with M^{pro}.

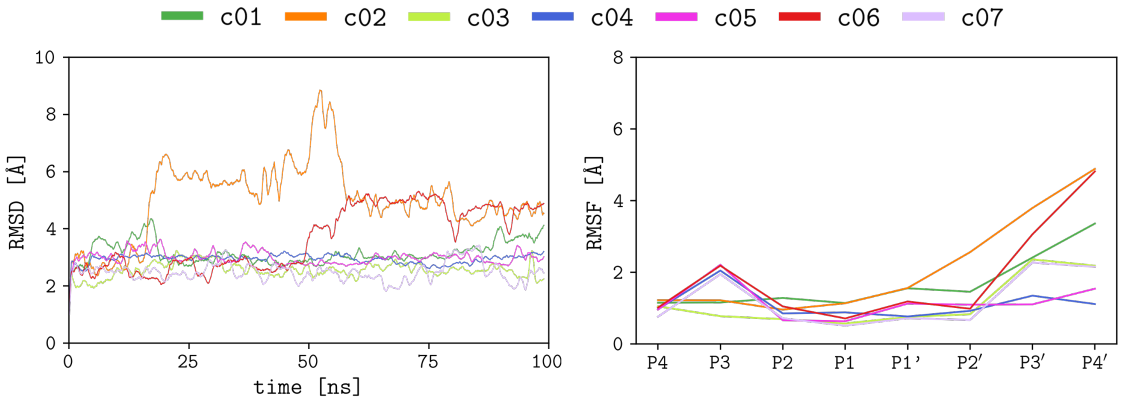


Figure A33.: Stability and flexibility of the peptide designs c01–c07 from strategy III in complex with M^{pro} during MD simulations. The RMSD and RMSF was computed for the heavy atoms of the peptide designs c01–c07, simulated for 100 ns in complex with M^{pro}. The per-residue RMSF is displayed for the peptide positions P4–P4′.

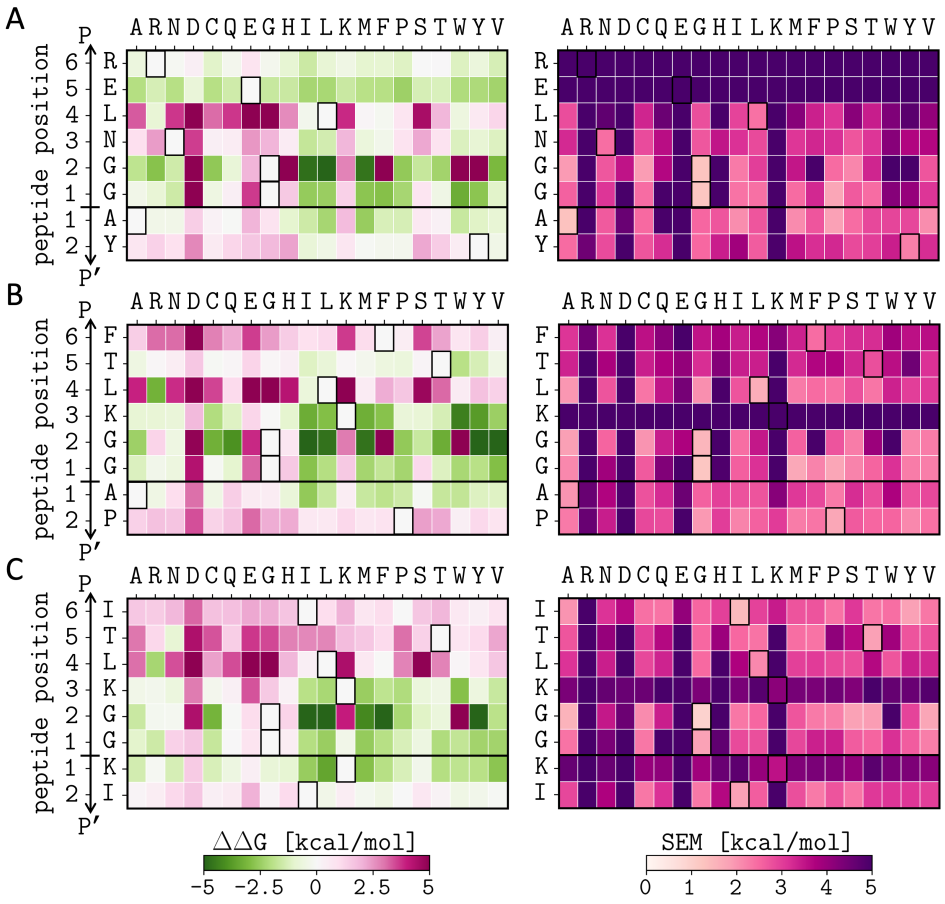


Figure A34.: CDSM with MD simulations of PL^{pro} in complex with p01–p03. The 100 ns MD simulations of PL^{pro} in complex with the substrate peptides (A) p01, (B) p02 and (C) p03 were used as input for CDSM using five snapshots each. The ΔΔG values for each peptide position of P6–P2′ and substitution are shown. The respective standard error of the mean (SEM) was determined across the five snapshot simulations. When the parent residue matches the mutated amino acid, the field is marked by a black frame.

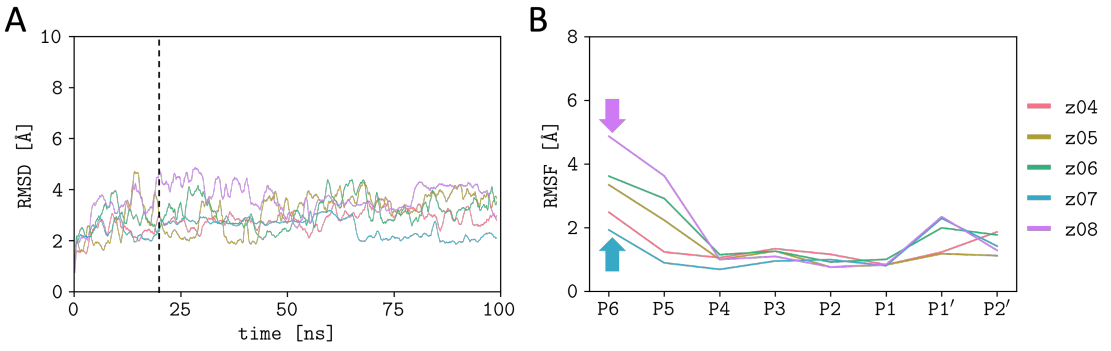


Figure A35.: Stability and flexibility of the peptide designs z04–z08 from strategy III in complex with PL^{pro} during MD simulations. The (A) RMSD and (B) RMSF was computed for the heavy atoms of the z04–z08 peptide designs that were simulated 100 ns in complex with PL^{pro}. The RMSF is displayed for the peptide positions P6–P2′. The arrows indicate low (blue) and high (violet) flexibility for the N-terminal P6–P5 positions in z07 and z08, respectively.

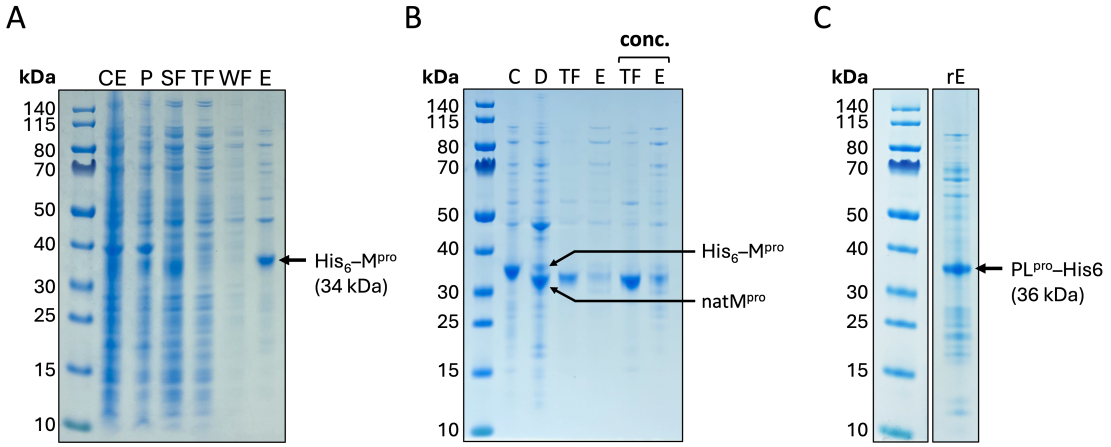


Figure A36.: SDS-PAGE with samples from the production steps of natural M^{pro}. (A) After expression in *E. coli* BL21(DE3) the His-tagged M^{pro} (His_6–M^{pro}) was purified by immobilized metal ion affinity chromatography (IMAC). The rows in the SDS gel from left to right contain CE: cell extract, SF: soluble fraction, TF: through-flow, WF: wash fraction, E: elution fraction. The purified His-tagged M^{pro} (His_6–M^{pro}) is marked in the elution fraction. (B) To obtain natural M^{pro} (natM^{pro}), the His_6–M^{pro} was digested with TEV protease. To separate natM^{pro} from the His-tagged M^{pro} and TEV protease, IMAC was applied. The rows in the SDS gel from left to right contain C: a control fraction without TEV protease, D: digestion fraction, TF: through-flow, and E: elution. The TF and E fractions were rebuffered and concentrated (conc.). The rebuffered TF fraction was concentrated to 2.5 mL with a final protease concentration of 0.59 mg/mL. (C) PL^{pro} was expressed in *E. coli* BL21(DE3) followed by purification via IMAC. The SDS gel shows the purified PL^{pro} in the rebuffered elution fraction (rE) with a final concentration of 0.54 mg/mL.

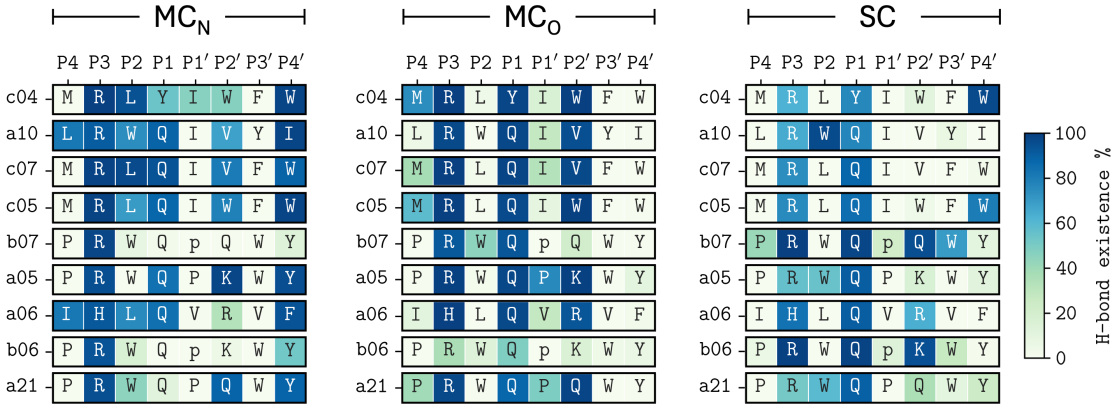


Figure A37.: Hydrogen bond existence for all M^{pro} peptide designs tested experimentally. The hydrogen bond (H-bond) existence between each peptide design and M^{pro} is shown for each peptide position (P4-P4′), and is further divided into mainchain (MC) hydrogen donor (nitrogen, N), hydrogen acceptor (oxygen, O) and variable sidechain (SC) hydrogen donors/acceptors. From left to right, the amino acid sequence of each peptide is displayed from N- to C-terminus. The existence is considered 100 % if at least one H-bond was detected for the respective peptide group in all analysed frames of the MD simulation.

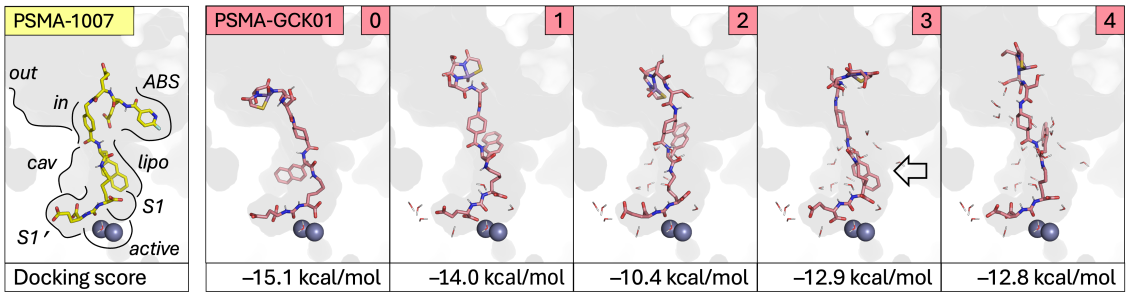


Figure A38.: Docking solutions of PSMA-GCK01 (2). The structure of PSMA-GCK01 was docked to PSMA (PDBs: 5O5T) using AutoDockFR (ADFR). Re-scored top-ranked docking poses are shown for hydration levels 0–4, with the docking scores provided below. For reference, the co-crystallized ligand structure of PSMA-1007 is displayed on the left including labeled binding subsites (Fig. 5.4).

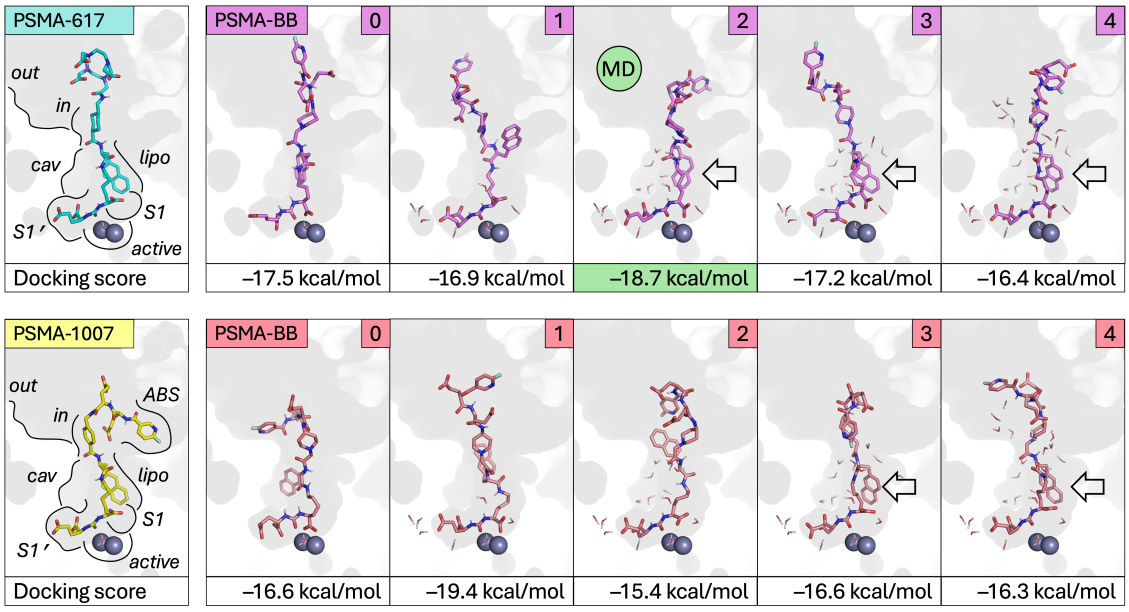


Figure A39.: Docking solutions of PSMA-BB. The ligand PSMA-BB was docked to both PSMA structures (PDBs: 8BOW and 5O5T) using ADFR. Re-scored top-ranked docking poses and scores are shown for hydration levels 0–4. For reference, the co-crystallized ligand structures of PSMA-617 and PSMA-1007 are displayed on the left. Arrows indicate the naphthyl group at the S1 lipophilic subsite. The pose selected for MD simulations is highlighted in green.

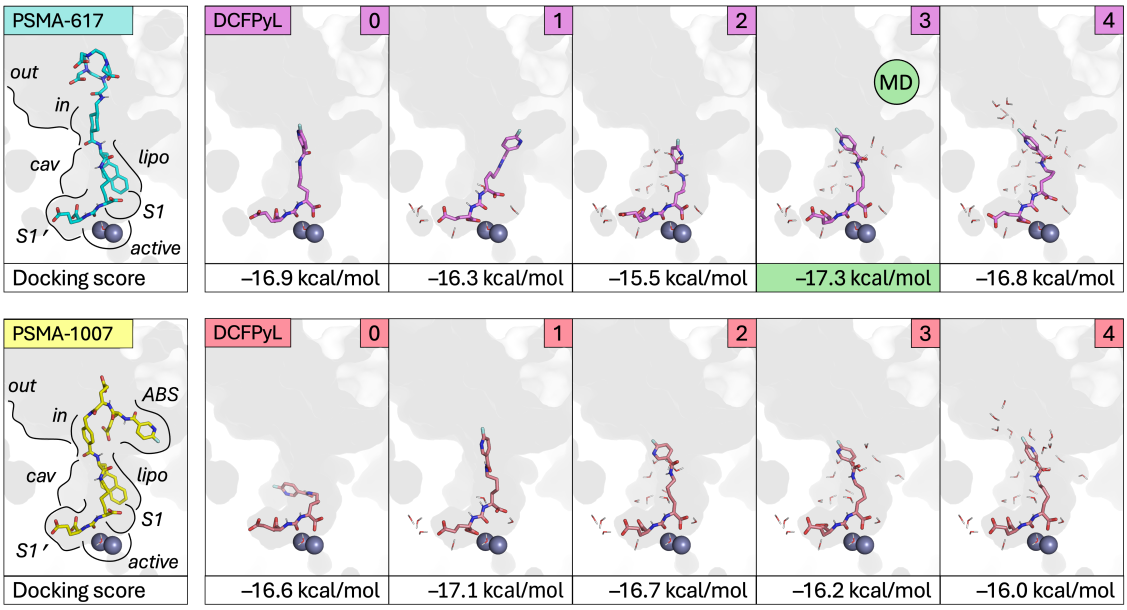


Figure A40.: Docking solutions of DCFPyL. The ligand DCFPyL was docked to both PSMA structures (PDB: 8BOW and 5O5T) using ADFR. Re-scored top-ranked docking poses and scores are shown for hydration levels 0–4. For reference, the co-crystallized ligand structures of PSMA-617 and PSMA-1007 are displayed on the left. The pose selected for MD simulations is highlighted in green.

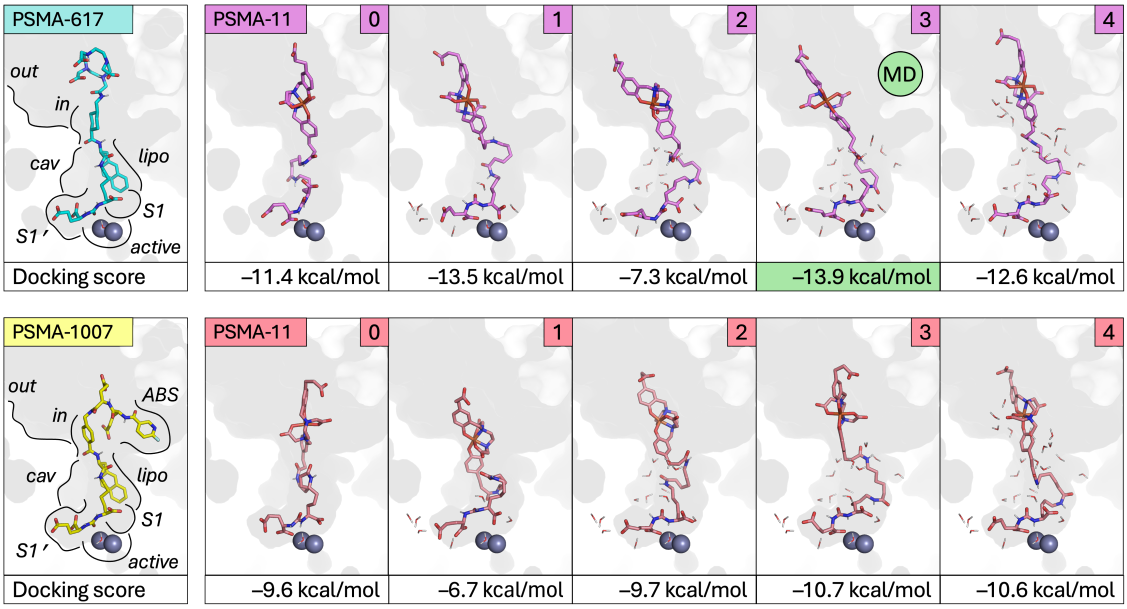


Figure A41.: Docking solutions of PSMA-11. The ligand PSMA-11 was docked to both PSMA structures (PDB: 8BOW and 5O5T) using ADFR. Re-scored top-ranked docking poses and scores are shown for hydration levels 0–4. Docking scores are provided below each pose. For reference, the co-crystallized ligand structures of PSMA-617 and PSMA-1007 are displayed on the left. The pose selected for MD simulations is highlighted in green.

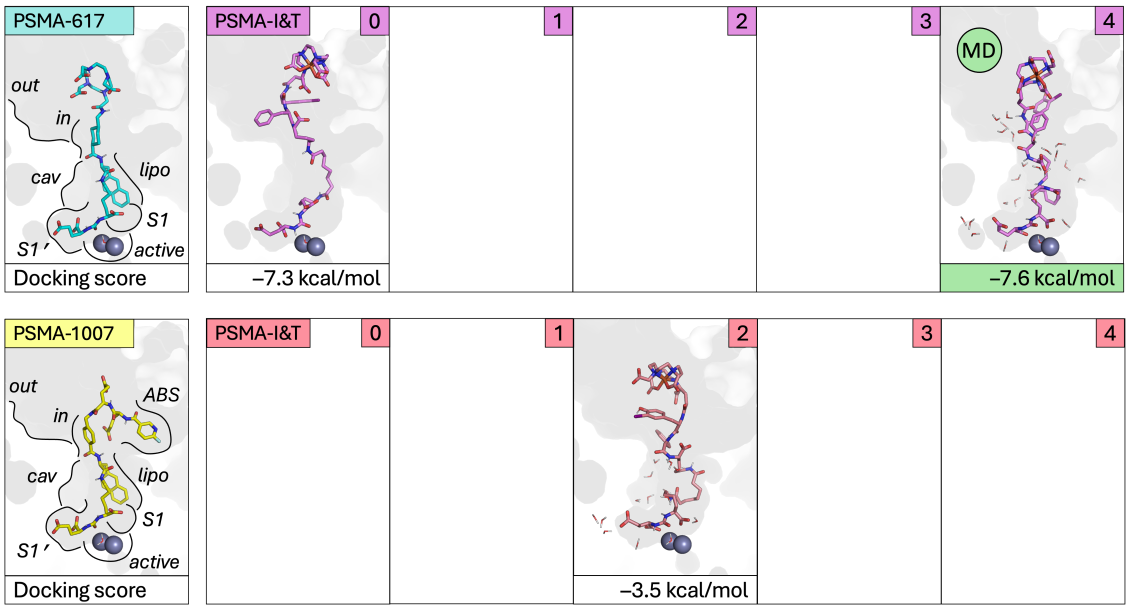


Figure A42.: Docking solutions of PSMA-I&T. The ligand PSMA-I&T was docked to both PSMA structures (PDB: 8BOW and 5O5T) using ADFR. Re-scored top-ranked docking poses and scores are shown for hydration levels 0–4. For reference, the co-crystallized ligand structures of PSMA-617 and PSMA-1007 are displayed on the left. The pose selected for MD simulations is highlighted in green.

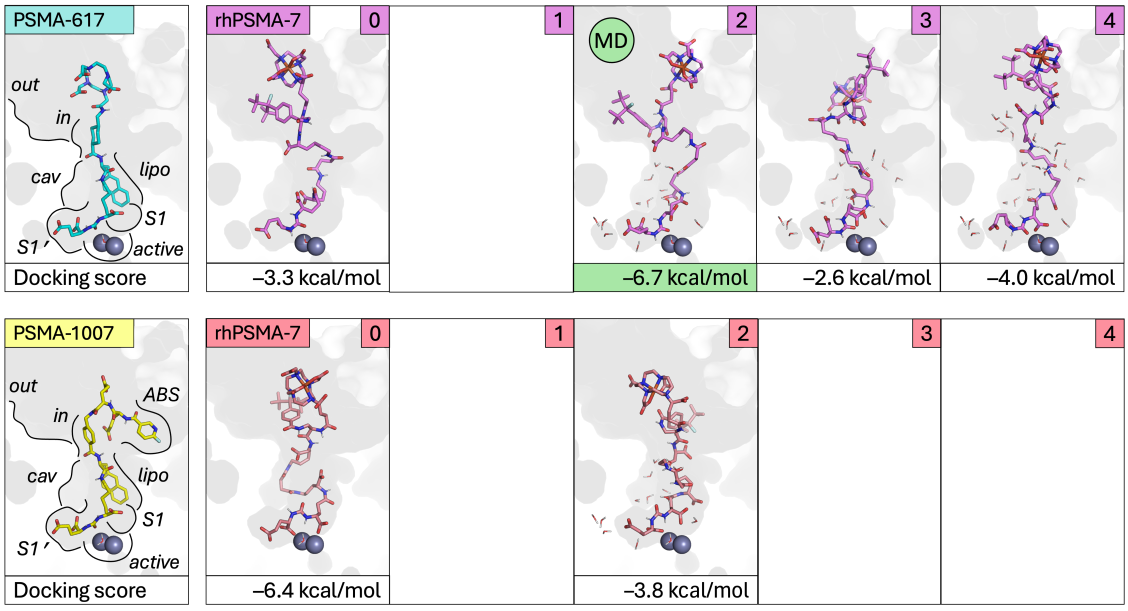


Figure A43.: Docking solutions of rhPSMA-7. The structure of rhPSMA-7 was docked to both PSMA structures (PDB: 8BOW and 5O5T) using ADFR. Re-scored top-ranked docking poses and scores are shown for hydration levels 0–4. Docking scores are provided below each pose. For reference, the co-crystallized ligand structures of PSMA-617 and PSMA-1007 are displayed on the left. The pose selected for MD simulations is highlighted in green.

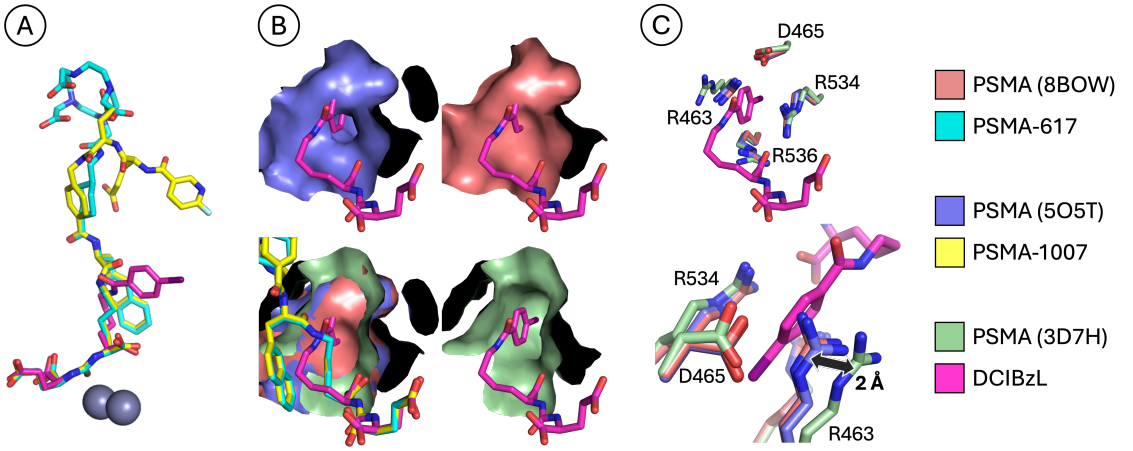


Figure A44.: Additional S1 lipophilic pocket (Arginine-patch pocket) addressed by PSMA inhibitor DCIBzL. The PSMA crystal structures co-crystallized with PSMA-617 (PDB: 8BOW), PSMA-1007 (PDB: 5O5T) and DCIBzL (PDB: 3D7H) aligned by C_α atoms using PyMOL[240]. (A) Superimposed stick models of PSMA-617, PSMA-1007, and DCIBzL. (B) PSMA surface around the iodobenzene moiety of DCIBzL in 5O5T (top left), 8BOW (top right), and 3D7H (bottom right), with all three inhibitors and PSMA structures combined (bottom left). The pocket harboring the iodobenzene moiety in 3D7H is absent in 5O5T and 8BOW. (C) Stick representation of PSMA residues surrounding the S1 pocket show that the pocket is formed through a 2 Å shift of the R463 sidechain.

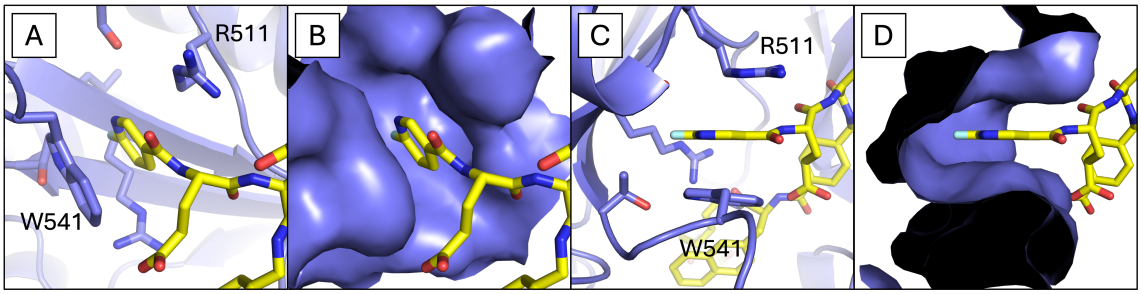


Figure A45.: The remote arene binding site accessed by PSMA-1007. The PSMA crystal structure co-crystallized with PSMA-1007 (PDB: 5O5T) is displayed as (A, B) side and (C, D) top views, focusing on the remote arene binding site accommodating the ligands 6-fluoronicotinic acid (FNA) moiety. PSMA-1007 is shown as sticks (yellow), while the protein and pocket-surrounding residues are depicted as cartoon and sticks (A, C) or as a surface (B, D).

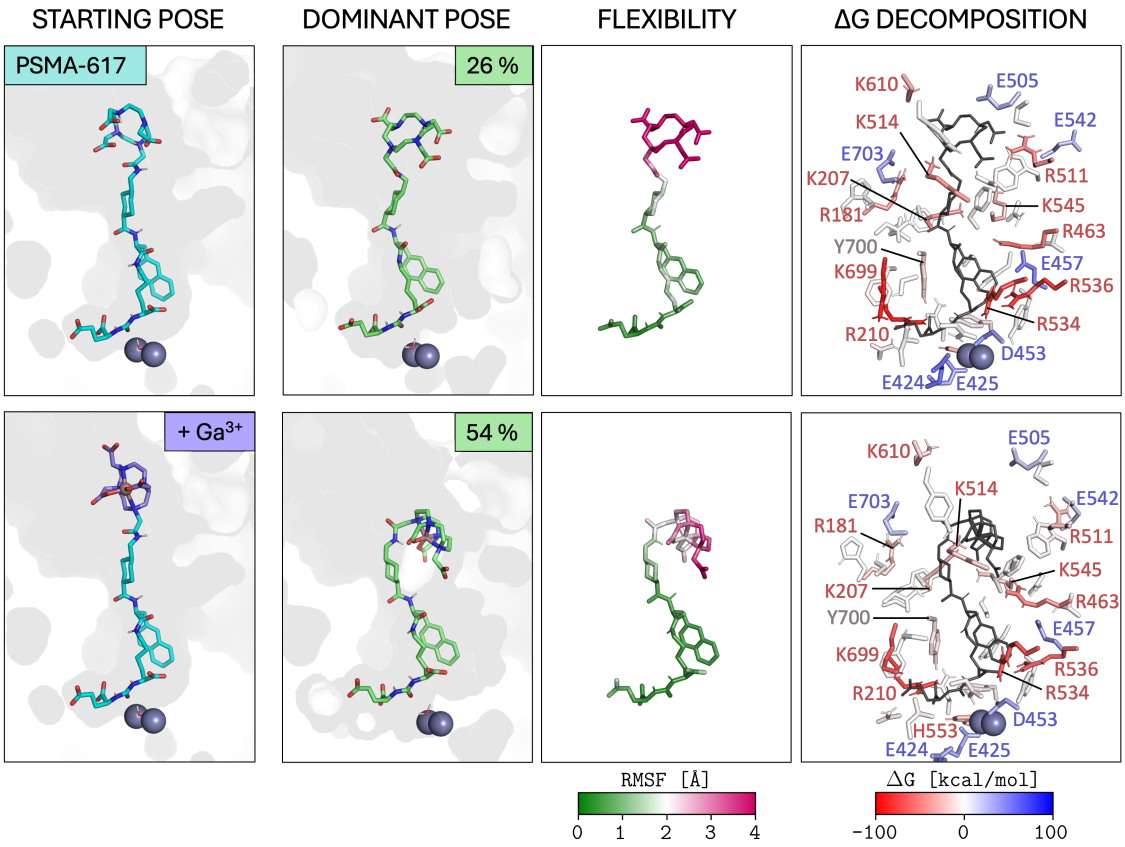


Figure A46.: Flexibility and interaction pattern of PSMA-617. The PSMA crystal structure (PDB: 8BOW) co-crystallized with non-labeled PSMA-617 (A) and its Ga-labeled analog (B)—obtained by replacing the chelator with a Ga(III) 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid (DOTA) crystal structure (violet)—were used as starting structures for 100 ns MD simulations. The predominant ligand conformations of the non-labeleld (26 %) and the Ga-labeled ligand (54 % sampled) were identified through RMSD-based clustering. The RMSF is represented by a color shift ranging from rigid (green) to flexible (pink). The per-residue ΔG_{bind} decomposition is displayed for the PSMA binding site residues defined by their presence of at least 80 % within 3.5 Å of all modeled PSMA ligands.

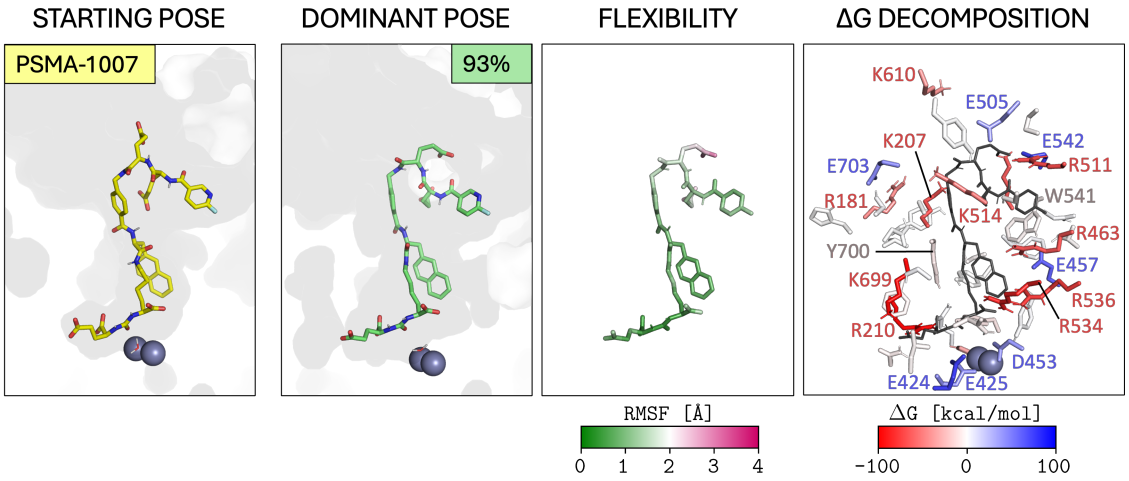


Figure A47.: Flexibility and interaction pattern of PSMA-1007. The PSMA crystal structure co-crystallized with PSMA-1007 (PDB: 5O5T) was used as starting structure for a 100 ns MD simulation. The predominant ligand conformation (93 % of the simulation) was identified through RMSD-based clustering. Ligand flexibility, as assessed by RMSF, is represented by a color shift ranging from rigid (green) to flexible (pink). The per-residue ΔG_{bind} decomposition is displayed for the PSMA binding site residues defined by their presence of at least 80 % within 3.5 Å of all modeled PSMA ligands.

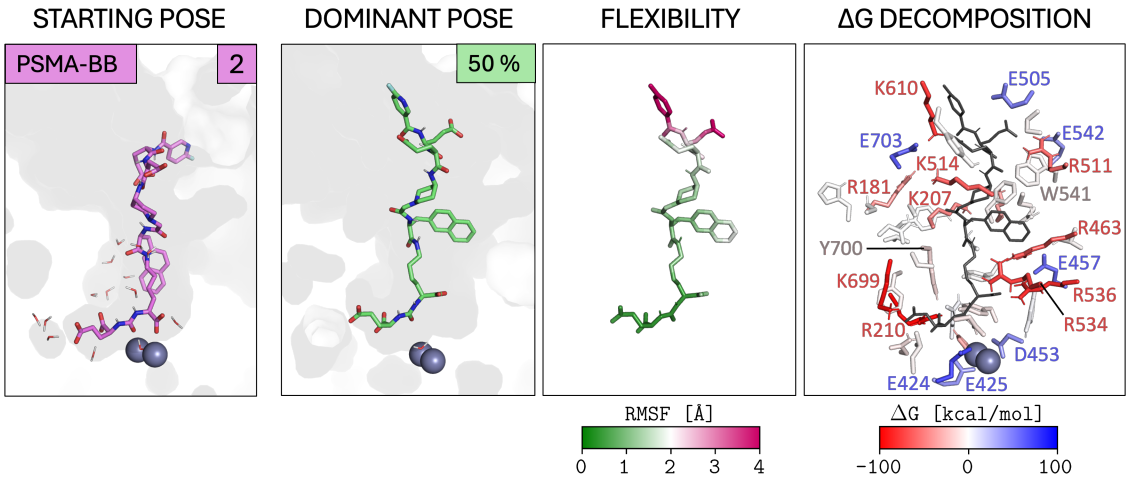


Figure A48.: Flexibility and interaction pattern of PSMA-BB. The top-ranked docking pose of PSMA-GCK01 from level 2, illustrated on the left, was used as the starting structure for a 100 ns MD simulation in complex with PSMA (PDB: 8BOW). The predominant ligand conformation (50 % of the simulation) was identified through RMSD-based clustering. Ligand flexibility, as assessed by RMSF, is represented by a color shift ranging from rigid (green) to flexible (pink). The per-residue ΔG_{bind} decomposition is displayed for the PSMA binding site residues defined by their presence of at least 80 % within 3.5 Å of all modeled PSMA ligands.

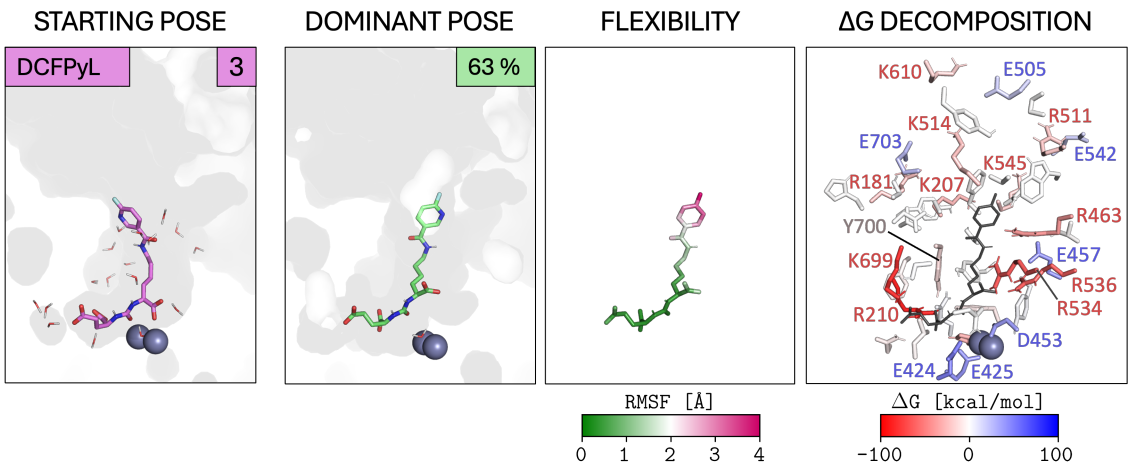


Figure A49.: Flexibility and interaction pattern of DCFPyL. The top-ranked docking pose of DCFPyL from level 3, illustrated on the left, was used as the starting structure for a 100 ns MD simulation in complex with PSMA (PDB: 8BOW). The predominant ligand conformation (63 % of the simulation) was identified through RMSD-based clustering. Ligand flexibility, as assessed by RMSF, is represented by a color shift ranging from rigid (green) to flexible (pink). The per-residue ΔG_{bind} decomposition is displayed for the PSMA binding site residues defined by their presence of at least 80 % within 3.5 Å of all modeled PSMA ligands.

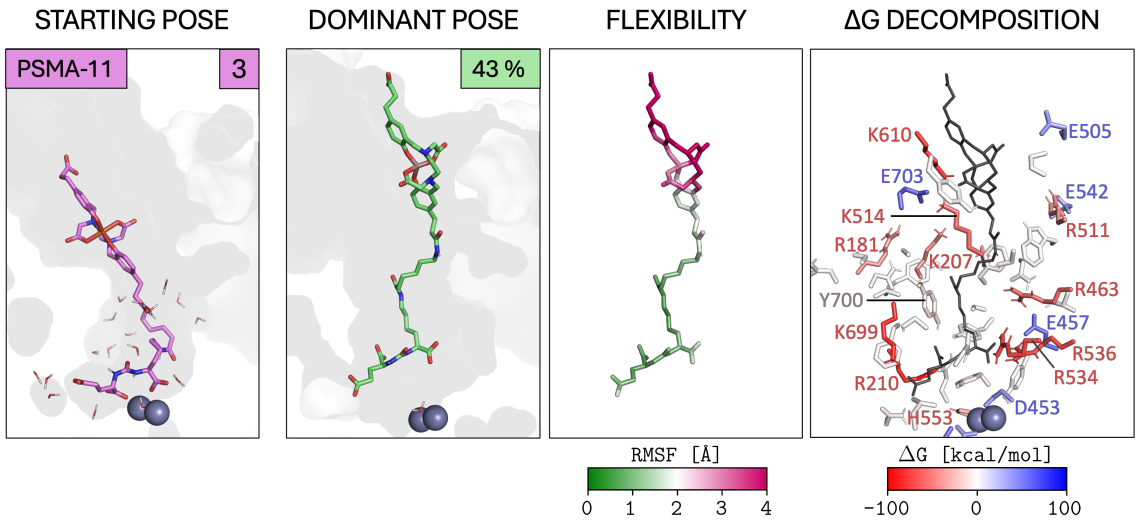


Figure A50.: Flexibility and interaction pattern of PSMA-11. The top-ranked docking pose of PSMA-11 from hydration level 3, illustrated on the left, was used as the starting structure for a 100 ns MD simulation in complex with PSMA (PDB: 8BOW). The predominant ligand conformation (43 % of the simulation) was identified through RMSD-based clustering. Ligand flexibility, as assessed by RMSF, is represented by a color shift ranging from rigid (green) to flexible (pink). The per-residue ΔG_{bind} decomposition is displayed for the PSMA binding site residues defined by their presence of at least 80 % within 3.5 Å of all modeled PSMA ligands.

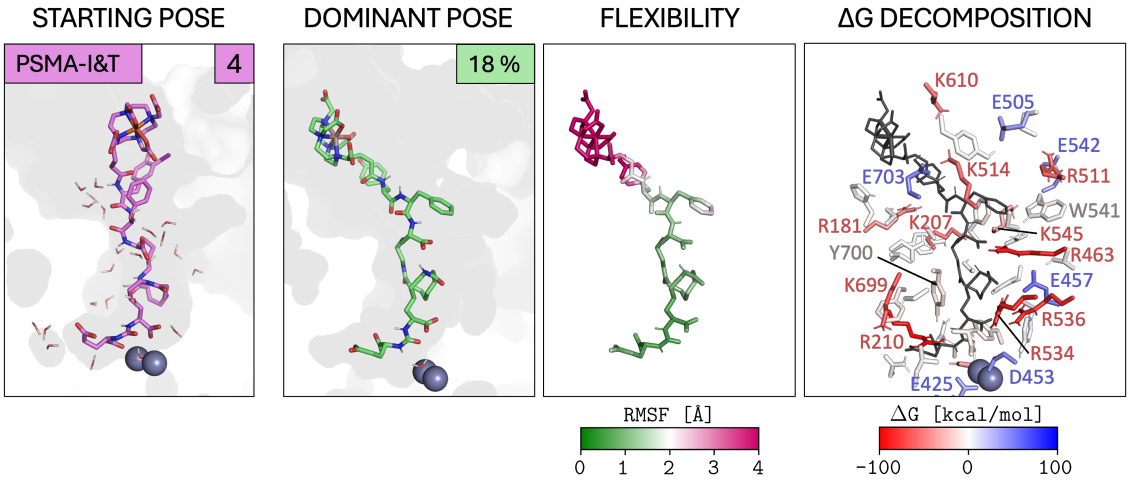


Figure A51.: Flexibility and interaction pattern of PSMA-P&T. The top-ranked docking pose of PSMA-P&T from level 4, illustrated on the left, was used as the starting structure for a 100 ns MD simulation in complex with PSMA (PDB: 8BOW). The predominant ligand conformation (18 % of the simulation) was identified through RMSD-based clustering. Ligand flexibility, as assessed by RMSF, is represented by a color shift ranging from rigid (green) to flexible (pink). The per-residue ΔG_{bind} decomposition is displayed for the PSMA binding site residues defined by their presence of at least 80 % within 3.5 Å of all modeled PSMA ligands.

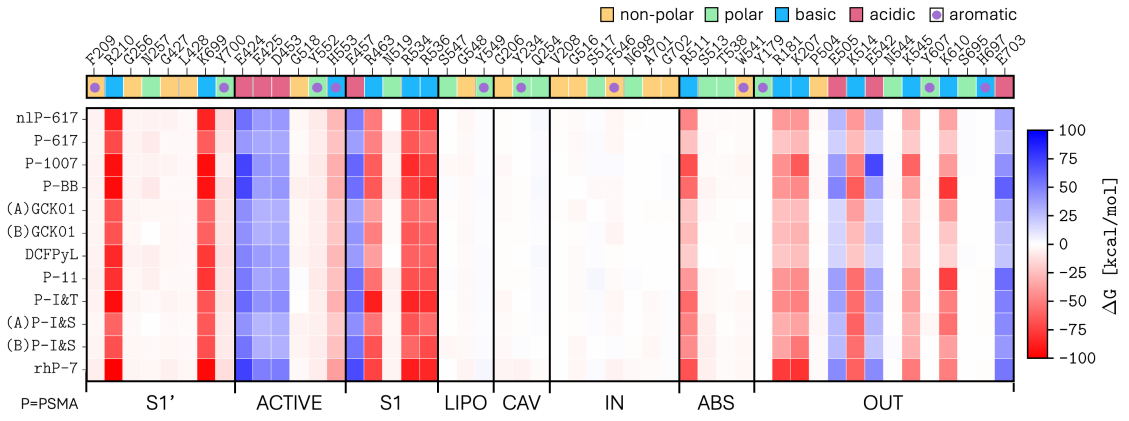


Figure A52.: Per-residue ΔG_{bind} decomposition of PSMA–ligand MD simulations. For each PSMA–ligand MD simulation, the per-residue decomposition of ΔG_{bind}, ranging from -100 to $+100$ kcal/mol, is shown for the binding site residues, which are categorized based on their location within the binding site: the S1′ glutamate pocket (S1′), the active site (including residues interacting with the urea moiety of the ligand, labeled ACTIVE), the S1 arginine patch (S1), the S1 lipophilic region (LIPOPHIL), the linker-surrounding area (LINKER), the arene binding pocket (ARENE), and the entrance of the binding funnel (ENTRANCE). Amino acid side chains are color-coded by property: non-polar (yellow), polar (green), basic (blue), and acidic (red). Aromatic residues are highlighted with a purple dot.

pubs.acs.org/acschemicalbiology

Articles

Discovery of All-D-Peptide Inhibitors of SARS-CoV-2 3C-like Protease

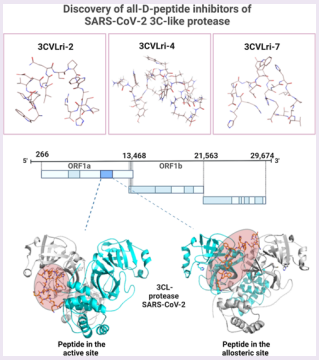
Raphael J. Eberle, Marc Sevenich, Ian Gering, Lara Scharbert, Birgit Strodel, Nils A. Lakomek, Karoline Santur, Jeannine Mohrlüder, Mônika A. Coronado,* and Dieter Willbold*

Cite This: *ACS Chem. Biol.* 2023, 18, 315−330

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: During the replication process of SARS-CoV-2, the main protease of the virus [3-chymotrypsin-like protease (3CL^{pro})] plays a pivotal role and is essential for the life cycle of the pathogen. Numerous studies have been conducted so far, which have confirmed 3CL^{pro} as an attractive drug target to combat COVID-19. We describe a novel and efficient next-generation sequencing (NGS) supported phage display selection strategy for the identification of a set of SARS-CoV-2 3CL^{pro} targeting peptide ligands that inhibit the 3CL protease, in a competitive or noncompetitive mode, in the low μM range. From the most efficient L-peptides obtained from the phage display, we designed all-D-peptides based on the retro-inverso (ri) principle. They had IC_{50} values also in the low μM range and in combination, even in the sub-micromolar range. Additionally, the combination with Rutinprivir decreases 10-fold the IC_{50} value of the competitive inhibitor. The inhibition modes of these D-ri peptides were the same as their respective L-peptide versions. Our results demonstrate that retro-inverso obtained all-D-peptides interact with high affinity and inhibit the SARS-CoV-2 3CL protease, thus reinforcing their potential for further development toward therapeutic agents. The here described D-ri peptides address limitations associated with current L-peptide inhibitors and are promising lead compounds. Further optimization regarding pharmacokinetic properties will allow the development of even more potent D-peptides to be used for the prevention and treatment of COVID-19.



1. INTRODUCTION

The spread of Coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), with increasing levels of infectivity and transmissibility, has strained human health and public safety worldwide.[1] By May 2022, the COVID-19 pandemic has resulted in more than 524 million confirmed cases and more than 6.2 million confirmed deaths, according to the World Health Organization (WHO).[2] To date, vaccination is considered the key strategy for ending the pandemic[3] and the worldwide vaccination campaign using clinical safe and efficient vaccines against SARS-CoV-2 have controlled the number of death. However, not yet the spread of the diseases.[4−7] So far, more than 11.8 billion vaccine doses have been administered.[2] Remdesivir, Dexamethasone, Favipiravir, Lopinavir/Ritonavir, Nirmatrelvir/Ritonavir (Main protease inhibitor), and Darunavir have been approved for emergency use to inhibit SARS-CoV-2 infection and replication.[8−12] Given a considerable limitation of direct-acting antivirals for COVID-19 and an increasing presence of SARS-CoV-2 variants (B.1.1.529, B.1.617.2, B.1.1.7, B.1.351, A.23.1, B.1.525, B.1.526 and P.1),[13] it remains a strategic priority to develop new drug candidates with minimal side effects and also targeting the new variants.

For the expression and replication of the CoV gene, proteases play essential roles involving the proteolytic processing of replicase polyproteins, which makes them attractive targets for drug development.[14,15] The main protease, also called 3CL protease (3CL^{pro}), is the 3-chymotrypsin-like cysteine protease that features a His41-Cys145 catalytic dyad. A glutamine (Gln) residue is commonly located at the P1 site on the protease substrate. No known human cysteine protease cleaves after Gln, thus offering potential selectivity for this viral target over human proteases.[16,17] Viral proteases have been attractive targets for oral small-molecule therapies in treating HIV and HCV.[18,19] The use of SARS-CoV-2 3CL^{pro} as a drug target has several advantages over other virus proteins: (i) its essential role in the viral replication process, (ii) its potential for mechanistic safety, and (iii) the expected lack of variant resistance challenges.[8] 3CL^{pro} inhibition represents an attractive approach for a safe and orally available antiviral therapy to treat COVID-19.

Drugs are conventionally classified into two molecule weight classes: "small-molecule" drugs (<500 Da) and protein/peptide-based drugs (>1000 Da).[20] Besides the obvious advantages of small molecules like favorable oral bioavailability and a rational design,[21] they often have low target selectivity, ultimately resulting in side effects. In comparison, peptides were often neglected as potential molecules for drug development, despite binding to their target proteins with high affinity and specificity. They are smaller than proteins and can be obtained synthetically

Received: September 22, 2022
Accepted: December 1, 2022
Published: January 17, 2023

Downloaded via 78.34.154.220 on April 14, 2024 at 16:05:32 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

ACS Publications

© 2023 The Authors. Published by American Chemical Society

https://doi.org/10.1021/acschembio.2c00735
ACS Chem. Biol. 2023, 18, 315−330

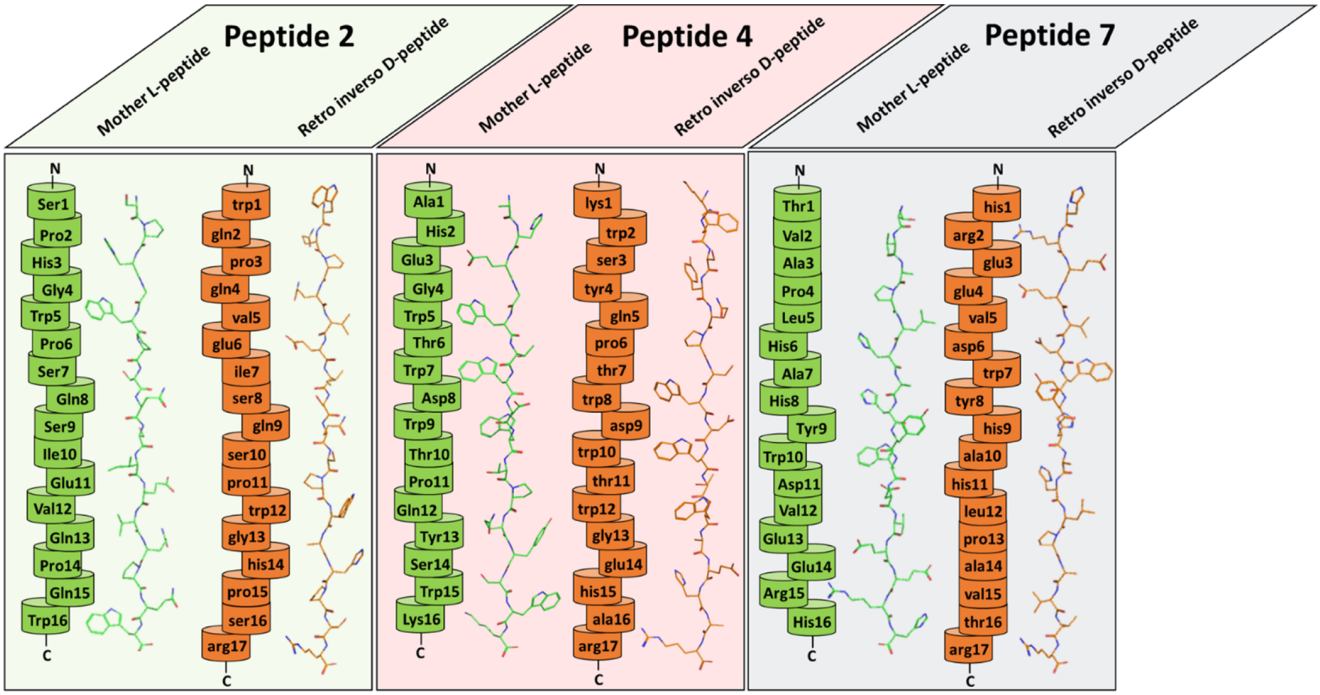


Figure 1. Schematic overview of the investigated D-ri peptides and their mother L-peptides.

by well-established and cost-efficient methods.[22] The disadvantages of peptide drugs are their potential immunogenicity and low bioavailability due to degradation and short half-lives. Several approaches have been developed to approach these problems. D-Enantiomeric amino acid residues effectively enhance the resistance to degradation since most proteolytic enzymes have substrate specificity for L-peptide bonds, allowing oral administration of D-enantiomeric peptides.[23,24] Furthermore, all-D-peptides are less immunogenic.[25]

Several studies described the development of the first all-D-peptide designed to treat Alzheimer's disease, and it proved to be efficient in animal models and safe in humans.[26−28] In the case of SARS-CoV-2, several studies describe the identification of D-peptide inhibitors against SARS-CoV-2 Spike protein and 3CL^{pro}.[29−32] Of particular interest are all-D-peptides composed of D-amino acid residues. However, many peptide-based drug development strategies, like phage display selection, end up with L-peptides. So-called D-retro-inverso (D-ri) peptides are composed solely of D-amino acid residues in the reverse sequence of their parental L-peptide (Figure 1). Such a peptide assembled in reversed sequence from D-amino acid residues will have almost the same structure, stability, and bioactivity as its parent peptide made of L-amino acid residues, but with the advantage of being more resistant to proteolytic degradation. This combination makes D-ri peptides attractive drug candidates.[33] The overall orientation of the side chains of the D-ri peptides is very similar to the original L-peptides, which, however, does not lead to perfect D-ri peptides, which is undoubtedly one reason why the retro-inverso approach was not successful in every case.[34−36]

However, in this paper, we demonstrate the inhibition potential of L-peptides selected using phage display targeting SARS-CoV-2 3CL^{pro}. Subsequently, its retro-inverso versions were explored, also in combination.

2. RESULTS

2.1. Characterization of SARS-CoV-2 3CL^{pro} Inhibition by L- and D-Retro-Inverso Peptides. All experiments were performed with recombinant produced SARS-CoV-2 3CL^{pro} protein (Supporting Information, Figure S1). To select binding L-peptides to the 3CL^{pro}, M13 phage display selection was performed (Figure 2A). Detailed information about the phage display process is described in Supporting Information, Text S1, Figures S2 and S3, and Tables S1−S3. To verify whether the selected peptides showed inhibition toward the target protein, they were primary tested using a final concentration of 10 μM of each peptide, separately, in an enzyme-based fluorescence assay (Figure 2B). With more than 60% inhibition, the peptides 3CVL-2 (SPHGWPSQSIEVQPQW), 3CVL-4 (AHEGWTWDWTPQYSWK), and 3CVL-7 (TVAPLHAHYWDVEERH) were selected for further analysis (Figure 2C−E). Concerning their potential to inhibit 100% of the protease activity, determination of the IC_{50} was performed using 0−140 μM (3CVL-2), 0−120 μM (3CVL-4), and 0−100 (3CVL-7). The 3CVL-2 at a final concentration of 140 μM showed 100% inhibition of the protease (Supporting Information, Figure S4A) with a calculated IC_{50} value of 4.90 ± 0.8 μM (Table 1 and Figure 2C). In a similar concentration range, the tested 3CVL-4 inhibited 100% of the protease activity at a final concentration of 120 μM (Supporting Information, Figure S4C), presenting an IC_{50} value of 2.44 ± 0.6 μM (Table 1 and Figure 2D). The selected peptide 3CVL-7 performs even better and inhibits 100% of the recombinant SARS-CoV-2 protease activity at a final concentration of 80 μM (Supporting Information, Figure S4E) with an IC_{50} value of 0.98 ± 0.2 μM (Table 1 and Figure 2E). Further inhibition mode assays identified 3CVL-2 and 3CVL-7 as competitive inhibitors. These peptides compete with the substrate for the active and/or binding sites of the protease (Supporting Information, Figure S4B,F). Remarkably, the inhibition mode has shown that 3CL^{pro} can be allosterically inhibited by 3CVL-4, which discloses a

https://doi.org/10.1021/acschembio.2c00735
ACS Chem. Biol. 2023, 18, 315−330

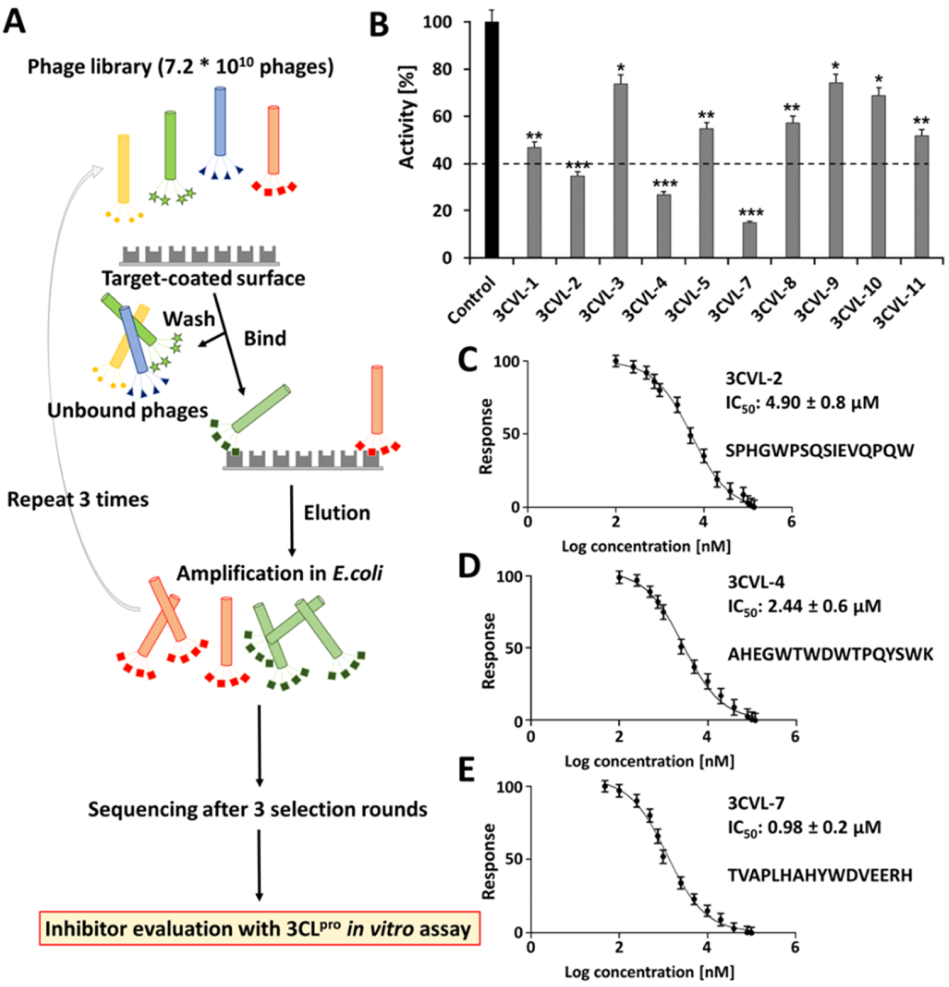


Figure 2. Phage display scheme and screening of phage display L-peptides against SARS-CoV-2 3CL^{pro} activity. (A) Schematic diagram of the phage display process. (B) Primary screen of the 11 selected peptides (10 μM) against SARS-CoV-2 3CL^{pro} activity. 3CVL-2, 3CVL-4, and 3CVL-7 inhibit the virus protease activity by more than 60%. (C−E) Selected peptide inhibiting SARS-CoV-2 3L^{pro}. Dose−response curves for IC_{50} values of 3CVL-2 (C), 3CVL-4 (D), and 3CVL-7 (E) were determined by nonlinear regression. Data shown are the mean ± standard deviation (SD) from three independent measurements ($n = 3$). Asterisks mean that the data differ from the control (0 μM inhibitor) significantly at $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***) levels according to analyses of variance (ANOVA) and Tukey's test.

Table 1. Summary of the SARS-CoV-2 3CL^{pro} Inhibition Experiments by 3CVL-2, -4, and -7

peptide	IC_{50} (μM)	inhibition mode
3CVL-2	4.90 ± 0.8	competitive
3CVL-4	2.44 ± 0.6	noncompetitive
3CVL-7	0.98 ± 0.2	competitive

noncompetitive inhibition mode (Supporting Information, Figure S4D).

To improve the poor proteolytic stability of L-enantiomeric peptides without losing affinity and binding specificity to their target protein, the peptides 3CVL-2, 3CVL-4, and 3CVL-7 were synthesized in retro-inverse D-enantiomeric (D-ri) form and designated as 3CVLri-2, 3CVLri-4, and 3CVLri-7, respectively. Glycine does not have a chiral center that was the reason that the D-ri peptides have identical amino acid sequences, except for glycine, which was replaced by alanine. Additionally, an arginine was added to the C-termini of each D-ri peptide to increase the permeability of the peptides. The representative scheme is shown in Figure 3A,B.

Like the L-peptides, the D-ri-peptides were tested using a concentration range of up to 140 μM (3CVLri-2 and 3CVLri-4) and up to 100 μM (3CVLri-7), which showed 100% inhibition of the protease by the corresponding used final concentration (Supporting Information, Figure S5A,C,E). Based on the enzymatic assay, the IC_{50} values were determined for each D-ri peptide, displaying for 3CVLri-2 (9.09 ± 0.5 μM), 3CVLri-4 (5.36 ± 1.7 μM) and 3CVLri-7 (1.57 ± 0.2 μM) an increase of the IC_{50} values of about 2-fold compared to the L-peptides (Figure 3C−E and Table 2). The mode of inhibition of the D-ri peptides was maintained, displaying a competitive mode of 3CLVri-2 and -7 and an allosteric mode of inhibition for 3CVLri-7 (Supporting Information, Figure S5B,D,F). A sequence analysis between the preferred 3CL^{pro} cleavage sequence pattern at the substrate-binding site and the identified inhibitors demonstrated that the competitive inhibitors have a higher similarity (ranging from 62.5 to 12.5%) compared to the noncompetitive inhibitors (no similarity) (Supporting Information, Figure S6).

2.2. Competitive and Noncompetitive D-Retro-Inverso Peptides Act Cooperatively to Inhibit SARS-CoV-2 3CL^{pro}. To presume the effect of the competitive (3CVLri-2

https://doi.org/10.1021/acschembio.2c00735
ACS Chem. Biol. 2023, 18, 315−330

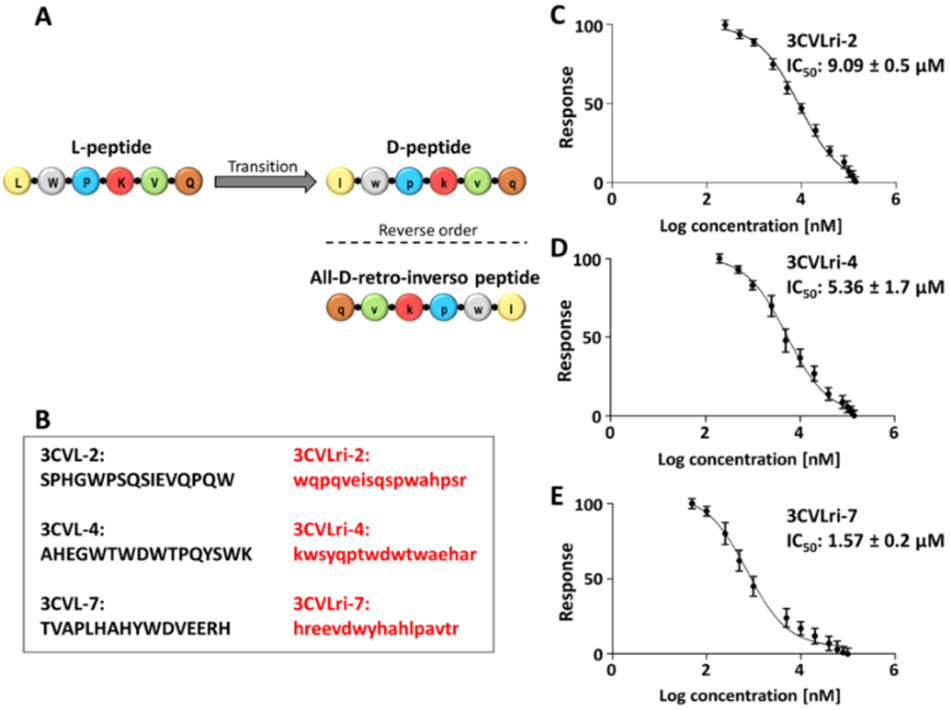


Figure 3. Principle for obtaining D-retro-inverso peptides and inhibitory effect against SARS-CoV-2 $3CL^{pro}$ activity. (A) Schematic diagram of D-retro-inverso (D-ri) peptides. (B) Sequences of 3CVLri-2, -4, -7 and their mother L-peptides. In the D-ri peptides, glycine (G) was replaced by alanine (a). (C–E) All-D-ri peptides inhibit SARS-CoV-2 $3CL^{pro}$ activity. Dose–response curves for IC_{50} values of 3CVLri-2 (C), 3CVLri-4 (D), and 3CVLri-7 (E) were determined by nonlinear regression. Data shown are the mean ± SD from three independent measurements (n = 3).

Table 2. Summary of the SARS-CoV-2 $3CL^{pro}$ Inhibition Experiments by 3CVLri-2, -4 and -7

peptide	IC_{50} (μM)	inhibition mode
3CVLri-2	9.09 ± 0.5	competitive
3CVLri-4	5.36 ± 1.7	noncompetitive
3CVLri-7	1.57 ± 0.2	competitive

and 3CVLri-7) and noncompetitive (3CVLri-4) D-retro-inverso peptides, a combined inhibitory assay was performed (3VLri-2 + 3CVLri-4 and 3CVL-7 + 3CVL-4). Both peptides were mixed in a 1:1 volume ratio and tested against the protease activity (Figure 4).

The calculated IC_{50} values for the 3CVLri-2 + 3CVLri-4 combination and the 3CVLri-7 + 3CVLri-4 combination were 0.87 ± 0.3 and 0.39 ± 0.1 μM (Figure 4A,B).

2.3. Cooperative Inhibition of SARS-CoV-2 $3CL^{pro}$ by Rupintrivir and 3CVLri-4. Rupintrivir is a known inhibitor of the 3C cysteine protease family but showed a weak inhibitory potential against the SARS-CoV-2 $3CL^{pro}$.[37−39] The determined IC_{50} value for Rupintrivir was 109 ± 14.40 μM (Figure 5A) and is in agreement with published data.[34,35] The IC_{50} value for the Rupintrivir/3CVLri-4 combination was 15.10 ± 1.60 μM (Figure 5B).

2.4. Binding Affinities of Selected $3CL^{pro}$ L- and D-Peptide Inhibitors. The interaction kinetics of 3CVL-2, -4, and -7 with $3CL^{pro}$ and its D-ri form was determined using surface plasmon resonance (SPR) experiments. $3CL^{pro}$ was immobilized via covalent primary amino group coupling, and 3CVL and 3CVLri peptides were injected as analytes. Supporting Information Figures S9 and S10 show the SPR sensorgrams for 3CVL peptides at assay concentrations of 50,

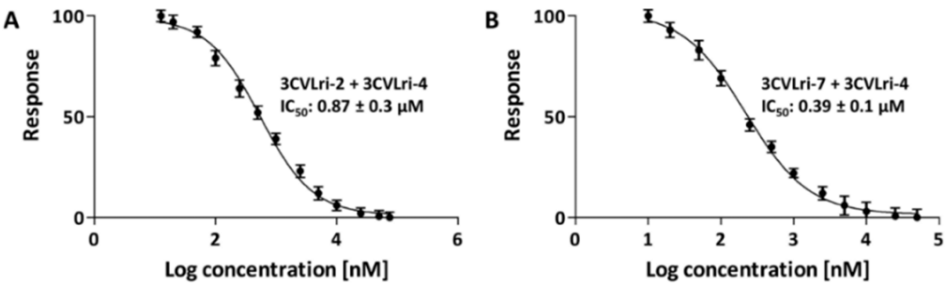


Figure 4. SARS-CoV-2 $3CL^{pro}$ inhibition by a combination of competitive and noncompetitive D-retro-inverso peptides simultaneously. Dose–response curves for IC_{50} values of the combination were determined by nonlinear regression. Data shown are the mean ± SD from three independent measurements (n = 3). The corresponding normalized activity and inhibition plots of SARS-CoV-2 $3CL^{pro}$ under the influence of the combined peptides are shown in the Supporting Information, Figure S7. (A) Dose–response curve of the 3CVLri-2 + 3CVLri-4 combination. (B) Dose–response curve of the 3CVLri-7 + 3CVLri-4 combination.

https://doi.org/10.1021/acschembio.2c00735
ACS Chem. Biol. 2023, 18, 315−330

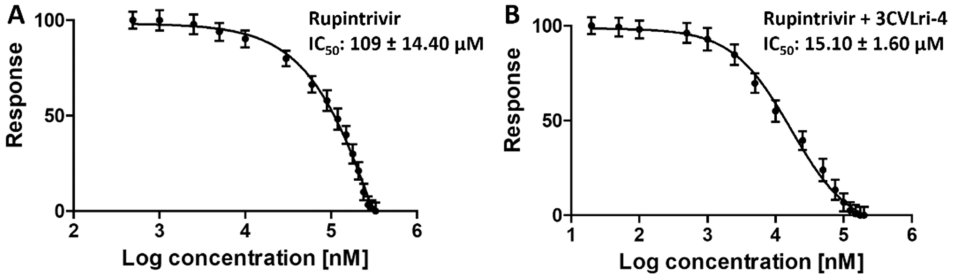


Figure 5. SARS-CoV-2 $3CL^{pro}$ inhibition by Rupintrivir and Rupintrivir/3CVLri-4 combination. Dose−response curves of the IC_{50} values were determined by nonlinear regression. Data shown are the mean ± SD from three independent measurements ($n = 3$). The corresponding normalized activity and inhibition plots of SARS-CoV-2 $3CL^{pro}$ under the influence of the combined peptides are shown in Supporting Information, Figure S8. (A) Dose−response curve of Rutinprivir. (B) Dose−response curve of the Rutinprivir/3CVLri-4 combination.

25, 12.5, 6.25, 3.12, 1.56, 0.78, and 0.39 μM on $3CL^{pro}$ immobilized on the CM5 sensor chip. The fit for the competitive inhibitors was made using a heterogeneous ligand model in which the two binding sites on the dimer are assumed to behave differently, as described for the $3CL^{pro}$ dimer active sites.[40−42] The fit for the noncompetitive inhibitor was made using a 1:1 stoichiometric kinetic fitting.

Heterogeneous ligand fitting revealed dissociation constants from the low μM to low nM range for the higher affine binding site. Global 1:1 fitting showed dissociation constants in the low μM range (Supporting Information, Figures S9 and S10 and Table 3).

Table 3. K_D Values Determined by SPR Experiments

heterogeneous ligand fitting		
peptide	K_D1	K_D2
3CVL-2	0.5 nM	6.5 μM
3CVL-7	1.8 μM	7.8 μM
3CVLri-2	46.8 μM	1.7 μM
3CVLri-7	2.4 nM	35 μM
1:1 ligand fitting		
peptide	K_D	
3CVL-4	5.1 μM	
3CVLri-4	25.2 μM	

The binding kinetic parameters of the association rate k_a, dissociation rate k_d, and dissociation constant K_D (k_d/k_a) for the interaction of $3CL^{pro}$ with peptides 3CVL-2, -4, -7, and 3CVLri-2, -4, and -7 are shown in Supporting Information, Figures S9 and S10.

2.5. 24 h Stability and Promiscuous Assays of 3CVLri Peptides. Peptide-based inhibitors targeting active proteolytic proteases can lose their inhibitory effect over time. Based on our focus, the retro-inverso peptides were tested regarding their stability over 24 h. The results demonstrated a constant inhibition of SARS-CoV-2 $3CL^{pro}$ over time, showing that the peptides are not prone to the 3CL protease digestion (Supporting Information, Figure S11). Furthermore, a detergent-based control was carried out to exclude peptide inhibitors that possibly act as an aggregator of $3CL^{pro}$, so-called "promiscuous" inhibitors. The experiment was performed by adding 0.001, 0.01, and 0.1% of Triton X-100 detergent to the reaction. Suppose that a molecule exhibits significant inhibition of $3CL^{pro}$, but this is diminished by detergent; it may act as an aggregation-based inhibitor. This was not observed for the tested 3CVLri peptides (Supporting Information, Figure S12).

2.6. Metabolic Stability of 3CVLri Peptides. To investigate the 3CVLri-2, 3CVLri-4, and 3CVLri-7 peptides' resistance against enzymatic degradation, the peptides were incubated in media simulating the gastrointestinal tract, blood, and liver, and the unmetabolized peptides were quantified by reversed-phase high-performance liquid chromatography (RP-HPLC). In simulated gastric fluid (SGF), the resistance of 3CVLri-2, 3CVLri-4, and 3CVLri-7 was monitored over 8 h. The results showed that the 3CVLri-4 and -7 peptides are relatively stable (≥80 ± 3%) within 8 h. In contrast, 3CVLri-2 was remarkably metabolized (∼75 ± 2%) (Figure 6A). The resistance of the D-ri peptides was also investigated using a simulated intestinal fluid (SIF). The peptides 3CVLri-2 and 3CVLri-4 remained rather stable over the selected time (≥90 ± 1.5%), as well as 3CVLri-7 (≥85 ± 5%) (Figure 6B). In human plasma, the stability of 3CVLri-2 and 3CVLri-7 was monitored for 48 h, with approximately 5% being metabolized (≥95 ± 4.5%). In contrast, the 3CVLri-4 showed about 21 ± 5.6% of degradation within 48 h (Figure 6C). Furthermore, human liver microsomes were monitored for 24 h, and 3CVLri-2 and 3CVLri-7 were about 5 ± 1.1% metabolized. However, 3CVLri-7 was almost 50% metabolized (≥45 ± 3.7%) in the same period (Figure 6D). The corresponding HPLC profiles of the stability tests are shown in Supporting Information, Figures S13−S16.

2.7. *In Silico* Structural Properties of 3CVL and 3CVLri Peptides and Their Similar Interaction Mode with SARS-CoV-2 $3CL^{pro}$. Protease-peptide docking and subsequent molecular dynamics (MD) simulations were performed to investigate the possible binding interface between $3CL^{pro}$-L and D-ri-peptide complexes. The secondary structure of the three-dimensional (3D) models of the L-peptides (3CVL-2, -4, and -7) and their D-ri counterparts corroborates with circular dichroism (CD) spectroscopy results.

The models are supported by the results of the inhibition assay. The inhibition mode assay identified two peptides as competitive inhibitors, which requires interaction in the active site and/or in the substrate-binding site. Another peptide was identified as a noncompetitive inhibitor, which requires interaction with an allosteric site. Indeed, the two competitive inhibitors docked in the active site formed a very stable interaction during the MD simulations. The noncompetitive inhibitor was docked in an allosteric site, which was described previously and showed a stable interaction through MD simulation.

During the 1 μs MD simulations of the unbound peptides, they mainly contained a random coiled secondary structure. However, the L- and D-ri-3CVL-4 peptides also frequently formed an α-helix (Supporting Information, Figure S17A−C).

https://doi.org/10.1021/acschembio.2c00735
ACS Chem. Biol. 2023, 18, 315−330

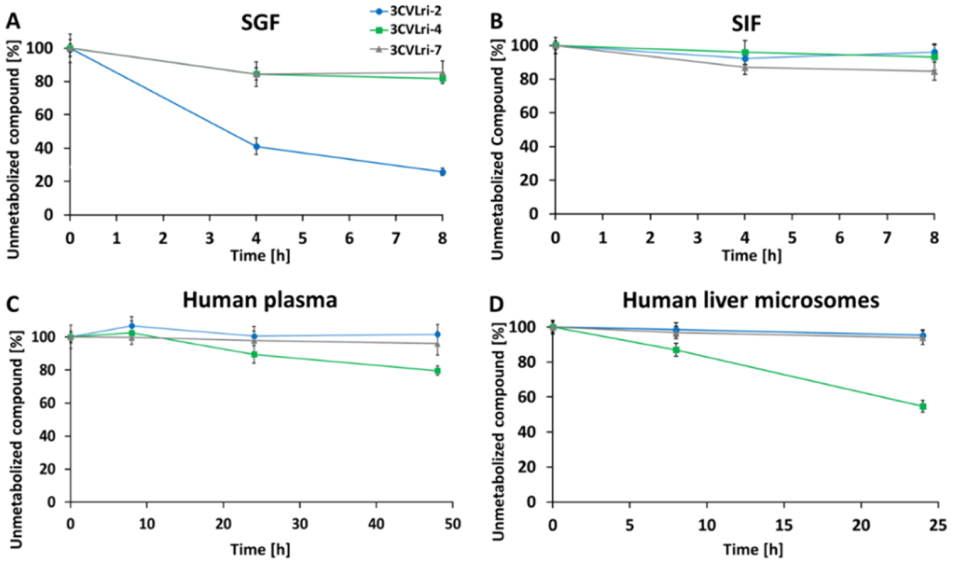


Figure 6. Stability of 3CVLri-2, 3CVLri-4, and 3CVLri-7 peptides incubated in SGF, SIF, human plasma, and human liver microsomes. (A) SGF, (B) SIF, (C) human plasma, and (D) human liver microsomes. Unmetabolized peptides were quantified by P-HPLC. The peak areas of the unmetabolized peptides after different incubation times were normalized to the peptides' peak areas after direct extraction from the media. Data are presented as mean ± SD (n = 3).

The atomic coordinates of SARS-CoV-2 3CL^{pro} dimer (PDB entry: 6M2N) and ten representative structures per peptide from the peptide simulations were submitted for docking using AutoDock Vina.[43]

The competitive peptides 3CVL-2/3CVLri-2 and 3CVL-7/3CVLri-7 were docked into the active site and substrate-binding site of the protease (Supporting Information, Table S4). The noncompetitive peptides 3CVL-4 and 3CVLri-4 were docked in a potential allosteric binding site (Supporting Information, Table S4). Previous studies have indicated that only one active site of the dimer is functional.[40,42] Based on that information, 200 ns control MD simulation of the dimer without ligands was performed to identify a potentially less stable protomer or collapsing binding sites. Based on the volume of the active sites, we did not identify any differences between both protomers (Supporting Information, Figure S18). Therefore, we docked the competitive peptides into the active site of chain A. After docking, 100 ns MD simulations with 3 different initial structures of the 3CL^{pro}/peptide complexes obtained from docking were conducted to further assess their stabilities. The initial structures were selected based on the calculated binding free energy and the distance to the critical catalytic dyad residues (Cys145 and His41). The flexibility of the peptides in complex with the protease was monitored by calculating the root-mean-square deviation (RMSD), root-mean-square fluctuations (RMSF), and ΔH as the enthalpy of binding (Supporting Information, Figures S19−S21). In all three simulations of each 3CL^{pro}/peptide complex, the peptide maintained its overall conformation in the binding pocket and no unbinding event was detected. Representative structures for all three simulations of each L- and D-ri-peptide are shown in Supporting Information, Figures S22 and S23. To investigate the 3CL^{pro} peptide interactions in more detail, we selected one of the three MD simulations per peptide based on several criteria, such as the RMSD, RMSF, and the binding free energy, elongated the simulations to 200 ns and conducted them in triplicate. Each replica of the elongated simulations showed a stable interaction between the L- and D-ri-3CVL peptides and their respective binding site of the 3CL^{pro} dimer throughout the simulation.

To identify the most popular binding modes sampled during the elongated simulations, we applied conformational clustering but considered only the nonflexible/interacting residues (RMSF ≤ 5 Å) of the peptides (Supporting Information, Figure S24).

Interactions between the residues of the peptides and the protease were analyzed by calculating the minimum distance between them for the first three clusters per complex (Table 4).

Table 4. Summary of MD Simulation Triplicates Considering Nonflexible/Interacting Residues of Peptides

peptide	#cluster	representative frames first 3 clusters (%)
3CVL-2	8	97
3CVL-4	12	88
3CVL-7	25	80
3CVLri-2	12	83
3CVLri-4	10	99
3CVLri-7	14	93

Analysis of the 3CL^{pro}-peptide complexes provides valuable information on the interaction interface similarities between 3CVL and 3CVLri peptides. Structural representation (Figure 7) demonstrated that 3CVL-2 extends along with S1′, S1, and S3 subpockets of the protease substrate-binding site (Figure 7A).

In contrast, the 3CVLri-2 interacts with the substrate-binding site differently. However, it extends along with S1′, S1, S2, and S3 (Figure 7B), which shows that one of the D-Gln residues of the 3CVLri-2 is very well accommodated. The 3CVL-7 and 3CVLri-7 interaction areas comprise residues of the S1′, S1, S2 and S3 subsites (Figure 7C,D). The graphs in Figure 7 that show the minimum distance between peptide and 3CL^{pro} residues reveal that all analyzed peptides formed several interactions with 3CL^{pro} active/binding site residues. The localization in the substrate-binding region of the 3CVL peptides and their D-ri counterparts, even when it does not interact with the catalytic

https://doi.org/10.1021/acschembio.2c00735
ACS Chem. Biol. 2023, 18, 315−330

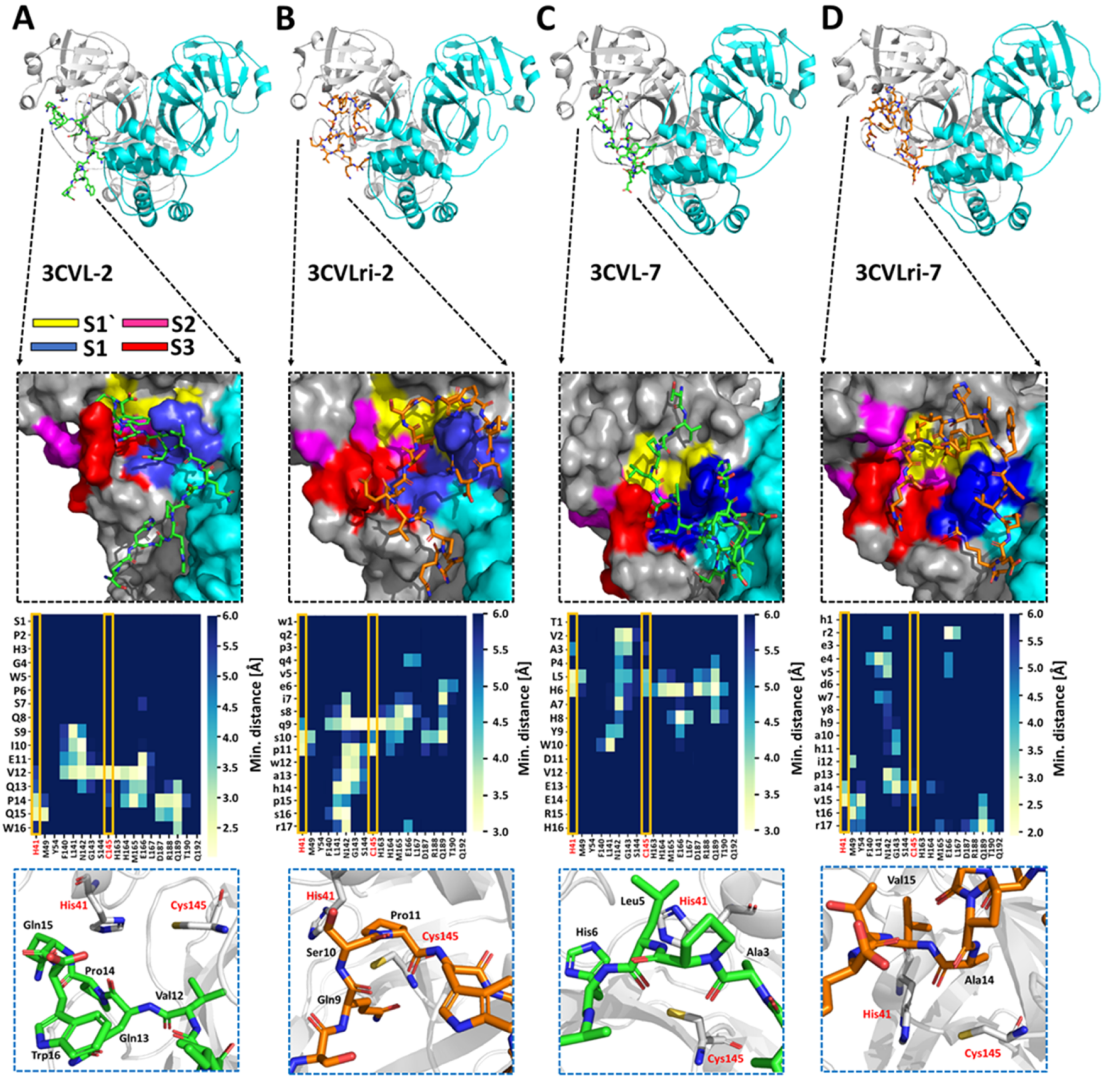


Figure 7. Comparison of the competitive 3CVL and 3CVLri peptide inhibitors binding position in the $3CL^{pro}$ active and substrate-binding site. $3CL^{pro}$ is shown in ribbon view: chain A (gray) and chain B (turquoise). The 3CVL (green) and 3CVLri (orange) peptides are shown as sticks. The surface view demonstrates a preferred position of the peptide in the substrate-binding region, and the single subsites are marked by different colors (S1′: yellow; S1: blue; S2: pink; S3: red). The distance between the $3CL^{pro}$ substrate-binding site residues and the peptides shows that active site residues His41 and Cys145 are blocked by the inhibitors: (A) 3CVL-2, (B) 3CVLri-2, (C) 3CVL-7, (D) 3CVLri-7.

dyad (His41 and Cys145), showed the potential to block the entrance of the substrate.

To investigate the impact of the competitive inhibitors 3CVL-2, 3CVLri-2, 3CVL-7, and 3CVLri-7, two-dimensional NMR 1H-15N correlation spectra of 100 μM 15N $3CL^{pro}$ were recorded in the presence of an equimolar amount of the respective peptides (1:1 ratio). Supporting Information, Figure S25A shows the reference spectrum of 15N $3CL^{pro}$ in the absence of the peptides. Spectral resonances are well resolved and display the properties of a globular protein containing both β-sheets and α-helices. Several other 15N $3CL^{pro}$ resonances appear in the presence of 3CVL-2 (orange spectrum) compared to the reference spectrum (black) (Supporting Information, Figure S25B). That could point either to a second $3CL^{pro}$ conformation becoming visible or previously invisible parts of $3CL^{pro}$ becoming more dynamic and visible in the presence of the peptide. The effect seems less intense for 3CVLri-2 (Supporting Information, Figure S25C, light blue spectrum).

The NMR spectra of 15N $3CL^{pro}$ in the presence of 3CVL-7 (Supporting Information, Figure S26B) look very similar to those for 3CVL-2; several new resonances of 15N $3CL^{pro}$ appear in the presence of 3CVL-7 and 3CVLri-7 (Supporting Information, Figure S26B,C), indicating higher conformational or overall dynamics.

Contrary, the noncompetitive inhibitors 3CVL-4 and 3CVLri-4 were docked in a suggested allosteric site at the interface of the protomers. Structural representation of the interfaces of $3CL^{pro}$ in complex with the 3CVL-4 and 3CVLri-4 peptides are shown in Figure 8A,B.

Analysis of the distance between residues of the allosteric site of each protomer demonstrated that several residues of both protomer chains interact with the protease, in particular Lys12, 97, and 100 (Figure 8).

MD simulations of 3CVL-4 and 3CVLri-4 in a proposed allosteric site showed that this interaction induces conformational changes in the protease's active site.

https://doi.org/10.1021/acschembio.2c00735
ACS Chem. Biol. 2023, 18, 315−330

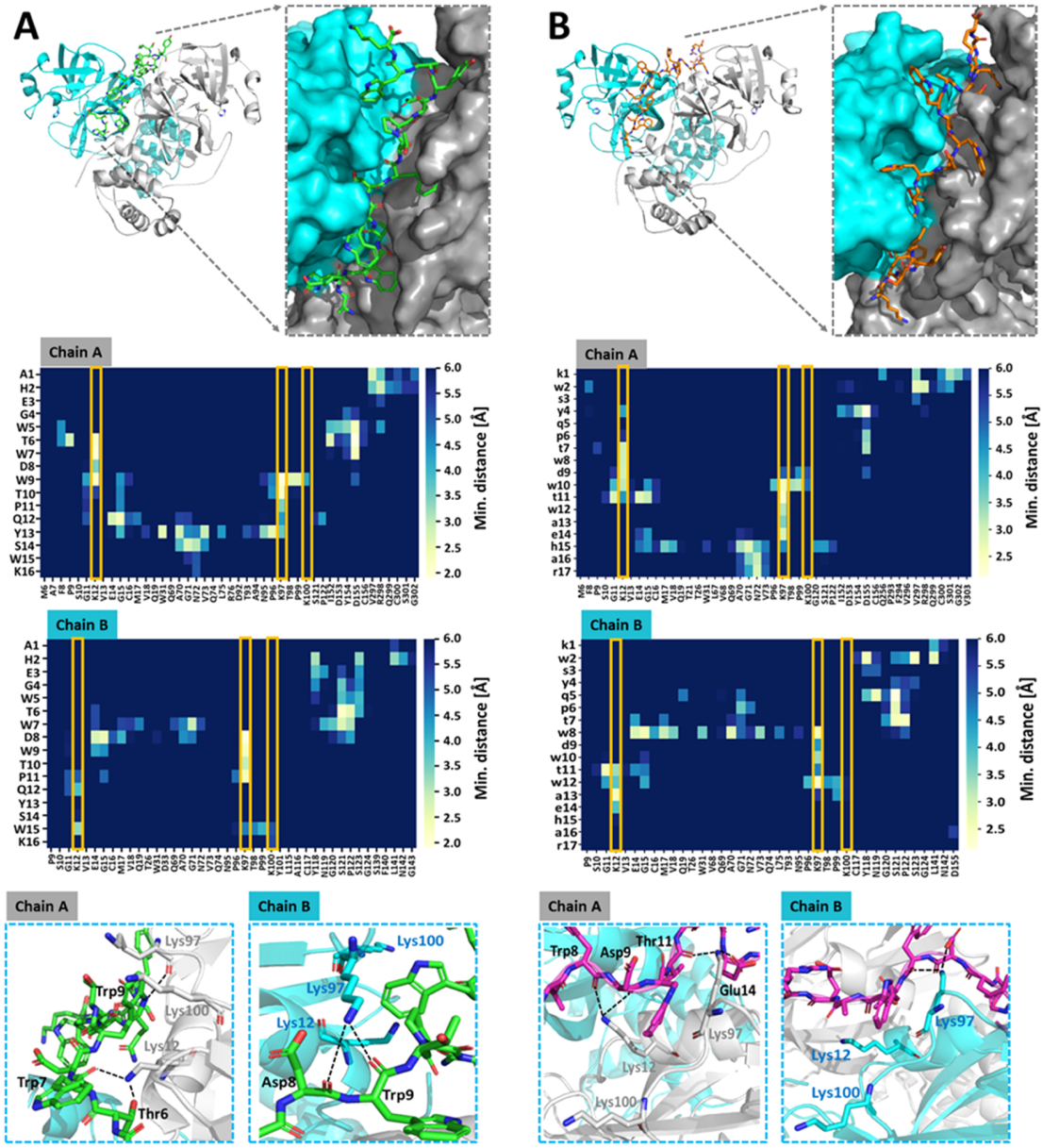


Figure 8. Comparison of the noncompetitive 3CVL-4 and 3CVLri-4 peptide inhibitors in a supposed 3CL^{pro} allosteric site. 3CL^{pro} is shown in ribbon view, and each protomer is colored differently: chain A (gray) and chain B (turquoise). 3CVL (green) and 3CVLri (orange) peptides are shown as sticks. The surface view demonstrates a preferred position of the peptide in the proposed allosteric site in a cleft between both protease protomers: (A) 3CVL-4 and (B) 3CVLri-4.

Determination of B-chain active site volume demonstrates a reduction compared to the control simulation due to the binding of 3CVL-4 and 3CVLri-4, respectively (Figure 9A−C and Supporting Information, Figure S27).

To determine the contribution of specific residues of the 3CVL and 3CVLri peptides to the interaction with 3CL^{pro} we performed a mutational analysis using alanine scanning. Alanine mutations that result in positive $\Delta\Delta H$ values relative to the parent structure suggest loss of binding, which indicates that the mutated amino acid was involved in an overall favorable network of electrostatic interactions. Mutations that result in negative $\Delta\Delta H$ suggest gain of binding, which indicates that the mutated amino acid was involved in an overall unfavorable network of electrostatic interactions.

In the case of 3CVLri-2 and its parental L-peptide Gln9 and Gln8, respectively, are important for the interaction with 3CL^{pro}. The Ser8 residue (Ser7 in the L-peptide) situated at the P2 position seems to have a weak effect on the interaction with 3CL^{pro} (Supporting Information, Figure S28A). The glutamine residues are in the P1 position of the peptides, as in the substrate already described for the protease, the P2 site has a preference for hydrophobic residues (Supporting Information, Figure S6). Important residues in the interaction between 3CVLri-7 (3CVL-7) and 3CL^{pro} are His9 (His6) and His11 (His8), unfavorable residues are Asp6 (Asp11) and Ala10 (Ala7) (Supporting Information, Figure S28A). In the case of the noncompetitive inhibitors, 3CVLri-4 (3CVL-4), Trp8 (Trp9), and Tyr4 (Tyr13) are mainly involved in the interaction with the virus protease. Interestingly, Asp8 in the mother L-peptide

https://doi.org/10.1021/acschembio.2c00735
ACS Chem. Biol. 2023, 18, 315−330

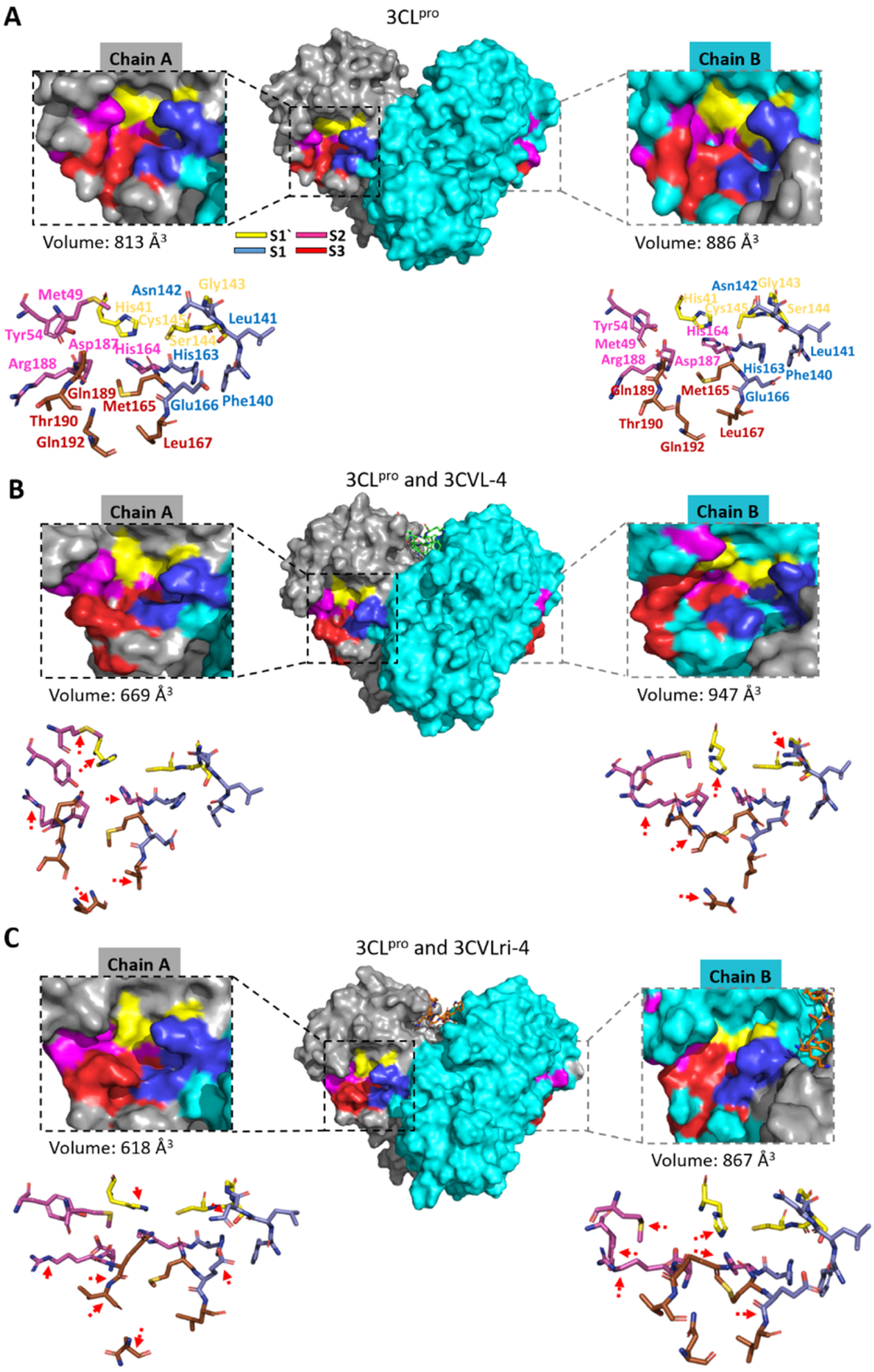


Figure 9. Comparison of the active site volumes of control and after binding of the noncompetitive 3CVL-4 and 3CVLri-4 peptide inhibitors in a supposed $3CL^{pro}$ allosteric site. $3CL^{pro}$ is shown in surface view, and each protomer is colored differently: chain A (gray) and chain B (turquoise). The substrate-binding subsites are colored differently. The active sites and substrate-binding sites of each protomer are shown enlarged with the corresponding volume in $Å^3$. Amino acids forming the active site and substrate-binding site are shown in sticks, red arrows demonstrate the change of the position of the residues because of the conformational change induced by the noncompetitive inhibitors: (A) $3CL^{pro}$, (B) $3CL^{pro}$ and 3CVL-4, (C) $3CL^{pro}$ and 3CVLri-4.

(Supporting Information, Figure S28B) is unfavorable for the interaction, but the corresponding residue in the D-ri peptide is involved in the interaction with $3CL^{pro}$.

3. DISCUSSION AND CONCLUSIONS

Even with the development of several vaccines against SARS-CoV-2, the development of suitable therapeutics against COVID-19 is important since no causative therapy is known.

https://doi.org/10.1021/acschembio.2c00735
ACS Chem. Biol. 2023, 18, 315−330

Instead of inhibiting the interaction of the spike protein with the cell receptors,[44,45] we intend to target one of the most important proteases in virus replication, the 3CL protease, by D-enantiomeric peptide ligands preventing the cleavage function of $3CL^{pro}$. To this end, a phage display selection was performed. The identified L-peptides showed a strong inhibitory effect in the enzyme activity assay, with IC_{50} values varying between 0.9 and 9 μM. As we demonstrated experimentally, 3CVL-4 is a noncompetitive inhibitor with an IC_{50} value of 2.4 μM, which was maintained in its D-enantiomer retro-inverso form; however, with an increased IC_{50} value of about 2-fold (5.4 μM). A 1:1 molar combination of the competitive and noncompetitive D-ri-peptides reduced the IC_{50} values remarkably. 3CVLri-2 + 3CVLri-4 combination reduces the IC_{50} value to 0.9 μM and the combination of 3CVLri-7 + 3CVLri-4 to 0.4 μM. It was clearly demonstrated that a combination of competitive and non-competitive inhibitors improved the inhibitory potential of all D-peptides. Similarly, in combination with Rupintrivir, a weak SARS-CoV-2 $3CL^{pro}$ competitive inhibitor ($IC_{50} > 100$ μM), the inhibitory effect improved after the combination with the noncompetitive 3CVLri-4 (IC_{50} of 15.10 $\pm$ 1.60 μM). The cooperative inhibition of SARS-CoV-2 $3CL^{pro}$ by competitive and noncompetitive inhibitors was also observed for Quinacrine/Suramin.[42] The structural analysis of $3CL^{pro}$ in complex with the studied peptides by docking followed by molecular dynamics simulation predicted a stable conformation for each complex, indicating significant intermolecular interaction. Of note, when not directly interacting with the catalytic dyad, amino acid residues of the peptide are located close to the pocket entry, blocking the substrate entrance. The predicted structures of $3CL^{pro}$ in complex with 3CVLri-2 indicate the occurrence of significant intermolecular interaction between the D-glutamine of the all-D-ri peptide and amino acid residues of the S1 subpocket, which explains the strong preference for D-glutamine at P1.[46]

A combination of competitive and noncompetitive D-ri-peptides (3CVLri-2/3CVLri-4 and 3CVLri-7/3CVLri-4) increased the inhibitory effect against $3CL^{pro}$.

The all-D-peptides 3CVLri-2 and -7 showed a remarkable resistance against metabolization over 8 h.

Two-dimensional 1H-15N TROSY-NMR correlation spectra of 15N $3CL^{pro}$ with 3CVL-2/3CVLri-2 and 3CVL-7/3CVLri-7 demonstrated the appearance of new and similar resonances, which may indicate the same binding region for the peptides.

Our study yielded all-D-peptides with a promising potential for further development toward therapeutic use against SARS-CoV-2 $3CL^{pro}$.

4. MATERIALS AND METHODS

4.1. Peptides. All peptides were purchased from Caslo (Lyngby, Denmark) as a lyophilized powder with >98% purity. Peptides 3CVL 1−10 consist of 16 amino acid residues in L-configuration with amidated C-termini. The D-retro-inverso (D-ri) peptides 3CVLri-2, 3CVLri-4, and 3CVLri-7 have identical amino acid sequences, except for glycine, which was replaced by alanine in the case of 3CVLri-2 and 3CVLri-4. All amino acid residues in D-configuration and with a reversal peptide bond. Additionally, an arginine was added to the C-termini of each ri-peptide resulting in 17 amino acid residues. The C-termini of the D-ri-peptides are amidated, and the N-termini are acetylated. An analysis certificate from Caslo (Lyngby, Denmark) demonstrated the purity of each peptide (Supporting Information, Figures S29−S32 and Table S5). High-performance liquid chromatography (HPLC) analysis was performed with a C18 column. For the mass spectrometry certificate, the matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) method was used.

4.2. Protein Expression and Purification. SARS-CoV-2 $3CL^{pro}$ (Uniprot entry: P0DTD1, virus strain: hCoV-19/Wuhan/WIV04/2019) was cloned, expressed, and purified as described previously.[41]

4.3. Phage Display. Two independent phage display selections (selection A and B), each with three rounds, were performed to obtain peptides that bind $3CL^{pro}$. Therefore, in both selections, 100 μg/mL purified $3CL^{pro}$ diluted in 100 μL of 20 mM potassium phosphate buffer pH 7.5 with 150 mM sodium chloride were immobilized on an amino plate (Nunc Amino Immobilizer 96-well plates, polystyrene; Thermo Scientific, Waltham). After 30 min at room temperature (RT), the $3CL^{pro}$ solution was removed, and the surface was quenched with 1 M ethanolamine pH 8.5 (Cytiva; Chicago) for 1 h at RT. Additionally, the surface was blocked with 200 μL of 10 mg/mL BSA in 20 mM potassium phosphate buffer pH 7.5 with 150 mM sodium chloride in the first round and with 200 μL of 10 mg/mL milk powder in 20 mM potassium phosphate buffer pH 7.5 with 150 mM sodium chloride in the second round to reduce unspecific binding events. In the third round, the surface was not additionally blocked to vary the selection conditions. The surface was washed six times with 200 μL of washing buffer (20 mM potassium phosphate buffer pH 7.5 with 150 mM sodium chloride and 0.05% tween-20). During the first and second rounds, 2 mg/mL BSA or milk powder was added to the washing buffer, respectively. Then, 7.2×10^{10} phages (TriCo-16 phage display peptide library; Creative Biolabs, New York) in 100 μL of the corresponding washing buffer were added and incubated for 20 min at RT with the exception of the third round of selection B. Here, the incubation time was extended to 30 min. After removing of nonbinding phages, the well was washed with 200 μL of the respective washing buffer of the corresponding round. The number of washing steps was varied in each selection round (selection A: 2-5-5; selection B: 5-8-10). Elution of phages was performed by incubation with 100 μL of 0.2 M Glycine-HCl pH 2.2 for 10 min at RT. The phage-containing solution was then removed from the plate and neutralized by transferring it to a tube containing 25 μL of 1 M Tris-HCl pH 9.1. To determine the output titer, 5 μL of eluted phages were used. Therefore, a dilution series from 10-2 to 10-8 was prepared with the eluted phages in a total volume of 100 μL of LB medium. Each dilution was mixed with 100 μL of *Escherichia coli* K12 ER2738 (OD600 of 0.6) and plated with 800 μL of top agar on plates (35 × 10 mm; Sarstedt, Nümbrecht, Germany) containing LB-Agar-IPTG-XGal. After overnight incubation at 37 °C, the plaques were counted to determine the output titer.

The remaining eluted phages (120 μL) were amplified in 20 mL of *E. coli* K12 ER2738, starting with an OD_{600} of 0.1 for 3.5 h at 37 °C and 120 rpm. The culture was then centrifuged for 20 min at 2700*g* and 4 °C. The pellet was discarded, and the supernatant was incubated with 7 mL of PEG-8000/2.5 M sodium chloride overnight at 4 °C for phage precipitation. Subsequently, the solution was centrifuged for 1 h at 4 °C, and 2700*g*. The phage-containing pellet was dissolved in 1 mL of 1 × phosphate-buffered saline (PBS) before another centrifugation step at 4 °C and 11,000*g* for 5 min to get rid of residual bacterial components. Afterward, the supernatant was added to 200 μL of PEG-8000/2.5 M sodium chloride and incubated for 1 h on ice, followed by final centrifugation for 45 min at 2700*g* and 4 °C. The phage-containing pellet was properly resuspended in 100 μL of 1 × PBS. The input titer was determined by spectrophotometry[40] in 1 × PBS using a 1:10 dilution. The resulting phages were then used in the next selection round with the same phage amount as before and further used for an enrichment enzyme-linked immunosorbent assay (ELISA) and prepared for next-generation sequencing (NGS) analysis.

Besides the primary selection with $3CL^{pro}$ as the target (Target Selection, TS), two control selections were performed as described before[46] to enable an efficient evaluation of the selection success and to facilitate further the identification of target-binding peptides over those peptides that have accumulated, e.g., because of an affinity to the surface. The first control selection, called Empty Selection (ES), was performed in the same way as TS but without $3CL^{pro}$. During immobilization, 20 mM potassium phosphate buffer pH 7.5 with 150 mM sodium chloride was used instead of $3CL^{pro}$. One further control,

https://doi.org/10.1021/acschembio.2c00735
ACS Chem. Biol. 2023, 18, 315−330

Direct Control (DC), was performed starting from the second round without $3CL^{pro}$, but in contrast to ES, the phages resulting from each round of TS were used for the following selection rounds.

4.4. Enrichment ELISA. An enrichment ELISA was performed to validate the success of the phage display selection. Therefore, 100 μg/mL $3CL^{pro}$ diluted in 100 μL of 20 mM potassium phosphate buffer pH 7.5 with 150 mM sodium chloride was immobilized on an amino plate (Nunc Immobilizer Amino 96-well plate, polystyrene; Thermo Scientific, Waltham) for 30 min at RT. Noncoated wells (20 mM potassium phosphate buffer pH 7.5 with 150 mM sodium chloride without $3CL^{pro}$) were used as a control for each selection round. The solution was removed, and the surface was quenched with 200 μL of 1 M ethanolamine pH 8.5 (Cytiva; Chicago) for 1 h at RT, followed by a blocking step with 200 μL of blocking solution (10 mg/mL BSA in 20 mM potassium phosphate buffer pH 7.5 with 150 mM sodium chloride) for 15 min at RT. After three washing steps with 200 μL of washing buffer (150 mM sodium chloride, 0.05% Tween-20, 2 mg/mL BSA in 20 mM potassium phosphate buffer pH 7.5), the amplified and purified phages of each target selection round were diluted in washing buffer to a total amount of 5×10^{10} phages in 100 μL. The wells immobilized with $3CL^{pro}$ and noncoated wells were incubated with the corresponding phage solution for 1 h at RT. Unbound phages were removed by five washing steps with 150 μL of washing buffer. The anti-M13 antibody (HRP conjugated mouse monoclonal antibacteriophage M13 antibody; Sino Biological, Peking, China) was diluted in washing buffer to a final concentration of 0.35 ng/μL, and 100 μL was added to each well for 1 h at RT. After 6 washing steps with 150 μL of washing buffer, the supernatant was removed entirely, and 100 μL of the 3,3′,5,5′-tetramethylbenzidine (TMB) solution (TMB was previously dissolved in 1 mL of dimethyl sulfoxide (DMSO) and diluted with 9 mL of 0.05 M phosphate citrate buffer pH 5) were added into each well. After 2.5 min incubation at RT, the reaction was stopped with 100 μL of 2 M H_2SO_4. The absorption was quantified at 450 nm using a microplate reader (BMG Labtech, Ortenberg, Germany).

In addition to the samples from the target selection (TS1, TS2, TS3), the library and the wild-type phage without peptides were analyzed. The measurement was performed in single determination to save protein samples. Noncoated wells that were incubated with washing buffer without phages were used as a control for the background signal.

4.5. Sequence Analysis. Extraction and the amplification of the single-stranded phage DNA from each sample from the phage display selection were purified and prepared for NGS analysis as described previously.[46] NGS analysis was performed by the Biologisch-Medizinisches Forschungszentrum at the Heinrich Heine University Düsseldorf, Germany. The resulting sequences were further processed and evaluated with the targeted sequencing analysis tool (TSAT) and the open-source software Hammock for subsequent sequence alignment.[47,48]

4.6. SARS-CoV-2 $3CL^{pro}$ Activity Assay. SARS-CoV-2 $3CL^{pro}$ activity assay was performed as described previously.[42,49−51] A fluorogenic substrate DABCYL-KTSAVLQ↓SGFRKME-EDANS (Bachem, Bubendorf, Switzerland) was used, and the assay buffer contained 20 mM Tris pH 7.2, 200 mM NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA), and 1 mM tris(2-carboxyethyl)-phosphine (TCEP). The reaction mixture was pipetted in a Corning 96-Well plate (Merck, Darmstadt, Germany) consisting of 0.5 μM $3CL^{pro}$. The assay was initiated with the addition of the substrate at a final concentration of 50 μM. The fluorescence intensities were measured at 60 s intervals over 30 min using an Infinite 200 PRO plate reader (Tecan, Männedorf, Switzerland), and the temperature was set to 37 °C. The excitation and emission wavelengths were 360 and 460 nm, respectively.

4.7. SARS-CoV-2 $3CL^{pro}$ Inhibition Assay. Inhibition of SARS-CoV-2 $3CL^{pro}$ activity by 3CVL-1 to 10 and 3CVLri-2, -4, and -7 was investigated using the described activity assay. The L-Peptides (10 μM) were used for a preliminary screening test. For the final inhibition assays, 0.5 μM protein was incubated with 0−140 μM 3CVL-2, 0−120 μM 3CVL-4, and 0−100 μM 3CVL-7. The corresponding 3CVLri peptides were incubated with the same inhibitor concentrations, except for 3CVLri-4 (0−140 μM). The mixtures were incubated for 30 min at RT. When the substrate with a final concentration of 50 μM was added to the mixture, the fluorescence intensities were measured at 60 s intervals over 30 min using an Infinite 200 PRO plate reader (Tecan, Männedorf, Switzerland). The temperature was set to 37 °C, and the excitation and emission wavelengths were 360 and 460 nm, respectively. Inhibition assays were performed as triplicates.

For the 3CVLri-2 + 3CVLri-4 and 3CVLri-4 + 3CVLri-7 combination tests, a 1:1 stock solution of the molecules was prepared. Protein (0.5 μM) was incubated with 0−75 μM (3CVLri-2 + 3CVLri-4) and 0−50 μM (3CVLri-4 + 3CVLri-7) of the combined molecules. The IC_{50} values were calculated by plotting the initial velocity against various concentrations of the peptides using a dose−response curve in GraphPad Prism software version 8 (San Diego, CA). Data are presented as mean ± SD.

4.8. Determination of Inhibition Mode. The inhibition mode was determined using different final concentrations of the inhibitors and substrate. Briefly, 0.5 μM SARS-CoV-2 $3CL^{pro}$ was incubated with the inhibitor, in various concentrations, for 30 min at RT. Subsequently, the reaction was initiated by adding the corresponding concentration series of the substrate. The data were analyzed using a Lineweaver−Burk plot; therefore, the reciprocal of velocity (1/V) vs the reciprocal of the substrate concentration (1/[S]) was compared.[52,53] All measurements were performed in triplicate, and data are presented as mean ± SD.

4.9. Inhibitor Stability over 24 h. Stable inhibition of SARS-CoV-2 $3CL^{pro}$ by 3CVLri-2, -4, and -7 was followed via a 24 h inhibition experiment. Briefly, 0.5 μM SARS-CoV-2 $3CL^{pro}$ was incubated with 5 μM (D-ri peptides 2 and 4) and 20 μM (D-ri peptide 7) and incubated for 1/2, 1, 2, 3, 4, 5, and 24 h at RT. The control was performed with $3CL^{pro}$ without the peptides and measured together after each time point. Subsequently, the reaction was initiated by the addition of the substrate. All measurements were performed in triplicate, and data are presented as mean ± SD.

4.10. Assay to Exclude D-ri Peptides as Promiscuous Inhibitors. A detergent-based control assay was performed to exclude inhibitors that possibly act as promiscuous aggregators of the $3CL^{pro}$ by adding 0.001, 0.01, and 0.1% of Triton X-100 to the reaction.[54] Four concentrations of 3CVLri-2 (1, 5, 10, and 20 μM), 3CVLri-4 (1, 5, 10, and 20 μM), and 3CVLri-7 (0.25, 0.5, 1, and 5 μM) were tested. All measurements were performed in triplicate, and data are presented as mean ± SD.

4.11. Circular Dichroism Spectroscopy. Circular dichroism (CD) measurements were carried out with a Jasco J-1100 spectropolarimeter (Jasco, Germany). Far-UV spectra were measured between 190 to 260 nm using 5 μM $3CL^{pro}$ in 20 mM K_2HPO_4/KH_2PO_4 pH 7.4 and a single peptide concentration of 30 μM in H_2O. The secondary structure of 3CVL-2, 3CVL-4, 3CVL-7, and the related D-ri peptides was checked. A 1 mm path length cell was used for the measurements; 15 repeat scans were obtained for each sample, and five scans were conducted to establish the respective baselines. The averaged baseline spectrum was subtracted from the averaged sample spectrum. The results are presented as molar ellipticity $[\theta]$, according to eq 1

$$[\theta]\lambda = \theta/(c^{*}0.001^{*}1^{*}n) \quad (1)$$

where θ is the ellipticity measured at the wavelength λ (deg), c is the peptide concentration (mol/L), 0.001 is the cell path length (cm), and n is the number of amino acids.

4.12. Surface Plasmon Resonance Measurements. Surface plasmon resonance (SPR) measurements with 3CVL peptides and SARS-CoV-2 $3CL^{pro}$ were performed using a Biacore 8K device (GE Healthcare, Uppsala, Sweden). Recombinant 3CLpro was immobilized on a CM5 sensor surface (GE Healthcare, Uppsala, Sweden) via covalent coupling of primary amino groups. The protease was diluted to a final concentration of 15 μg/mL in 10 mM NaAc pH 5.5 and injected for 600 s at 10 μL/min to reach a final immobilization level of 2700 to 3000 RU. After quenching with Ehanolamine at pH 8.5, kinetic experiments were performed as parallel injections on all eight flow channels. 3CVL and 3CVLri peptides were diluted in 20 mM Tris pH 7.4, 100 mM NaCl, 1 mM EDTA and 0.05% Tween-20 in the

https://doi.org/10.1021/acschembio.2c00735
ACS Chem. Biol. 2023, 18, 315−330

concentration range of 0.03−50 μM with a 1:1 dilution series and injected at 30 μL/min. Surface regeneration was performed after dissociation ended for 2 × 45 s at 30 μL/min using an acidic regeneration solution consisting of 150 mM oxalic acid, 150 mM phosphoric acid, 150 mM formic acid, and 150 mM malonic acid at pH 5.0. Data evaluation was performed using a heterogeneous ligand fit model for the competitive inhibitors with Biacore Insight Evaluation Software v3.0.12 (GE Healthcare, Uppsala, Sweden). The heterogeneous ligand model assumes analyte binding at two independent ligand sites. Each ligand site binds the analyte independently and with a different rate constant. A global 1:1 kinetic fit model was used for the noncompetitive inhibitor.

4.13. Nuclear Magnetic Resonance. Two-dimensional 1H-15N TROSY-NMR correlation spectra[55] of 100 μM 15N 3CL^{pro} dissolved in 20 mM MES pH 6.5 buffer containing 150 mM NaCl were recorded in the presence and absence of 3CVL-2 and 3CVLri-2 as well as 3CVL-7 and 3CVLri-7 peptides. Experiments for 3CVL-4 and 3CVLri-4 could not be performed because of the peptide's instability at pH 6.5. Spectra were recorded on a Bruker 1.2 GHz AVANCE neo spectrometer equipped with a triple-resonance 1H,15N,13C cryogenic probe. The experimental temperature was 25 °C. Spectral dimensions were 16.02 ppm (1H) × 32.89 ppm (15N), with 1024 complex points recorded in the 1H dimension and 64 complex points recorded in the 15N dimension, and the Echo/Anti-echo scheme used for quadrature detection, resulting in an acquisition time of 53.2 ms in the 1H dimension and 16 ms in the 15N dimension. The carrier was positioned at 4.7 ppm in the 1H dimension and 117 ppm in the 15N dimension; 512 scans were recorded for each increment, with a recovery delay of 1 s between scans, resulting in an overall experimental time of 19.7 h per spectrum.

To study the interaction between 3CL^{pro} and the various peptides, only the protease was 15N-labeled and, therefore, NMR-visible, while the peptides were unlabeled and invisible in the 1H-15N correlation spectra. For the interaction study, 100 μM 15N-labeled 3CL^{pro} was mixed with an equimolar amount of the respective peptide. Spectra were overlayed and compared to the reference spectrum of 3CL^{pro} (equal final concentration, in the absence of the peptide). Spectra analysis was performed using the software Bruker Topspin 4.1.3. and CCPN Analysis 3.0.[56]

4.14. Metabolic Stability. Metabolic stability experiments of the 3CVLri peptides were performed as described before.[57,58] Solutions simulating gastric and intestinal fluid (SGF and SIF) were prepared according to the European Pharmacopoeia 7.0. Human plasma samples were purchased from Innovative Research (Novi, Michigan), and human liver microsomes were purchased from Sekisui XenoTech (Kansas City; H1000). The liver microsomes were diluted in an NADPH regenerating system (NRS). For the stability tests, 150 μM 3CVLri peptides were incubated in SGF, SIF, human plasma, and human liver microsomes in triplicate at 37 °C with slight shaking for different periods. The peptides were extracted by precipitating the proteins with a solution containing acetonitrile and 0.5% formic acid. Afterward, the mixture was mixed and heated for 5 min at 95 °C. The sample was subsequently centrifuged at 14.000g for 10 min at 4 °C. The extracted peptides are in the supernatant and were analyzed by reversed-phase high-performance liquid chromatography (RP-HPLC). The RP-HPLC system (Agilent Technologies, Santa Clara; 1260 series) consisted of an autosampler, quaternary pump, a thermostatted column compartment, and a variable-wavelength detector. Chromatography was performed with a C18 column (Agilent Technologies, Santa Clara; ZORBAX 300SB-C18 5 μm, 4.6 × 250 mm) at 25 °C and 214 nm with a flow rate of 1 mL/min. The sample injection volume was 20 μL. Chromatograms were recorded and analyzed with the Agilent software OpenLab version 2−5. Mobile phases were acetonitrile + 0.1% trifluoroacetic acid (TFA) (A) and water + 0.1% TFA (B) for sample analysis from stability tests. The samples were measured with an initial isocratic step at 15% solvent A for 3 min followed by a gradient elution to 45% A in 15 min. All measurements were performed in triplicate, and data are presented as mean ± SD.

4.15. Statistical Analysis. All data are expressed as the mean ± standard deviations (SDs). The statistical significance of the mean values' differences was assessed with one-way analyses of variance (ANOVA), followed by Tukeys' multiple comparison test. Significant differences were considered at $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***). All statistical analyses were performed with GraphPad Prism software version 8 (San Diego, CA).

4.16. Docking and Molecular Dynamics. *4.16.1. Peptide and Protein Structure Preparation.* The peptide models were constructed as linear chains using the python module PeptideBuilder.[59] To obtain D-retro-inverso (D-ri) peptides, the peptides were first generated as L-retro-inverso molecules, followed by the inversion of their stereochemical configuration by flipping the structure files with cartesian coordinates along the x-axis. The sequences of the peptides are shown in Table 5.

Table 5. Sequences of 3CVL and 3CVLri Peptide Models

peptide	sequence
3CVL-2	SPHGWPSQSIEVQPQW
3CVL-4	AHEGWTWDWTPQYSWK
3CVL-7	TVAPLHAHYWDVEERH
3CVLri-2	wqpqveisqspwahpsr
3CVLri-4	kwsyqptwdwtwaehar
3CVLri-7	hreevdwyhahlpavtr

The initial structural data of the 3CL^{pro} dimer were obtained from a crystal structure (PDB: 6M2N).

4.16.2. Docking of the Peptides against the 3CL^{pro} Dimer. 3CVL-2, 3CVL-7, 3CVLri-2, and 3CVLri-7 were docked against the active site of chain A of the 3CL^{pro} dimer (PDB: 6M2N), which can be defined by the residues His41, Met49, Tyr54, Phe140, Leu141, Asn142, Gly143, Ser144, Cys145, His163, His164, Met165, Glu166, Leu167, Asp187, Arg188, Gln189, Thr190, and Gln192.[60]

3CVL-4 and 3CVLri-4 bind 3CL^{pro} allosterically, and both peptides were docked against the potential binding site residues Lys12, Lys97, Lys100, Tyr101, Lys102, Phe103, Val104, and Arg105.[36] Autodock Tools[61,62] were used to generate a docking grid around the positions of the binding site residues of the 3CL^{pro} dimer. The grid was centered at x,y,z-position of −33.572, −62.854, 40.9 (Active site) and −49.16, −47.323, 38.911 (Allosteric binding site). After adding hydrogen atoms to the 3CL^{pro} dimer and the peptides, Gasteiger partial charges were computed and added to the dimer and the peptides using AutoDock Tool. The protein and peptide models were saved in the PDBQT format. The docking was then performed using AutoDock Vina,[43] which treats the ligand as flexible while keeping the receptor rigid.

4.16.3. Simulation Setup and Production Run. All MD simulations and analyses were performed with GROMACS 2021.4.[63] AMBER14SB[64] was used as a force field, and the TIP3P water model[65] for explicit water simulation. The peptides and the protein−peptide complexes were centered in a dodecahedron box with 240.8−370.5 and 872.5 nm^3, solvated with water, and neutralized with Na^+ and Cl^- ions resulting in a system size of ∼23,500−36,000 and ∼86,000 atoms, respectively.

The systems' energy was minimized using the steepest descent algorithm.[66] The systems were then equilibrated in two steps. First, a 0.1 ns simulation was performed in the NVT ensemble, and second, a 1 ns simulation in the NpT ensemble at 310 K—velocity-rescaling thermostat[67] and 1.0 bar—Berendsen barostat.[68] First, 1 μs peptide MD simulations were performed. After the peptides were docked against the 3CL^{pro} dimer, we performed 100 ns of protein−peptide complexes with the temperature set to 310 K (Nosé−Hoover thermostat)[69] and the pressure set to 1 bar (Parinello−Rahman barostat).[70] One selected simulation of each protein−peptide complex was then extended to 200 ns and ran in triplicates.

Electrostatic interactions were processed with the particle-mesh Ewald method,[71,72] combined with periodic boundary conditions and a real-space cutoff of 12 Å. The same cutoff was applied to Lennard-Jones (LJ) interactions. The leapfrog stochastic dynamics integrator was used with a time step of 2 fs to integrate the equations of motion. During the

https://doi.org/10.1021/acschembio.2c00735
ACS Chem. Biol. 2023, 18, 315−330

MD simulations, all bond lengths were constrained using the LINCS algorithm.[73]

4.16.4. Molecular Dynamics Analysis. The peptide's root means square deviation (RMSD) indicates how flexible the ligand is within the binding site. The protein structures sampled during the MD simulations were aligned to the MD starting structure, and then the RMSD was calculated for the ligand using *gmx rms*.

To investigate the effect of the allosteric peptides and the occupation of the peptides that bind to the active site, we used the VMD plugin Epock 1.0.5 to evaluate the pocket volume of both active sites of the dimer for 1000 snapshots of the whole MD simulation after fitting the protein to the starting structure.[74] The following simulations were conducted for the second half of the MD simulations: The root-mean-square fluctuation (RMSF) was determined for each peptide with *gmx rmsf* to identify the flexible and rigid areas of the peptide within the protein−peptide complex, and the average minimum distance between specific residues of the protein and peptide was determined using *gmx mindist*.

We used the MM/PBSA method as implemented in gmx_mmpbsa to calculate the binding free energy ΔG_{bind} between the protein and the peptide and to perform alanine scanning.[75] The linear PB equation was solved using the level set function to build the dielectric interface between the solvent and solute.[76] The ionic strength was set to 150 mM. The total nonpolar solvation free energy was modeled as a single term linearly proportional to the solvent-accessible surface area.[77] Here, the binding free energy is defined as

$$\Delta G_{bind} = \langle G_{complex} \rangle - \langle G_{receptor} \rangle - \langle G_{ligand} \rangle$$

with $\langle\rangle$ indicating the average of over 500 snapshots. The free energy for each entity is given as

$$\langle G_x \rangle = \langle E_{bonded} \rangle + \langle E_{nonbonded} \rangle + \langle G_{polar} \rangle + \langle G_{nonpolar} \rangle - \langle TS \rangle$$

where E_{bonded} describes the bonded interactions, $E_{nonbonded}$ indicates the nonbonded, i.e., the Coulomb and Lennard-Jones interactions, and G_{polar} and $G_{nonpolar}$ are the polar and nonpolar contributions to the solvation free energy, respectively. The TS term includes the absolute temperature, T, and the entropy, S.

ΔG_{bind} can also be stated as

$$\Delta G_{bind} = \Delta H - T\Delta S$$

with ΔH the enthalpy of binding and $-T\Delta S$ the conformational entropy after ligand binding. To determine the entropic component, gmx_mmpbsa provides different methods, such as quasi-harmonic (QH) approximation, interaction entropy (IE), or C2 entropy approach. However, these methods were inappropriate for our system, as they either required a large number of snapshots ($\geq 3N$, N = number of complex atoms) or a standard deviation of the interaction energy (sIE) of <3.6 kcal/mol to obtain realistic entropies. Thus, we decided to omit the entropic term, as ΔH is usually sufficient to compare a set of similar ligands by their relative binding free energies.[75] To elucidate the contribution of each peptide residue to the binding free energy, we performed alanine scanning using gmx_mmpbsa. Each residue of the 3CVL and 3CVL-ri peptides (Except Gly and Pro) was substituted by Ala and the enthalpy of binding difference $\Delta\Delta H^{X->A}$ upon alanine mutation was calculated in the following way

$$\Delta\Delta H^{X->A} = \Delta H^{A} - \Delta H^{X}$$

To identify the peptide structures with the most populated configuration or binding pose in the protein−peptide complex, the algorithm of Daura et al. was used as implemented in GROMACS (*gmx cluster*).[78]

The cutoff for clustering the peptide configurations was set to 3.5 Å, and the average structures of the 10 most populated clusters were used for docking against the dimer of the main protease. The binding poses of the peptides in the elongated protein−peptide complex simulations were clustered using a cutoff of 3.0 Å. To avoid biasing of nonbinding flexible areas of the peptide, only those peptide residues with an RMSF of <6 Å were considered for clustering. The three most prominent clusters were then analyzed for interacting residues by generating a contact map.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acschembio.2c00735.

Additional experiments and supplementary figures (phage display outcome, inhibitor tests, SPR analysis, HPLC profiles, NMR spectra, and MD analysis) and tables (analysis of peptide sequences selected by phage display and list of amino acids forming the $3CL^{pro}$ substrate-binding site and allosteric site) (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Mônika A. Coronado − *Institute of Biological Information Processing, Structural Biochemistry (IBI-7), Forschungszentrum Jülich, 52425 Jülich, Germany*; Email: m.coronado@fz-juelich.de

Dieter Willbold − *Institute of Biological Information Processing, Structural Biochemistry (IBI-7), Forschungszentrum Jülich, 52425 Jülich, Germany; Institut für Physikalische Biologie, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany*; orcid.org/0000-0002-0065-7366; Email: d.willbold@fz-juelich.de

Authors

Raphael J. Eberle − *Institute of Biological Information Processing, Structural Biochemistry (IBI-7), Forschungszentrum Jülich, 52425 Jülich, Germany; Institut für Physikalische Biologie, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany*; orcid.org/0000-0002-8763-3884

Marc Sevenich − *Institute of Biological Information Processing, Structural Biochemistry (IBI-7), Forschungszentrum Jülich, 52425 Jülich, Germany; Institut für Physikalische Biologie, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany; Priavoid GmbH, 40225 Düsseldorf, Germany*

Ian Gering − *Institute of Biological Information Processing, Structural Biochemistry (IBI-7), Forschungszentrum Jülich, 52425 Jülich, Germany*

Lara Scharbert − *Institute of Biological Information Processing, Structural Biochemistry (IBI-7), Forschungszentrum Jülich, 52425 Jülich, Germany; Institut für Physikalische Biologie, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany*

Birgit Strodel − *Institute of Biological Information Processing, Structural Biochemistry (IBI-7), Forschungszentrum Jülich, 52425 Jülich, Germany; Institut für Physikalische Biologie, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany*; orcid.org/0000-0002-8734-7765

Nils A. Lakomek − *Institute of Biological Information Processing, Structural Biochemistry (IBI-7), Forschungszentrum Jülich, 52425 Jülich, Germany; Institut für Physikalische Biologie, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany*

Karoline Santur − *Institute of Biological Information Processing, Structural Biochemistry (IBI-7), Forschungszentrum Jülich, 52425 Jülich, Germany; Institut für Physikalische Biologie, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany*

Jeannine Mohrlüder − *Institute of Biological Information Processing, Structural Biochemistry (IBI-7), Forschungszentrum Jülich, 52425 Jülich, Germany;* orcid.org/0000-0002-9922-2986

Complete contact information is available at: https://pubs.acs.org/10.1021/acschembio.2c00735

Author Contributions

Conceptualization: R.J.E., J.M., and M.A.C.; methodology: R.J.E., M.S., I.G., L.S., B.S., N.A.L., K.B.S., J.M., and M.A.C.; validation: R.J.E., M.S., I.G., L.S., B.S., N.A.L., K.B.S., J.M., and M.A.C.; formal analysis: R.J.E., M.S., I.G., L.S., B.S., N.A.L., K.B.S., J.M., and M.A.C.; investigation: R.J.E. and M.A.C.; resources: B.S. and D.W.; writing—original draft preparation: R.J.E. and M.A.C.; writing—review and editing: R.J.E., M.S., I.G., L.S., B.S., N.A.L., K.B.S., J.M., M.A.C., and D.W. All authors have read and agreed to the published version of the manuscript.

Funding

D.W. received funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Project-ID 267205415, SFB 1208. B.S. and L.S. received funding from the VolkswagenStiftung for that project (project number 99761). N.A.L. thanks the German Science Foundation (DFG, grant project number 433700474) for funding as well as support by the project "Virological and immunological determinants of COVID-19 pathogenesis—lessons to get prepared for future pandemics (KA1-Co-02 "COVIPA")", a grant from the Helmholtz Association's Initiative and Networking Fund.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Dedicated to Karl-Erich Jaeger for his pioneering contributions in the field of molecular enzyme technology. The authors thank the support of the Institute of Biological Information Processing (IBI-7) Forschungszentrum Jülich, Germany. L.S. and B.S. gratefully acknowledge the computing time granted by the JARA Vergabegremium and provided under project "3cldtp" on the JARA Partition part of the supercomputer JURECA at Forschungszentrum Jülich.

■ REFERENCES

(1) Hu, B.; Guo, H.; Zhou, P.; Shi, Z. L. Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* **2021**, *19*, 141−154.

(2) WHO. *Coronavirus Disease 2019 (COVID-19) Dashboard*; World Health Organization, 2022.

(3) Mathieu, E.; Ritchie, H.; Ortiz-Ospina, E.; Roser, M.; Hasell, J.; Appel, C.; Giattino, C.; Rodés-Guirao, L. A global database of COVID-19 vaccinations. *Nat. Hum. Behav.* **2021**, *5*, 947−953.

(4) Polack, F. P.; Thomas, S. J.; Kitchin, N.; Absalon, J.; Gurtman, A.; Lockhart, S.; Perez, J. L.; Marc, G. P.; Moreira, E. D.; Zerbini, C.; Bailey, R.; Swanson, K. A.; Gruber, W. C.; et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *N. Engl. J. Med.* **2020**, *383*, 2603−2615.

(5) Baden, L. R.; El Sahly, H. M.; Essink, B.; Kotloff, K.; Frey, S.; Novak, R.; Diemert, D.; Spector, S. A.; Rouphael, N.; Creech, C. B.; McGettigan, J.; Khetan, S.; Zaks, T.; et al. Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *N. Engl. J. Med.* **2021**, *384*, 403−416.

(6) Voysey, M.; Clemens, S. A. C.; Madhi, S. A.; Weckx, L. Y.; Folegatti, P. M.; Aley, P. K.; Bijker, E.; et al. Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *Lancet* **2021**, *397*, 99−111.

(7) Livingston, E. H.; Malani, P. N.; Creech, C. B. The Johnson & Johnson Vaccine for COVID-19. *JAMA* **2021**, *325*, 1575.

(8) Owen, D. R.; Allerton, C. M.; Anderson, A. S.; Aschenbrenner, L.; Avery, M.; Berritt, S.; Boras, B.; Cardin, R. D.; Carlo, A.; Coffman, K. J.; Dantonio, A.; et al. An oral SARS-CoV-2 Mpro inhibitor clinical candidate for the treatment of COVID-19. *Science* **2021**, *374*, 1586−1593.

(9) Madsen, L. W. Remdesivir for the Treatment of Covid-19-Final Report. *N. Engl. J. Med.* **2020**, *338*, 1813−1826.

(10) Hosseinzadeh, M. H.; Shamshirian, A.; Ebrahimzadeh, M. A. Dexamethasone Vs. COVID-19: An Experimental Study in Line with the Preliminary Findings of a Large Trial. *Int. J. Clin. Pract.* **2020**, *75*, No. e13943.

(11) Nangaku, M.; Kadowaki, T.; Yotsuyanagi, H.; Ohmagari, N.; Egi, M.; Sasaki, J.; Sakamoto, T.; Hasegawa, Y.; Ogura, T.; Chiba, S.; Node, K.; Suzuki, R.; Yamaguchi, Y.; Murashima, A.; Ikeda, N.; Morishita, E.; Yuzawa, K.; Moriuchi, H.; Hayakawa, S.; Nishi, D.; Irisawa, A.; Miyamoto, T.; Suzuki, H.; Sone, H.; Fujino, Y. The Japanese Medical Science Federation COVID-19 Expert Opinion English Version. *JMA J.* **2021**, *4*, 148−162.

(12) Italian Society of Infectious and Tropical Diseases (SIMIT). *Guidelines for the Teatment and Support Management of Patients with COVID-19 Coronavirus Infection*, 2nd ed.; Italian Society of Infectious and Tropical Diseases (SIMIT): Italy, 2020.

(13) Rambaut, A.; Holmes, E. C.; O'Toole, Á.; Hill, V.; McCrone, J. T.; Ruis, C.; du Plessis, L.; Pybus, O. G. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* **2020**, *5*, 1403−1407.

(14) Pillaiyar, T.; Manickam, M.; Namasivayam, V.; Hayashi, Y.; Jung, S. H. An overview of Severe Acute Respiratory Syndrome-Coronavirus (SARS-CoV) 3CL protease inhibitors: Peptidomimetics and small molecule chemotherapy. *J. Med. Chem.* **2016**, *59*, 6595−6628.

(15) Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; Duan, Y.; Yu, J.; Wang, L.; Yang, K.; Liu, F.; Jiang, R.; Yang, X.; You, T.; Liu, X.; Yang, X.; Bai, F.; Liu, H.; Liu, X.; Guddat, L. W.; Xu, W.; Xiao, G.; Qin, C.; Shi, Z.; Jiang, H.; Rao, Z.; Yang, H. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **2020**, *582*, 289−293.

(16) Rut, W.; Groborz, K.; Zhang, L.; Sun, X.; Zmudzinski, M.; Pawlik, B.; Wang, X.; Jochmans, D.; Neyts, J.; Młynarski, W.; Hilgenfeld, R.; Drag, M. SARS-CoV-2 Mpro inhibitors and activity-based probes for patient-sample imaging. *Nat. Chem. Biol.* **2021**, *17*, 222−228.

(17) Anand, K.; Ziebuhr, J.; Wadhwani, P.; Mesters, J. R.; Hilgenfeld, R. Coronavirus main proteinase (3CLpro) structure: Basis for design of anti-SARS drugs. *Science* **2003**, *300*, 1763−1767.

(18) 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*, 5172−5208.

(19) Cannalire, R.; Barreca, M. L.; Manfroni, G.; Cecchetti, V. A journey around the medicinal chemistry of Hepatitis C virus inhibitors targeting NS4B: From target to preclinical drug candidates. *J. Med. Chem.* **2016**, *59*, 16−41.

(20) Craik, D. J.; Fairlie, D. P.; Liras, S.; Price, D. The future of peptide-based drugs. *Chem. Biol. Drug Des.* **2013**, *81*, 136−147.

(21) Mikitsh, J. L.; Chacko, A. M. Pathways for small molecule delivery to the central nervous system across the blood-brain barrier. *Perspect. Med. Chem.* **2014**, *6*, 11−24.

(22) Goodwin, D.; Simerska, P.; Toth, I. Peptides as therapeutics with enhanced bioactivity. *Curr. Med. Chem.* **2012**, *19*, 4451−4461.

(23) Jiang, N.; Leithold, L. H.; Post, J.; Ziehm, T.; Mauler, J.; Gremer, L.; Cremer, M.; Schartmann, E.; Shah, N. J.; Kutzsche, J.; Langen, K. J.; et al. Preclinical pharmacokinetic studies of the tritium labelled D-enantiomeric peptide D3 developed for the treatment of Alzheimer s disease. *PLoS One* **2015**, *10*, No. e0128553.

(24) Leithold, L. H. E.; Jiang, N.; Post, J.; Ziehm, T.; Schartmann, E.; Kutzsche, J.; Shah, N. J.; Breitkreutz, J.; Langen, K. J.; Willuweit, A.; Willbold, D. Pharmacokinetic properties of a novel D-peptide developed to be therapeutically active against toxic β-amyloid oligomers. *Pharm. Res.* **2016**, *33*, 328−336.

(25) Chong, P.; Sia, C.; Tripet, B.; James, O.; Klein, M. Comparative immunological properties of enantiomeric peptides. *Lett. Pept. Sci.* **1996**, *3*, 99−106.
(26) Zhang, T.; Gering, I.; Kutzsche, J.; Nagel-Steger, L.; Willbold, D. Toward the mode of action of the clinical stage all-D-enantiomeric peptide RD2 on Aβ42 aggregation. *ACS Chem. Neurosci.* **2019**, *10*, 4800−4809.
(27) Kutzsche, J.; Jürgens, D.; Willuweit, A.; Adermann, K.; Fuchs, C.; Simons, S.; Windisch, M.; Hümpel, M.; Rossberg, W.; Wolzt, M.; Willbold, D. Safety and pharmacokinetics of the orally available antiprionic compound PRI-002: A single and multiple ascending dose phase I study. *Alzheimer's Dement.* **2020**, *6*, No. e12001.
(28) Schemmert, S.; Camargo, L. C.; Honold, D.; Gering, I.; Kutzsche, J.; Willuweit, A.; Willbold, D. In Vitro and In Vivo Efficacies of the Linear and the Cyclic Version of an All-d-Enantiomeric Peptide Developed for the Treatment of Alzheimer's Disease. *Int. J. Mol. Sci.* **2021**, *22*, 6553.
(29) Valiente, P. A.; Wen, H.; Nim, S.; Lee, J.; Kim, H. J.; Kim, J.; Perez-Riba, A.; Paudel, Y. P.; Hwang, I.; Kim, K. D.; Kim, S.; Kim, P. M. Computational Design of Potent D-Peptide Inhibitors of SARS-CoV-2. *J. Med. Chem.* **2021**, *64*, 14955−14967.
(30) Valiente, P. A.; Nim, S.; Lee, J.; Kim, S.; Kim, P. M. Targeting the Receptor-Binding Motif of SARS-CoV-2 with D-Peptides Mimicking the ACE2 Binding Helix: Lessons for Inhibiting Omicron and Future Variants of Concern. *J. Chem. Inf. Model.* **2022**, *62*, 3618−3626.
(31) Hernández González, J. E.; Eberle, R. J.; Willbold, D.; Coronado, M. A. A computer-aided approach for the discovery of D-peptides as inhibitors of SARS-CoV-2 main protease. *Front. Mol. Biosci.* **2022**, *8*, No. 816166.
(32) Eberle, R. J.; Gering, I.; Tusche, M.; Ostermann, P. N.; Müller, L.; Adams, O.; Schaal, H.; Olivier, D. S.; Amaral, M. S.; Arni, R. K.; Willbold, D.; Coronado, M. A. Design of D-Amino Acids SARS-CoV-2 Main Protease Inhibitors Using the Cationic Peptide from Rattlesnake Venom as a Scaffold. *Pharmaceuticals* **2022**, *15*, 540.
(33) Cardoso, M. H.; Cândido, E. S.; Oshiro, K. G. N.; Rezende, S. B.; Franco, O. L.*Peptide Applications in Biomedicine, Biotechnology and Bioengineering*, Koutsopoulos, S., Ed.; Woodhead Publishing, 2018.
(34) Fischer, P. M. The design, synthesis and application of stereochemical and directional peptide isomers: a critical review. *Curr. Protein Pept. Sci.* **2003**, *4*, 339−356.
(35) Li, C.; Pazgier, M.; Li, J.; Li, C.; Liu, M.; Zou, G.; Li, Z.; Chen, J.; Tarasov, S. G.; Lu, W. Y.; Lu, W. Limitations of peptide retro-inverso isomerisation in molecular mimicry. *J. Biol. Chem.* **2010**, *285*, 19572−19581.
(36) Li, C.; Zhan, C.; Zhao, L.; Chen, X.; Lu, W. Y.; Lu, W. Functional consequences of retro-inverso isomerisation of a miniature protein inhibitor of the p53−MDM2 interaction. *Bioorg. Med. Chem.* **2013**, *21*, 4045−4050.
(37) Cheng, L.; Boland, S.; Scholle, M. D.; Bardiot, D.; Marchand, A.; Chaltin, P.; Blatt, L. M.; Beigelman, L.; Symons, J. A.; Raboisson, P.; Gurard-Levin, Z. A.; Vandyck, K.; Deval, J. Dual inhibition of SARS-CoV-2 and human rhinovirus with protease inhibitors in clinical development. *Antiviral Res.* **2021**, *187*, No. 105020.
(38) Antonopoulou, I.; Sapountzaki, E.; Rova, U.; Christakopoulos, P. Inhibition of the main protease of SARS-CoV-2 (Mpro) by repurposing/designing drug-like substances and utilising nature's toolbox of bioactive compounds. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 1306−1344.
(39) Lockbaum, G. J.; Henes, M.; Lee, J. M.; Timm, J.; Nalivaika, E. A.; Thompson, P. R.; Yilmaz, K. N.; Schiffer, C. A. Pan-3C Protease Inhibitor Rupintrivir Binds SARS-CoV-2 Main Protease in a Unique Binding Mode. *Biochemistry* **2021**, *60*, 2925−2931.
(40) Chen, H.; Wei, P.; Huang, C.; Tan, L.; Liu, Y.; Lai, L. Only one protomer is active in the dimer of SARS 3C-like proteinase. *J. Biol. Chem.* **2006**, *281*, 13894−13898.
(41) Kidera, A.; Moritsugu, K.; Ekimoto, T.; Ikeguchi, M. Allosteric regulation of 3CL protease of SARS-CoV-2 and SARS-CoV observed in the crystal structure ensemble. *J. Mol. Biol.* **2021**, *433*, No. 167324.
(42) Eberle, R. J.; Olivier, D. S.; Amaral, M. S.; Gering, I.; Willbold, D.; Arni, R. K.; Coronado, M. A. The Repurposed Drugs Suramin and Quinacrine Cooperatively Inhibit SARS-CoV-2 3CLpro In Vitro. *Viruses* **2021**, *13*, 873.
(43) Trott, O.; Olson, A. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimisation, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455−461.
(44) Bojadzic, D.; Alcazar, O.; Chen, J.; Chuang, S. T.; Capcha, J. M. C.; Shehadeh, L. A.; Buchwald, P. Small-Molecule Inhibitors of the Coronavirus Spike: ACE2 Protein−Protein Interaction as Blockers of Viral Attachment and Entry for SARS-CoV-2. *ACS Infect. Dis.* **2021**, *7*, 1519−1534.
(45) Sevenich, M.; Thul, E.; Lakomek, N. A.; Klünemann, T.; Schubert, M.; Bertoglio, F.; van den Heuvel, J.; Petzsch, P.; Mohrlüder, J.; Willbold, D. Phage Display-Derived Compounds Displace hACE2 from Its Complex with SARS-CoV-2 Spike Protein. *Biomedicines* **2022**, *10*, 441.
(46) Singh, E.; Khan, R. J.; Jha, R. K.; Amera, G. M.; Jain, M.; Singh, R. P.; Muthukumaran, J.; Singh, A. K. A comprehensive review on promising antiviral therapeutic candidates identified against main protease from SARS-CoV-2 through various computational methods. *J. Genet. Eng. Biotechnol.* **2020**, *18*, 1−12.
(47) Santur, K.; Reinartz, E.; Lien, Y.; Tusche, M.; Altendorf, T.; Sevenich, M.; Tamgüney, G.; Mohrlüder, J.; Willbold, D. Ligand-Induced Stabilisation of the Native Human Superoxide Dismutase 1. *ACS Chem. Neurosci.* **2021**, *12*, 2520−2528.
(48) Krejci, A.; Hupp, T. R.; Lexa, M.; Vojtesek, B.; Muller, P. Hammock: A hidden Markov model-based peptide clustering algorithm to identify protein-interaction consensus motifs in large datasets. *Bioinformatics* **2016**, *32*, 9−16.
(49) 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 α-ketoamide inhibitors. *Science* **2020**, *368*, 409−412.
(50) Zhang, L.; Lin, D.; Kusov, Y.; Nian, Y.; Ma, Q.; Wang, J.; De Wilde, A.; et al. α-Ketoamides as broad-spectrum inhibitors of coronavirus and enterovirus replication: Structure-based design, synthesis, and activity assessment. *J. Med. Chem.* **2020**, *63*, 4562−4578.
(51) Ma, C.; Sacco, M. D.; Hurst, B.; Townsend, J. A.; Hu, Y.; Szeto, T.; Zhang, X.; Tarbet, B.; Marty, M. T.; Chen, Y.; Wang, J. Bo-ceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. *Cell Res.* **2020**, *30*, 678−692.
(52) Roy, A.; Lim, L.; Srivastava, S.; Lu, Y.; Song, J. Solution conformations of Zika NS2B-NS3pro and its inhibition by natural products from edible plants. *PLoS One* **2017**, *12*, No. e0180632.
(53) Motulsky, H.; Christopoulos, A.*Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting*; Oxford University Press, 2004.
(54) Feng, B. Y.; Shoichet, B. K. A detergent-based assay for the detection of promiscuous inhibitors. *Nat. Protoc.* **2006**, *1*, 550−553.
(55) Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 12366−12371.
(56) Skinner, S. P.; Fogh, R. H.; Boucher, W.; Ragan, T. J.; Mureddu, L. G.; Vuister, G. W. CcpNmr AnalysisAssign: a flexible platform for integrated NMR analysis. *J. Biomol. NMR* **2016**, *66*, 111−124.
(57) Elfgen, A.; Santiago-Schübel, B.; Gremer, L.; Kutzsche, J.; Willbold, D. Surprisingly high stability of the Aβ oligomer eliminating all-d-enantiomeric peptide D3 in media simulating the route of orally administered drugs. *Eur. J. Pharm. Sci.* **2017**, *107*, 203−207.
(58) Elfgen, A.; Hupert, M.; Bochinsky, K.; Tusche, M.; González de San Román, M. E.; Gering, I.; Sacchi, S.; Pollegioni, L.; Huesgen, P. F.; Hartmann, R.; Santiago-Schübel, B.; et al. Metabolic resistance of the D-peptide RD2 developed for direct elimination of amyloid-β oligomers. *Sci. Rep.* **2019**, *9*, No. 5715.

https://doi.org/10.1021/acschembio.2c00735
ACS Chem. Biol. 2023, 18, 315−330

(59) Tien, M. Z.; Sydykova, D. K.; Meyer, A. G.; Wilke, C. O. PeptideBuilder: A simple Python library to generate model peptides. *Peer J.* **2013**, *1*, No. e80.
(60) Tang, B.; He, F.; Liu, D.; Fang, M.; Wu, Z.; Xu, D.*AI-Aided Design of Novel Targeted Covalent Inhibitors against SARS-CoV-2*; bioRxiv, 2020.
(61) Goodsell, D.; Morris, G.; Olson, A. Automated docking of flexible ligands: Applications of AutoDock. *J. Mol. Recognit.* **1996**, *9*, 1−5.
(62) Santos-Martins, D.; Forli, S.; Ramos, M.; Olson, A. AutoDock4Zn: An Improved AutoDock Force Field for Small-Molecule Docking to Zinc Metalloproteins. *J. Chem. Inf. Model.* **2014**, *54*, 2371−2379.
(63) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1−2*, 19−25.
(64) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14sb: Improving the accuracy of protein side chain and backbone parameters from ff99sb. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.
(65) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(66) Cauchy, M. A. Méthode généerale pour la ystemson des ystems d' éequa-tions simultanéees. *C. R. Hebd. Seances Acad. Sci.* **1847**, *25*, 536−538.
(67) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, No. 014101.
(68) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(69) Nosé, S. Molecular-dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **1984**, *52*, 255−268.
(70) Parrinello, M.; Rahman, A. Polymorphic transitions in single-crystals—a new molecular-dynamics method. *Mol. Phys.* **1981**, *52*, 7182−7190.
(71) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald—an N. Log(N) method for ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(72) Essmann, U.; Perera, L.; Berkowitz, M.; et al. A smooth particle mesh ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(73) Hess, B.; Bekker, H.; Berendsen, H.; Fraaije, J. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(74) Laurent, B.; Chavent, M.; Cragnolini, T.; Dahl, A. C. E.; Pasquali, S.; Derreumaux, P.; Sansom, M. S. P.; Baaden, M. Epock: rapid analysis of protein pocket dynamics. *Bioinformatics.* **2015**, *31*, 1478−1480.
(75) Valdés-Tresanco, M. S.; Valdés-Tresanco, M. E.; Valiente, P. A.; Moreno, E. gmx_MMPBSA: a new tool to perform end-state free energy calculations with GROMACS. *J. Chem. Theory Comput.* **2021**, *17*, 6281−6291.
(76) Wang, J.; Cai, Q.; Xiang, Y.; Luo, R. Reducing grid dependence in finite-difference Poisson−Boltzmann calculations. *J. Chem. Theory Comput.* **2012**, *8*, 2741−2751.
(77) Tan, C.; Tan, Y. H.; Luo, R. Implicit nonpolar solvent models. *J. Phys. Chem. B* **2007**, *111*, 12263−12274.
(78) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W.; Mark, A. E. Peptide folding: when simulation meets experiment. *Angew. Chem., Int. Ed.* **1999**, *38*, 236−240.

Recommended by ACS

Development of Highly Potent Noncovalent Inhibitors of SARS-CoV-2 3CLpro
Ningke Hou, Qi Hu, *et al.*
JANUARY 25, 2023
ACS CENTRAL SCIENCE
READ

AI-Accelerated Design of Targeted Covalent Inhibitors for SARS-CoV-2
Rajendra P. Joshi, Neeraj Kumar, *et al.*
FEBRUARY 21, 2023
JOURNAL OF CHEMICAL INFORMATION AND MODELING
READ

Identification of Human Host Substrates of the SARS-CoV-2 M^{pro} and PL^{pro} Using Subtiligase N-Terminomics
Shu Y. Luo, Olivier Julien, *et al.*
APRIL 03, 2023
ACS INFECTIOUS DISEASES
READ

Scalable Inhibitors of the Nsp3–Nsp4 Coupling in SARS-CoV-2
Abdul-Rahman Azizogli, Vivek A. Kumar, *et al.*
FEBRUARY 06, 2023
ACS OMEGA
READ

Get More Suggestions >

https://doi.org/10.1021/acschembio.2c00735
ACS Chem. Biol. 2023, 18, 315−330

Supplementary Material

Discovery of all-D-peptide inhibitors of SARS-CoV-2 3C-like protease

Raphael J. Eberle[†,‡], Marc Sevenich[†,‡,ª], Ian Gering[†], Lara Scharbert[†,‡], Birgit Strodel[†,‡], Nils A. Lakomek[†,‡], Karoline Santur[†,‡], Jeannine Mohrlüder[†], Mônika A. Coronado[†*] and Dieter Willbold[†,‡*]

[†]Institute of Biological Information Processing: Structural Biochemistry (IBI-7), Forschungszentrum Jülich, 52425 Jülich, Germany

[‡]Institut für Physikalische Biologie, Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany

[ª]Priavoid GmbH, Merowingerplatz 1, 40225 Düsseldorf, Germany

[*]Corresponding authors: Tel: +49-2461-619505, m.coronado@fz-juelich.de;

Tel.: +49-2461-612100, Fax: +49-2461-612023, d.willbold@fz-juelich.de

Table of contents

Supplementary figure S1. Purification of SARS-CoV-2 $3CL^{pro}$.

Supplementary figure S2. Phage display selection on SARS-CoV-2 $3CL^{pro}$.

Supplementary figure S3. Enrichment ELISA with input titer samples on the immobilized SARS-CoV-2 $3CL^{pro}$ selection target protein.

Supplementary figure S4. Inhibition effect and inhibition mode of 3CVL peptides against SARS-CoV-2 $3CL^{pro}$.

Supplementary figure S5. Inhibition effect and inhibition mode of 3CVLri peptides against SARS-CoV-2 $3CL^{pro}$.

Supplementary figure S6. Sequence comparison between the preferred cleavage sequence pattern of SARS-CoV-2 $3CL^{pro}$ and the investigated peptide inhibitors.

Supplementary figure S7. Inhibition effect and inhibition mode of combined 3CVLri peptides against SARS-CoV-2 $3CL^{pro}$

Supplementary figure S8. Inhibition effect of Rupintrivir and Rupintrivir combined with 3CVLri-4 against SARS-CoV-2 $3CL^{pro}$.

Supplementary figure S9. SPR measurements of the 3CVL-2, - 4 and 3CVLri-2, -4 peptides

Supplementary figure S10. SPR measurements of the 3CVL- 4 and 3CVLri-4 peptides.

Supplementary figure S11. Stability and promiscuous assays of 3CVLri peptides.

Supplementary figure S12. Effect of Triton X-100 on the 3CVLri-2, -4 and -7 inhibition against SARS-CoV-2 $3CL^{pro}$.

Supplementary figure S13. HPLC profiles of 3CVLri-2, -4 and -7 incubated in simulated gastric fluid (SGF).

Supplementary figure S14. HPLC profiles of 3CVLri-2, -4 and -7 incubated in simulated intestinal fluid (SIF).

Supplementary figure S15. HPLC profiles of 3CVLri-2, -4 and -7 incubated in human plasma.

Supplementary figure S16. HPLC profiles of 3CVLri-2, -4 and -7 incubated in liver microsomes.

Supplementary figure S17. Circular dichroism (CD) spectroscopy of 3CVL and 3CVLri peptides.

Supplementary figure S18. Active site pocket fluctuation of the SARS-CoV-2 $3CL^{pro}$ during 200 ns of MD simulation.

Supplementary figure S19. RMSD values for the MD simulation of $3CL^{pro}$ in complex with 3CVL and 3CVLri peptides.

Supplementary figure S20. RMSF profiles for the MD simulation of $3CL^{pro}$ in complex with 3CVL and 3CVLri peptides.

Supplementary figure S21. RMSF profiles for the MD simulation of $3CL^{pro}$ in complex with 3CVL and 3CVLri peptides.

Supplementary figure S22. Representative $3CL^{pro}$-L-peptide complex structures.

Supplementary figure S23. Representative $3CL^{pro}$-D-retro-inverse peptide complex structures.

Supplementary figure S24. RMSD values, RMSF values and Hbonds between peptides and protein for $3CL^{pro}$ in complex with the 3CVL and 3CVLri peptides after clustering of the non-flexible/interacting peptide region.

Supplementary figure S25. Two-dimensional 1H-15N correlation spectra of 100 µM 15N $3CL^{pro}$, in the presence and absence of 3CVL-2 or 3CVLri-2.

Supplementary figure S26. Two-dimensional 1H-15N correlation spectra of 100 µM 15N $3CL^{pro}$, in the presence and absence of 3CVL-7 or 3CVLri-7.

Supplementary figure S27. Volume of the SARS-CoV-2 $3CL^{pro}$ active site with and without the D-ri-peptides.

Supplementary figure S28. Computational alanine scanning of 3CVL and 3CVLri peptides to determine important residues for the interaction with $3C^{Lpro}$.

Supplementary figure S29. HPLC chromatogram and mass spectrometry of 3CVL peptides 1-4.

Supplementary figure S30. HPLC chromatogram and mass spectrometry of 3CVL peptides 5-8.

Supplementary figure S31. HPLC chromatogram and mass spectrometry of 3CVL peptides 9-11.

Supplementary figure S32. HPLC chromatogram and mass spectrometry of 3CVLri peptides.

Supplementary table S1. List of selected 3CL^{pro} phage display peptides.

Supplementary table S2. Evaluation of 3CVL-2, -4 and -7 sequences with TSAT.

Supplementary table S3. Evaluation of 3CVL-2, -4 and -7 sequences with Hammock.

Supplementary table S4. Contributing amino acids in the Substrate-binding site and proposed allosteric site of SARS-CoV-2 3CL^{pro}.

Supplementary table S5. Purity and molecular weight of purchased L- and D-ri-peptides.

Supplementary Text S1. Selection of SARS-CoV-2 3CL^{pro} binding peptides using phage display.

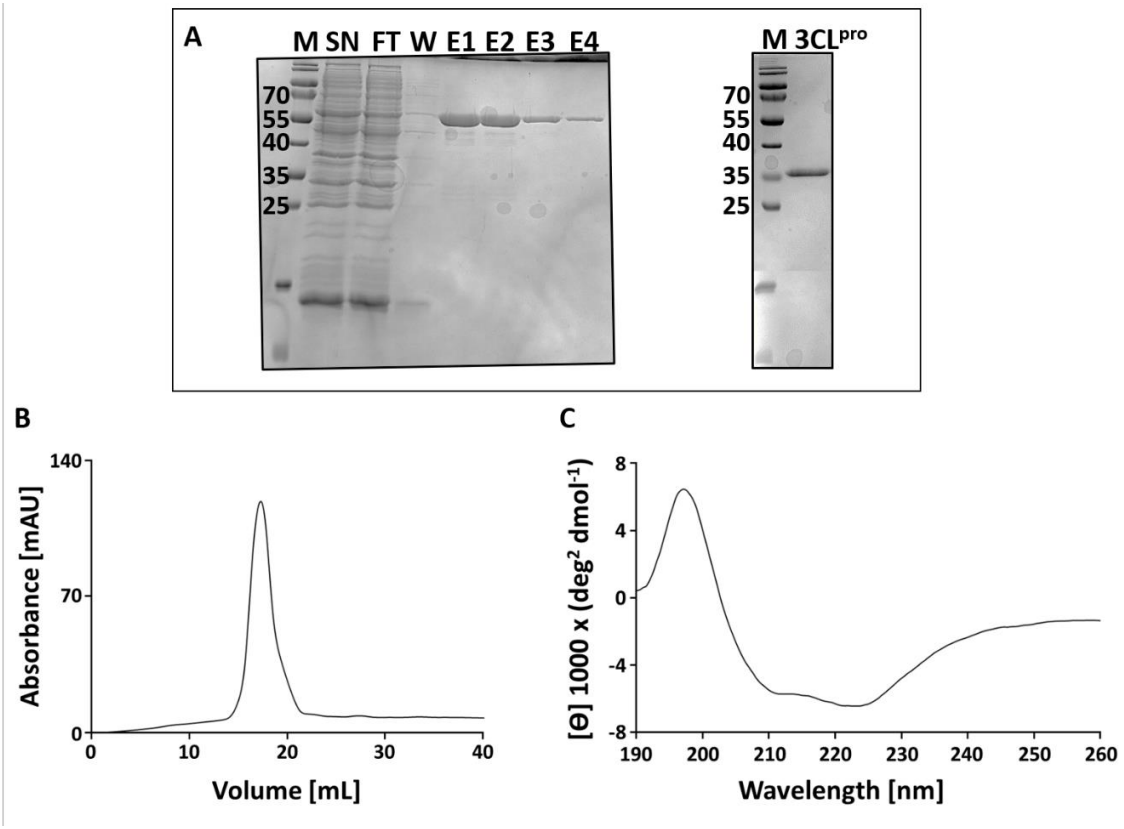


Figure S1. Purification of SARS-CoV-2 3CL^{pro}. (A) SDS 15% Gel of 3CL^{pro}_GST purification with GSH sepharose. M: protein marker, SN: supernatant, FT: flow-through, W: washing step, E1-E4: Elution. SDS 15% Gel of pure 3CL^{pro} after cleavage of the GST tag using PreScission protease. **(B)** 3CL^{pro} Chromatogram after size exclusion chromatography. **(C)** CD spectrum of 3CL^{pro}.

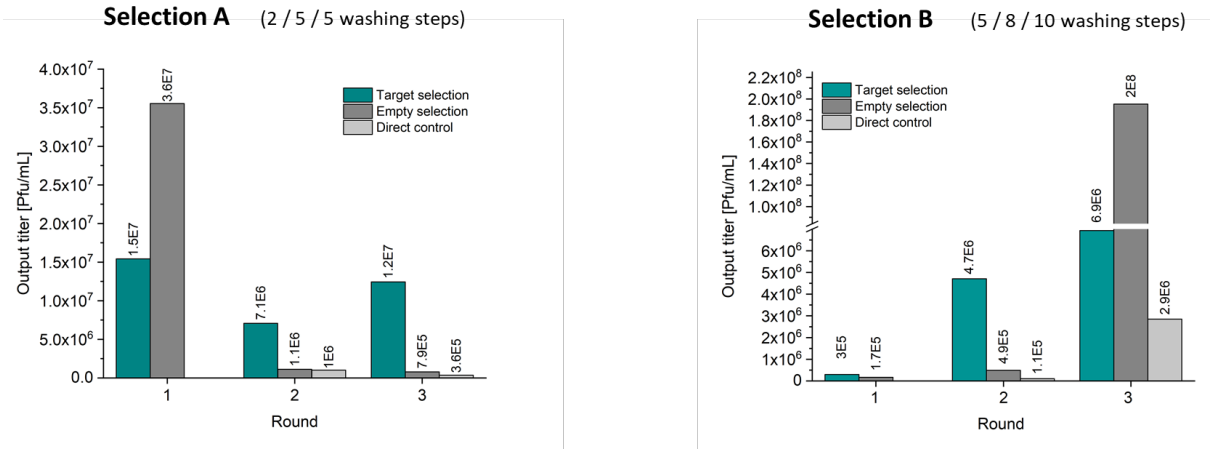


Figure S2. Phage display selection on SARS-CoV-2 3CL^{pro}. Selection A and B in three consecutive selection rounds. Output phage titer in PFU/mL for selection samples of TS, ES and DC as measured past phage elution after biopanning.

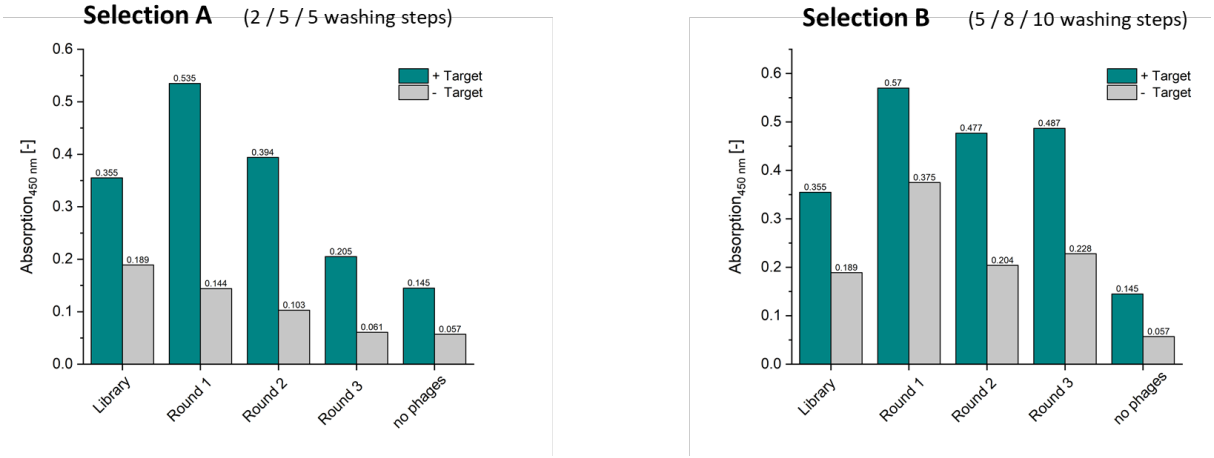


Figure S3. Enrichment ELISA with input titer samples on the immobilized SARS-CoV-2 3CL^{pro} selection target protein. Enrichment ELISA of the 3CL^{pro} phage display was performed to verify the enrichment of 3CL^{pro}-binding peptides during the selection. Relative binding affinities of the phage samples (5 * 10^{10} phages) of the wild type, library, and each selection round to immobilized 3CL^{pro} (dark green) and to empty wells without 3CL^{pro} (gray) were investigated in a double determination (points: corresponding values) as well as a control without phages.

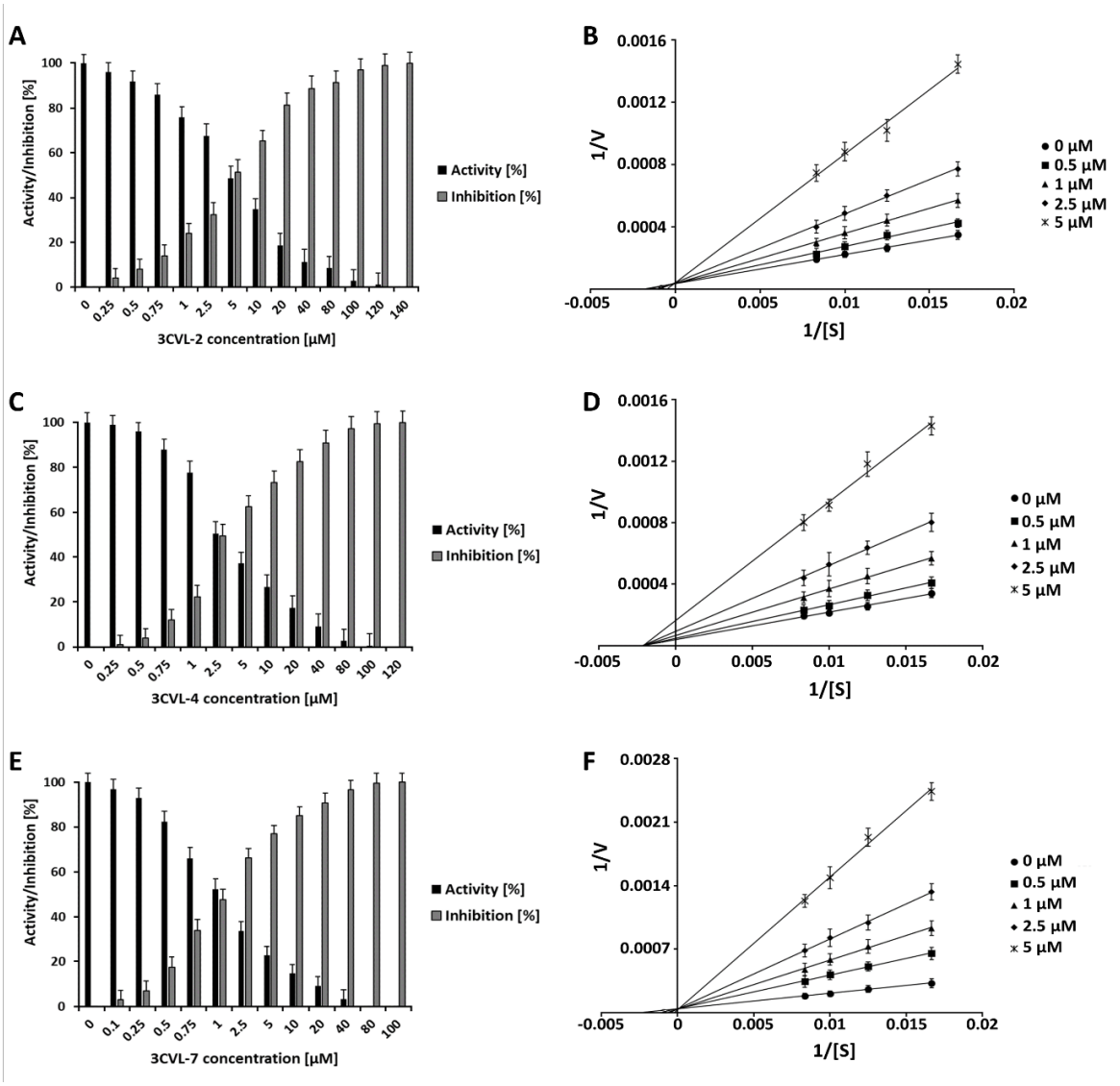


Figure S4. Inhibition effect and inhibition mode of 3CVL peptides against SARS-CoV-2 3CL^{pro}. Normalized activity and inhibition of SARS-CoV-2 3CL^{pro} under inhibitor influence. **(A)** 3CVL-2, **(C)** 3CVL-4 and **(E)** 3CVL-7. Data shown are the mean ± SD from three independent measurements (n=3). Lineweaver-Burk plots are shown to determine the inhibition mode. [S] is the substrate concentration; v is the initial reaction rate. **(B)** 3CVL-2, **(D)** 3CVL-4 and **(F)** 3CVL-7. Data shown are the mean ± SD from 3 independent measurements (n=3).

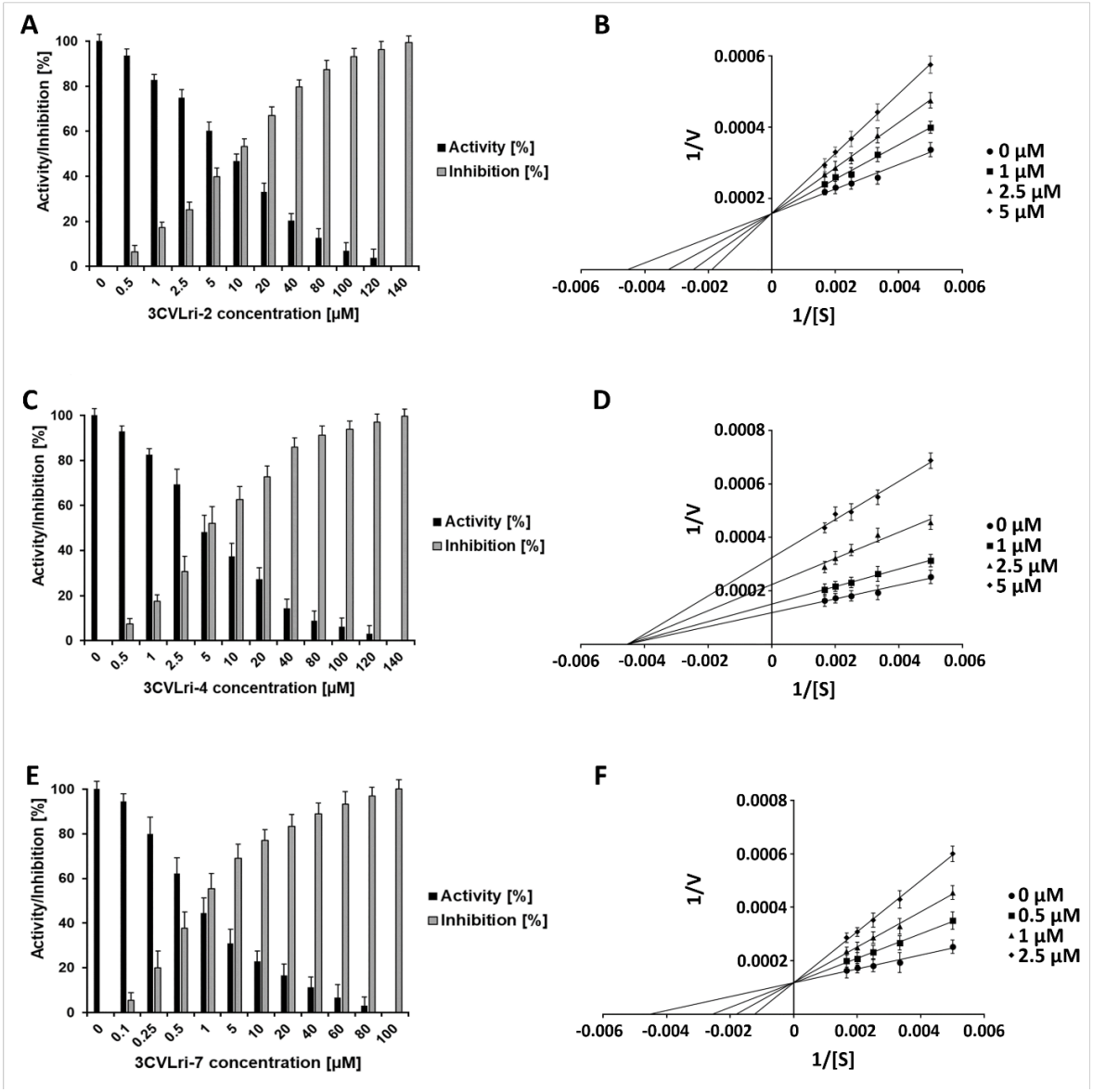


Figure S5. Inhibition effect and inhibition mode of 3CVLri peptides against SARS-CoV-2 3CL^{pro}. Normalized activity and inhibition of SARS-CoV-2 3CL^{pro} under inhibitor influence. **(A)** 3CVLri-2, **(C)** 3CVLri-4 and **(E)** 3CVLri-7. Data shown are the mean ± SD from three independent measurements (n=3). Lineweaver-Burk plots are shown to determine the inhibition mode. [S] is the substrate concentration; v is the initial reaction rate. **(B)** 3CVLri-2, **(D)** 3CVLri-4 and **(F)** 3CVLri-7. Data shown are the mean ± SD from 3 independent measurements (n=3).

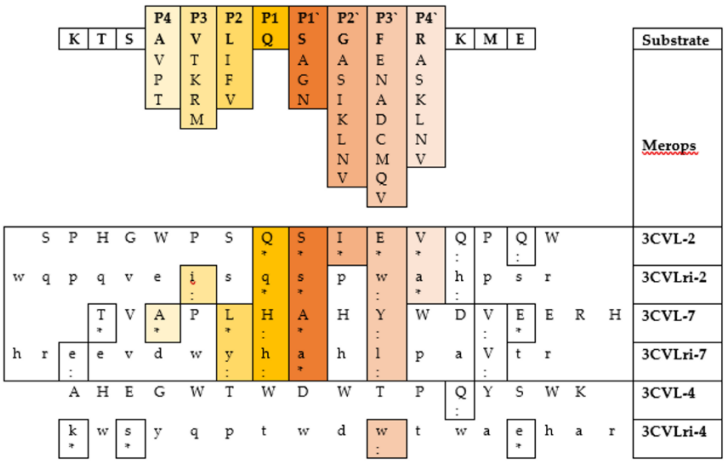


Figure S6. Sequence comparison between the preferred cleavage sequence pattern of SARS-CoV-2 3CL^{pro} and the investigated peptide inhibitors. 3CL^{pro} substrate binding site and preferred amino acids are shown based on the used substrate sequence and published sequences on the Merops database[1]. Asterisks label the appearance of the same residue in the sequence and colon similar residues. The competitive inhibitors (3CVL-2, 3CVLri-2, 3CVL-7 and 3CVLri-7) shows a higher similarity to the preferred residues in the 3CL^{pro} substrate binding site compared to the noncompetitive inhibitors (3CVL-4 and 3CVLri-4).

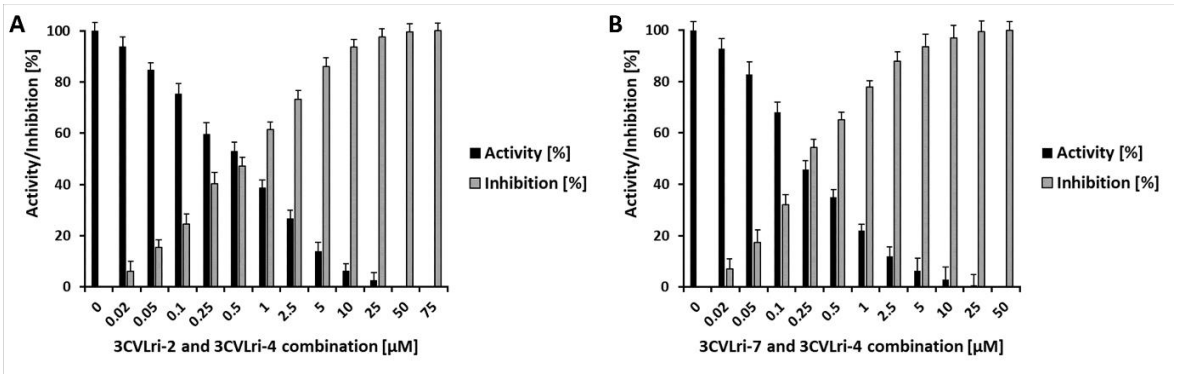


Figure S7. Inhibition effect and inhibition mode of combined 3CVLri peptides against SARS-CoV-2 3CL^{pro}. Normalized activity and inhibition of SARS-CoV-2 3CL^{pro} under the influence of competitive and noncompetitive 3CVLri peptide combinations. **(A)** 3CVLri-2 + 3CVLri-4. **(B)** 3CVLri-7 + 3CVLri-4. Data shown are the mean ± SD from three independent measurements (n=3).

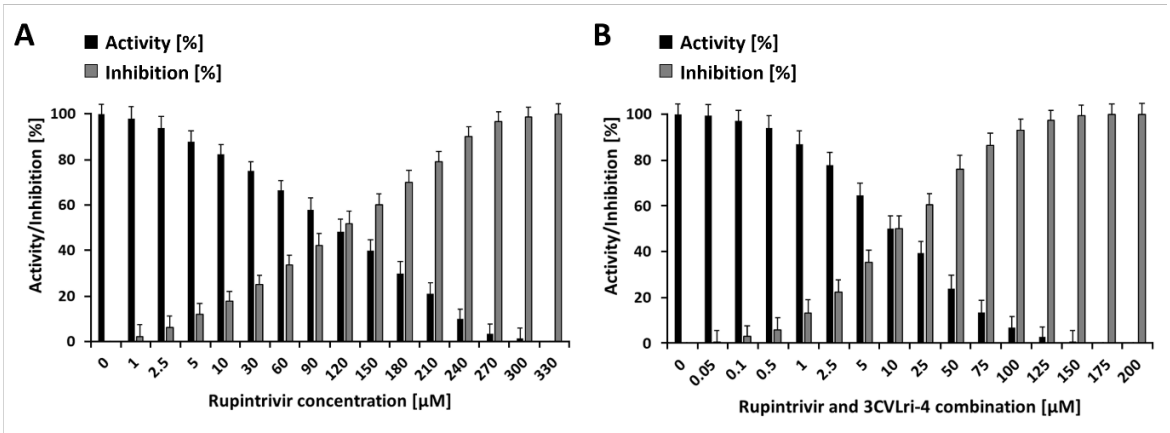
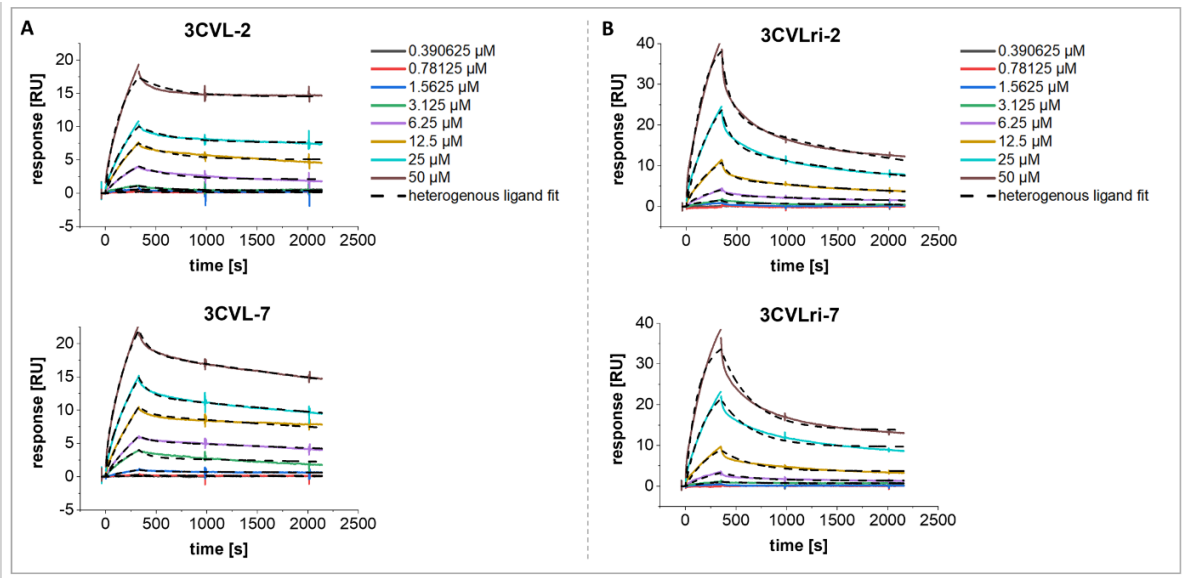


Figure S8. Inhibition effect of Rupintrivir and Rupintrivir combined with 3CVLri-4 against SARS-CoV-2 3CL^{pro}. Normalised activity and inhibition of SARS-CoV-2 3CL^{pro} under the influence of Rupintrivir and Rupintrivir combined with 3CVLri-4. **(A)** Rupintrivir, **(B)** Rupintrivir and 3CVLri-4 combination. Data shown are the mean ± SD from three independent measurements (n=3).



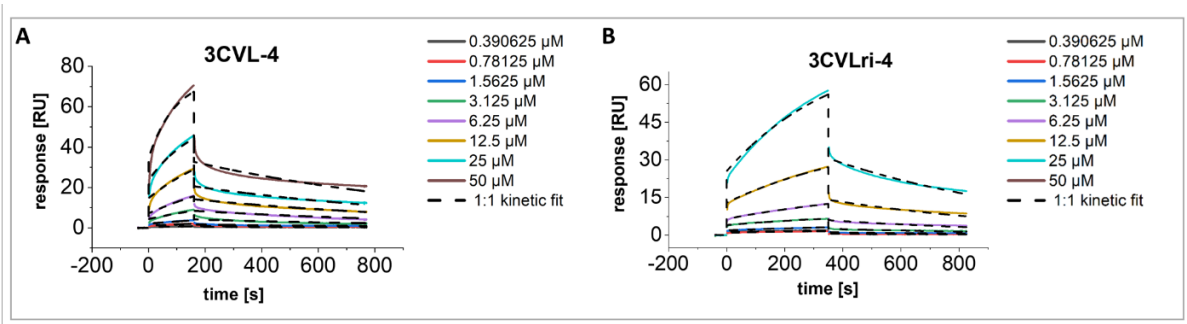
C

ID	k_{a1} [1/Ms]	k_{d1} [1/s]	K_{D1} [M]	k_{a2} [1/Ms]	k_{d2} [1/s]	K_{D2} [M]
3CVL-2	9.45E1	5.05E-8	5.34E-10	4.60E2	3.05E-3	6.50E-6
3CVL-7	7.15E1	1.28E-4	1.8E-6	1.29E3	1.00E-2	7.79E-6

D

ID	k_{a1} [1/Ms]	k_{d1} [1/s]	K_{D1} [M]	k_{a2} [1/Ms]	k_{d2} [1/s]	K_{D2} [M]
3CVLri-2	3.68E6	1.72E2	4.68E-5	2.63E2	4.59E-4	1.74E-6
3CVLri-7	1.75E2	4.25E-7	2.43E-9	9.36E1	3.28E-3	3.50E-5

Figure S9. SPR measurements of the 3CVL-2, - 7 and 3CVLri-2, -7 peptides. Heterogenous ligand based fitting model. Sensograms showing binding response (colored, solid lines) and heterogenous ligand fit (black, dotted lines). **(A)** 3CVL sensograms and **(B)** 3CVLri sensograms. Kinetic rates and dissociations constants as determined by heterogenous ligand fitting for **(C)** 3CVL peptides and **(D)** 3CVLri peptides.



C

ID	ka [1/Ms]	kd [1/s]	K_D [µM]
3CVL-4	1.93E+2	9.73E-4	5.1

D

ID	ka [1/Ms]	kd [1/s]	K_D [µM]
3CVLri-4	5.25E+1	1.32E-3	25.2

Figure S10. SPR measurements of the 3CVL- 4 and 3CVLri-4 peptides. Global kinetic fitting of 1:1 stoichiometric Langmuir based fitting model. Sensograms showing 3CVLri binding response (colored, solid lines) and 1:1 kinetic fitting model (black, dotted lines). **(A)** 3CVL-4 and **(B)** 3CVLri-4. Kinetic rates and dissociations constants as determined by global 1:1 fitting. **(C)** 3CVL-4 and **(D)** 3CVLri-4

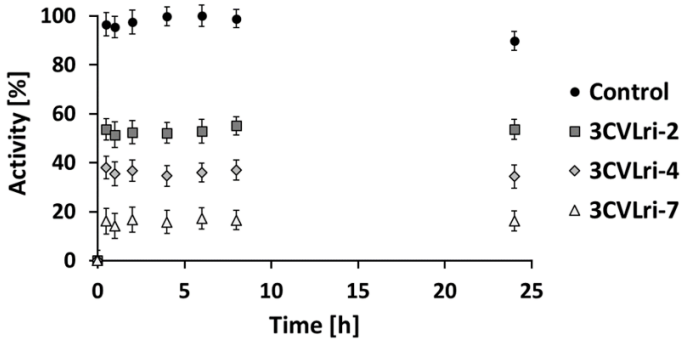


Figure S11. Stability and promiscuous assays of 3CVLri peptides. 24h stability experiment of 3CVLri peptides against SARS-CoV-2 $3CL^{pro}$ activity.

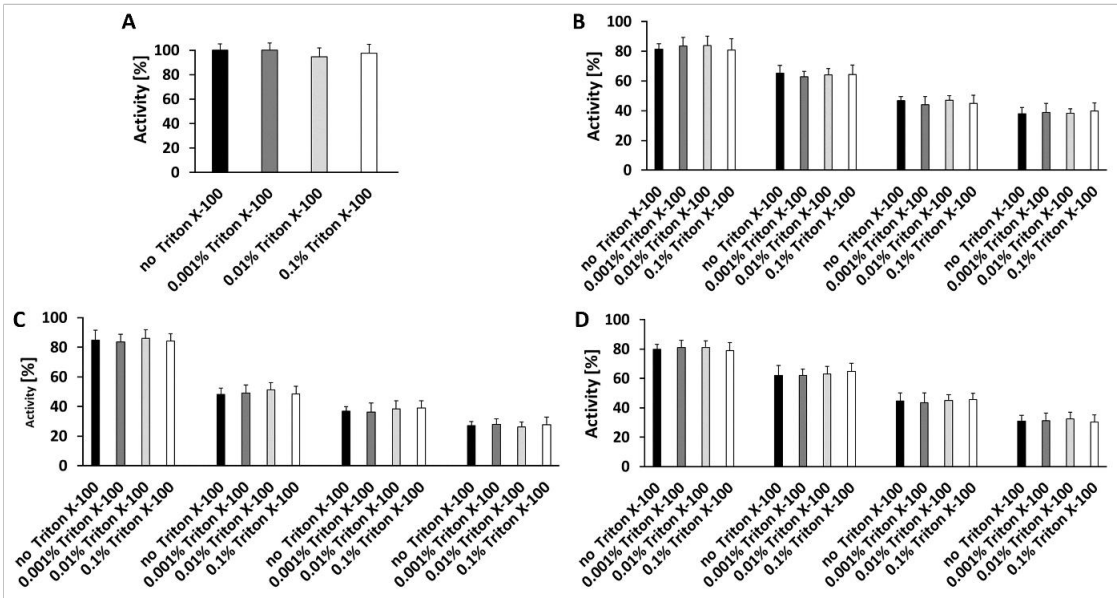


Figure S12. Effect of Triton X-100 on the 3CVLri-2, -4 and -7 inhibition against SARS-CoV-2 3CL^{pro}. Three Triton X-100 concentrations (0.001%, 0.01% and 0.1%) were tested with four different inhibitor concentrations and compared with the results without detergent. Data shown are the mean ± SD from three independent measurements (n=3). **(A)** Effect of Triton X-100 on 3CL^{pro} activity. **(B)** Effect of Triton X-100 on 3CVLri-2 inhibition against 3CL^{pro}. **(C)** Effect of Triton X-100 on 3CVLri-4 inhibition against 3CL^{pro}. **(D)** Effect of Triton X-100 on 3CVLri-7 inhibition against 3CL^{pro}.

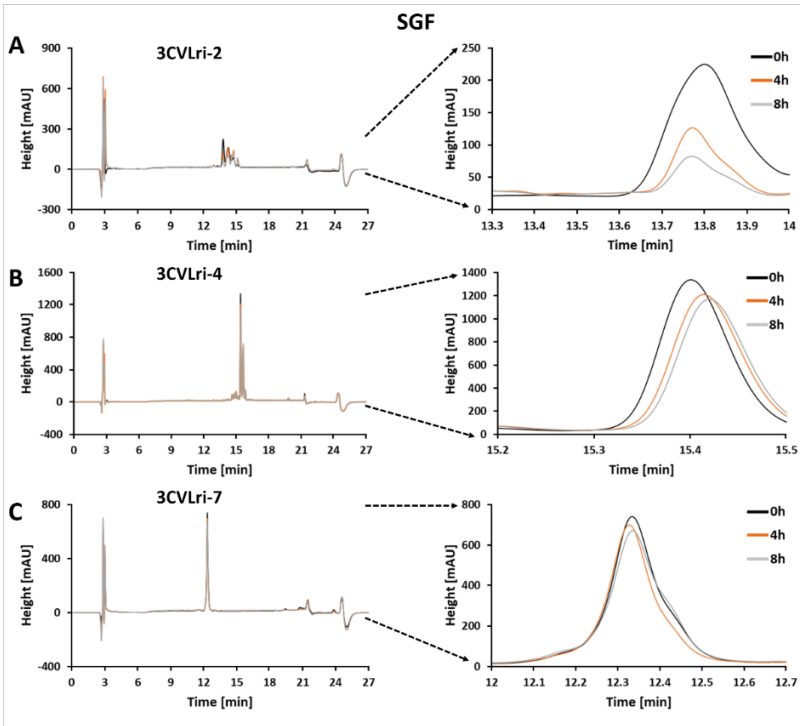


Figure S13. HPLC profiles of 3CVLri-2, -4 and -7 incubated in simulated gastric fluid (SGF). Incubation time was 0h, 4h and 8h. **(A)** HPLC profile of 3CVLri-2, **(B)** 3CVLri-4 and **(C)** 3CVLri-7.

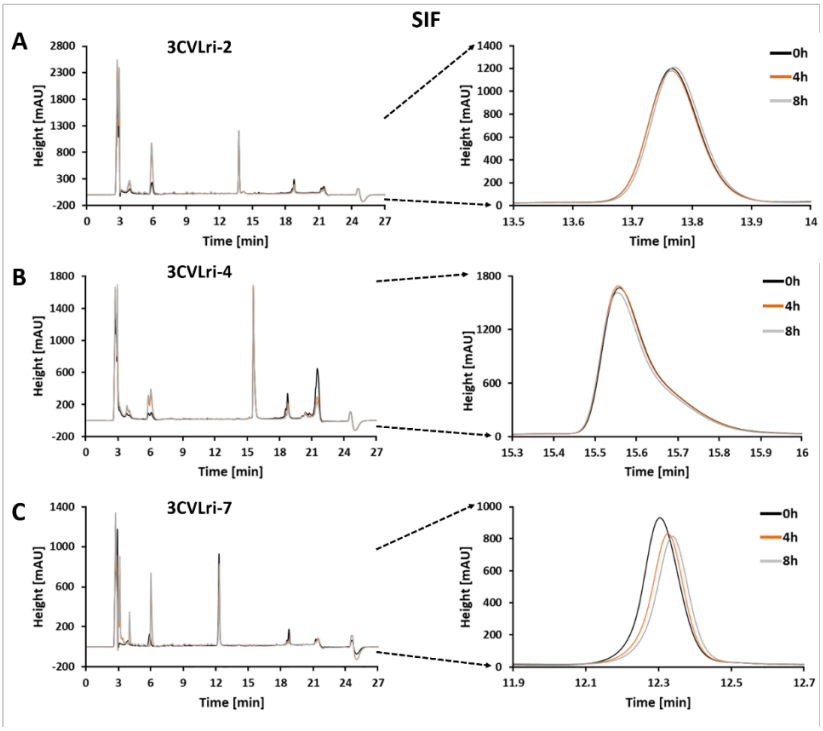


Figure S14. HPLC profiles of 3CVLri-2, -4 and -7 incubated in simulated intestinal fluid (SIF). Incubation time was 0h, 4h and 8h **(A)** HPLC profile of 3CVLri-2, **(B)** 3CVLri-4 and **(C)** 3CVLri-7.

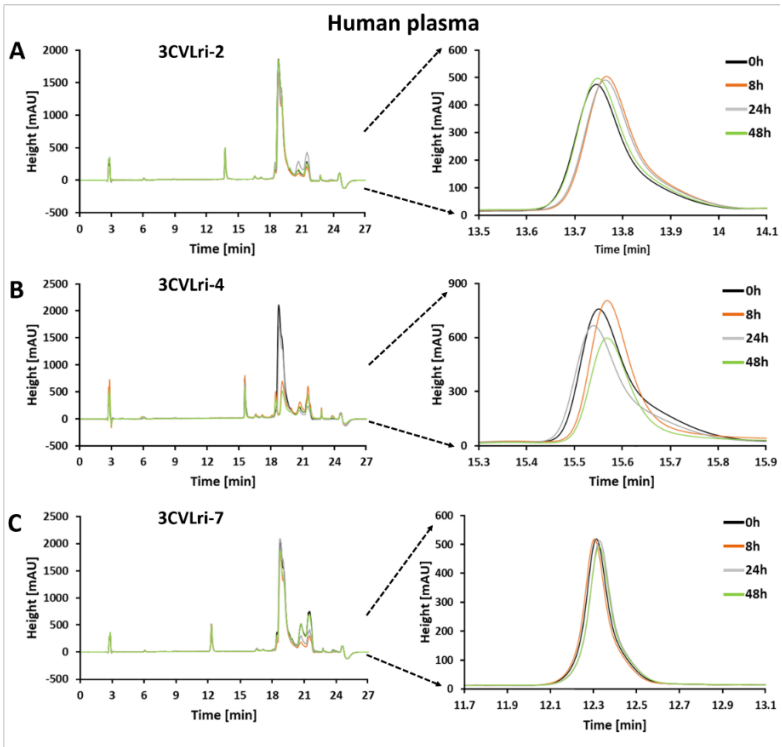


Figure S15. HPLC profiles of 3CVLri-2, -4 and -7 incubated in human plasma. Incubation time was 0h, 8h, 24h and 48h **(A)** HPLC profile of 3CVLri-2, **(B)** 3CVLri-4 and **(C)** 3CVLri-7.

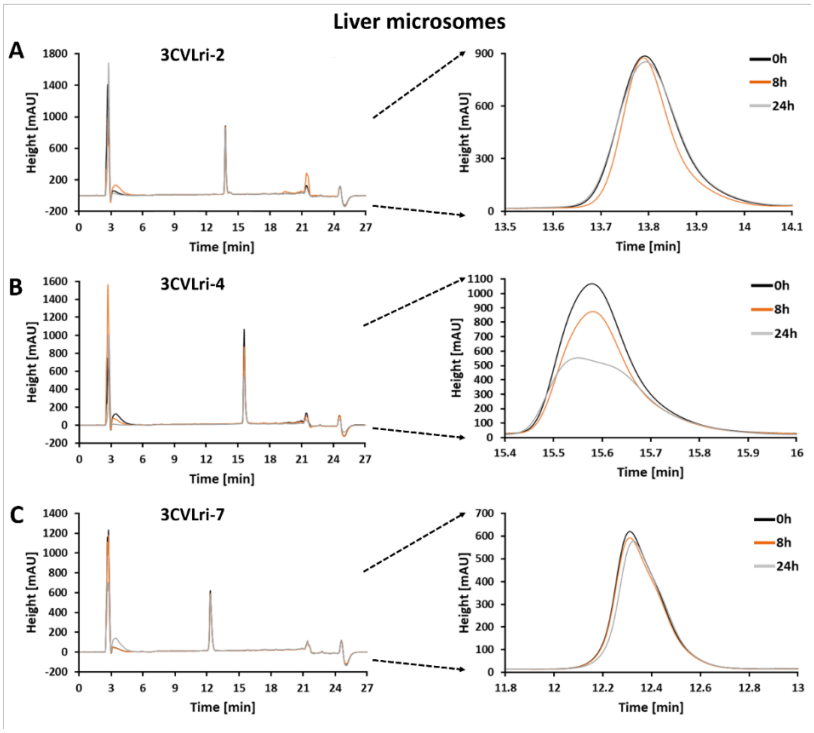


Figure S16. HPLC profiles of 3CVLri-2, -4 and -7 incubated in liver microsomes. Incubation time was 0h, 8h and 24h **(A)** HPLC profile of 3CVLri-2, **(B)** 3CVLri-4 and **(C)** 3CVLri-7.

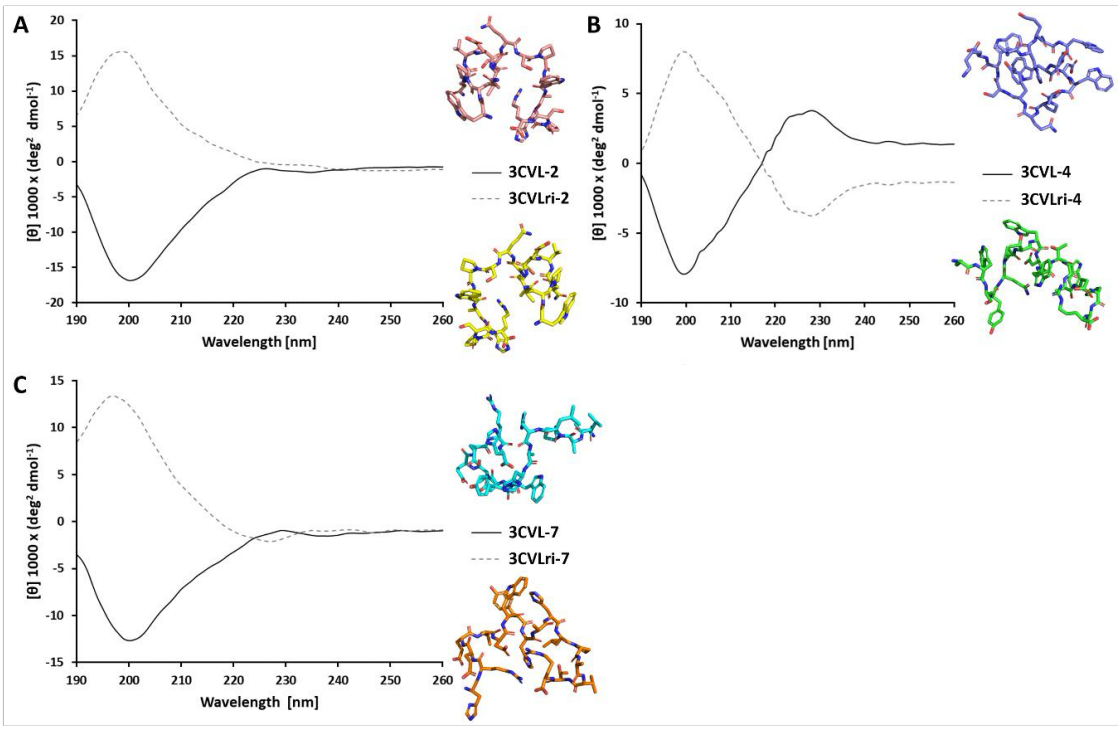


Figure S17. Circular Dichroism (CD) spectroscopy of 3CVL and 3CVLri peptides. The CD spectrum of the peptides in solution are presented as molar ellipticity [θ]. 3D models of the peptides are shown together with the corresponding CD-spectrum of the peptides. **(A)** 3CVL-2 and 3VLri-2, **(B)** 3CVL-4 and 3CVLri-4, **(C)** 3CVL-7 and 3CVLri-7.

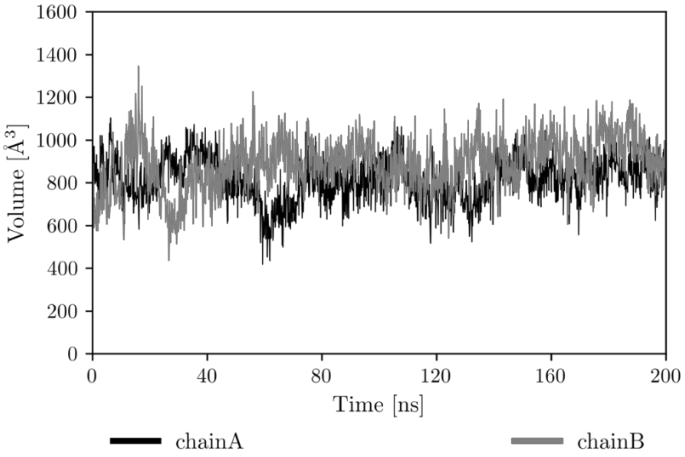


Figure S18. Active site pocket fluctuation of the SARS-CoV-2 3CL^{pro} during 200 ns of MD simulation. The fluctuation of the chain A (black) and chain B (grey) are shown.

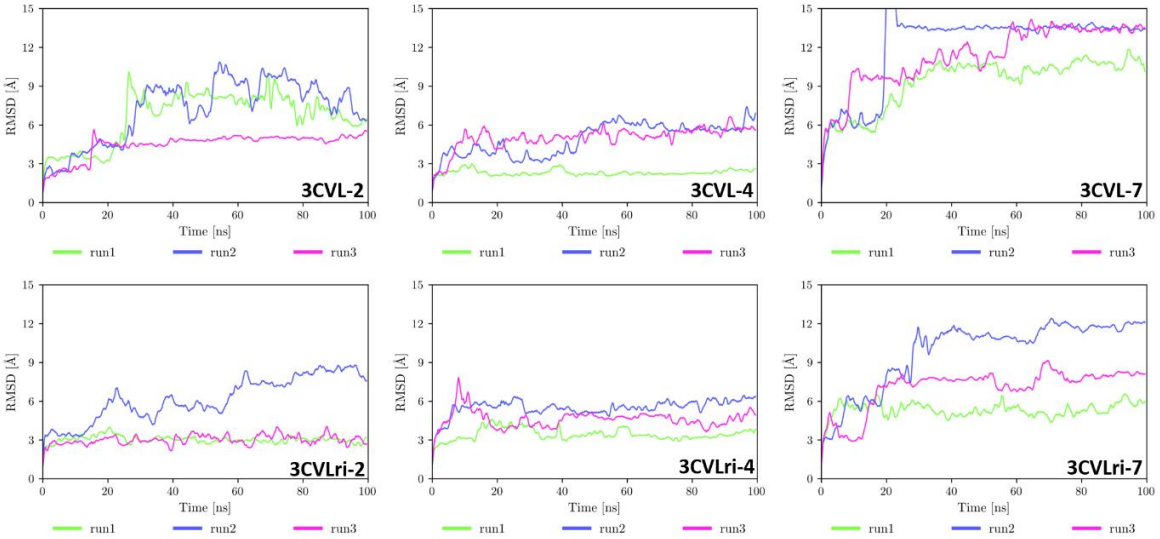


Figure S19. RMSD values for the MD simulation of 3CL^{pro} in complex with 3CVL and 3CVLri peptides (Three simulations with different initial structures per peptide). 3CVL-2, 3CVL-4, 3CVL-7, 3CVLri-2, 3CVLri-4 and 3CVLri-7, three different start structures generated by docking were considered: run1 (green), run2 (blue) and run3 (pink) are shown.

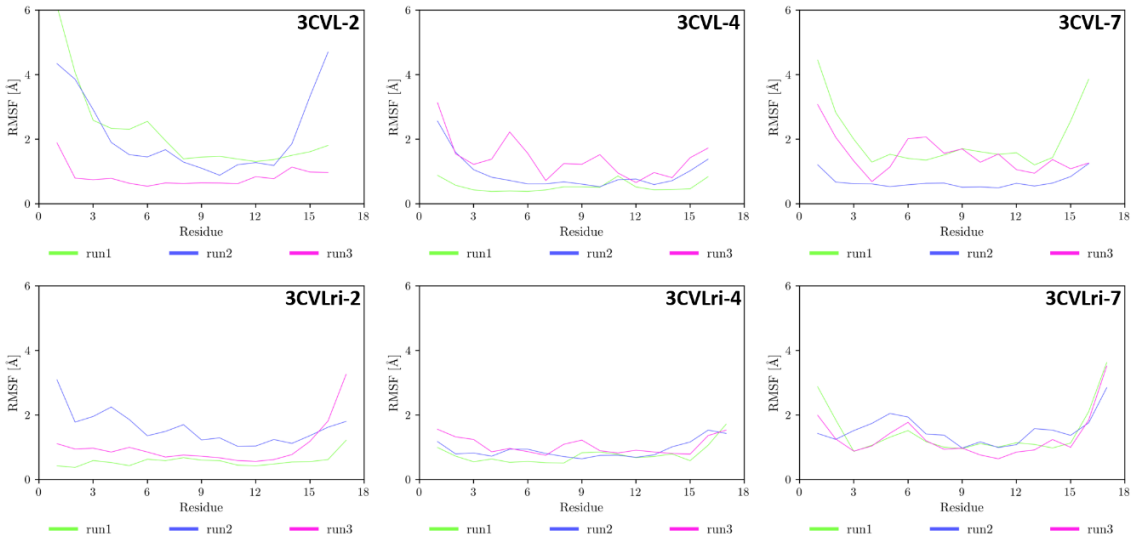


Figure S20. RMSF profiles for the MD simulation of 3CL^{pro} in complex with 3CVL and 3CVLri peptides (Three simulations with different initial structures per peptide). 3CVL-2, 3CVL-4, 3CVL-7, 3CVLri-2, 3CVLri-4 and 3CVLri-7. Run1 (green), run2 (blue) and run3 (pink) are shown.

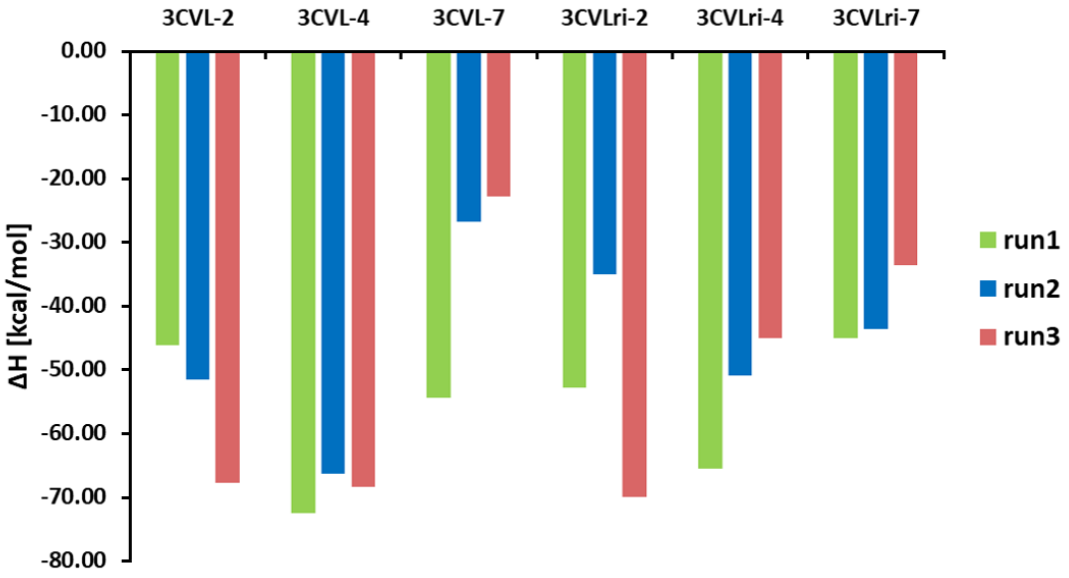


Figure S21. RMSF profiles for the MD simulation of 3CL^{pro} in complex with 3CVL and 3CVLri peptides (Three simulations with different initial structures per peptide). 3CVL-2, 3CVL-4, 3CVL-7, 3CVLri-2, 3CVLri-4 and 3CVLri-7. Run1 (green), run2 (blue) and run3 (pink) are shown.

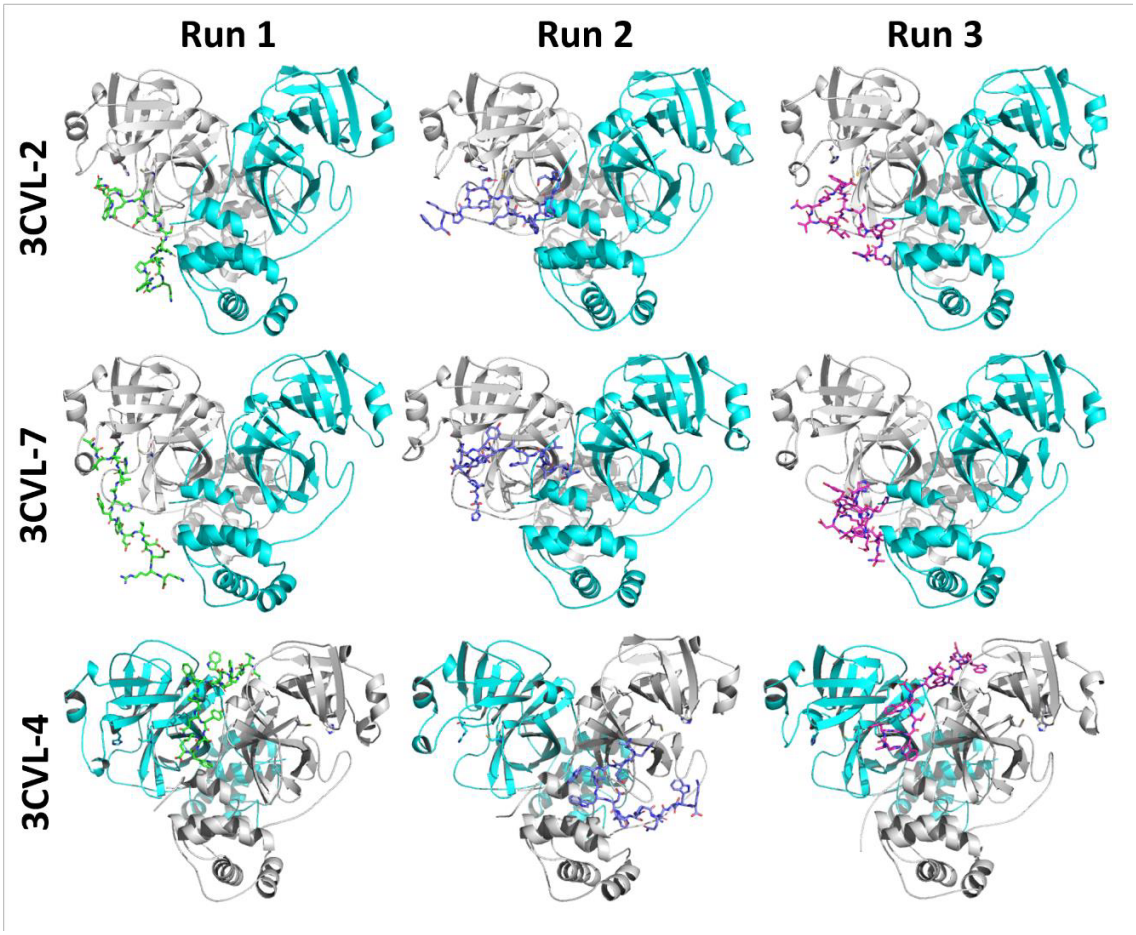


Figure S22. Representative 3CL^{pro}-L-peptide complex structures related to three simulations with different initial structures per peptide

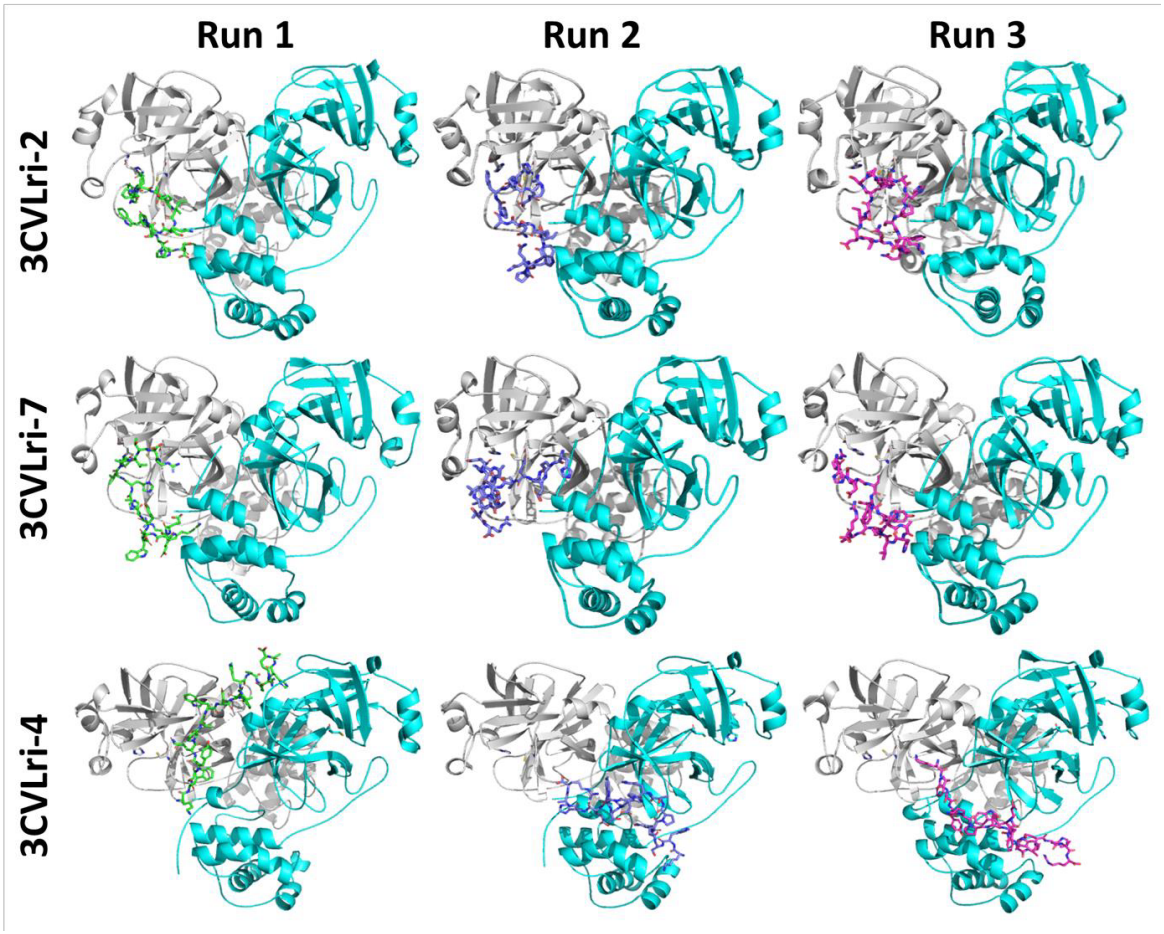


Figure S23. Representative 3CL^{pro}-D-retro-inverse peptide complex structures related to three simulations with different initial structures per peptide

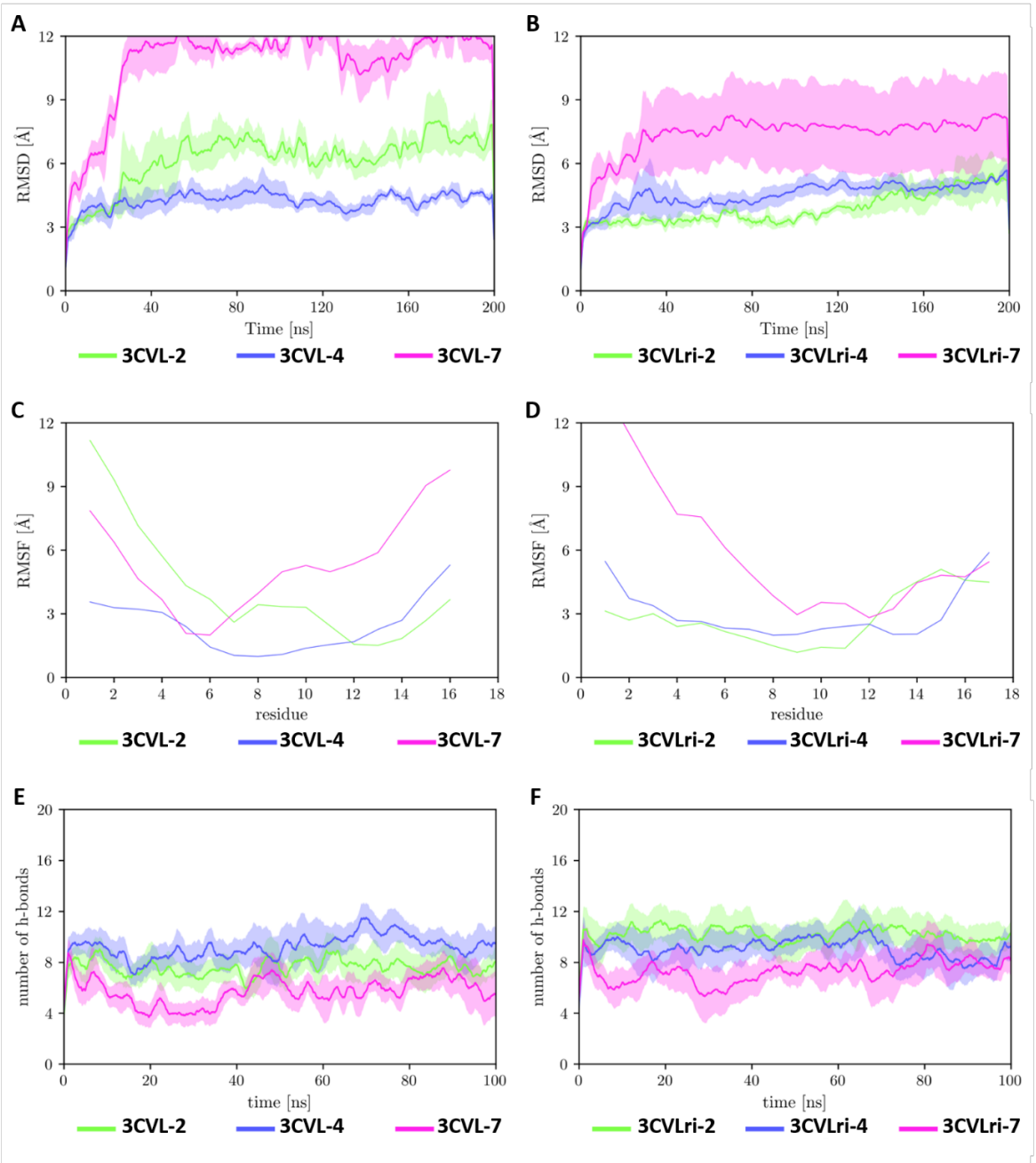


Figure S24. RMSD values, RMSF values and Hbonds between peptides and protein for 3CL^{pro} in complex with the 3CVL and 3CVLri peptides after clustering of the non-flexible/interacting peptide region. (A) RMSD of 3CVL-2, -4 and -7. **(B)** RMSD of 3CVLri-2, ri-4 and ri-7. **(C)** RMSF of 3CVL-2, -4 and -7. **(D)** RMSF of 3CVLri-2, ri-4 and ri-7. **(E)** Hbonds between the 3CL^{pro} dimer and 3CVL-2, -4 and -7. **(F)** Hbonds between the 3CL^{pro} dimer and 3CVLri-2, ri-4 and ri-7. The results are presented as the means ± SEM from elongated 200 ns simulations.

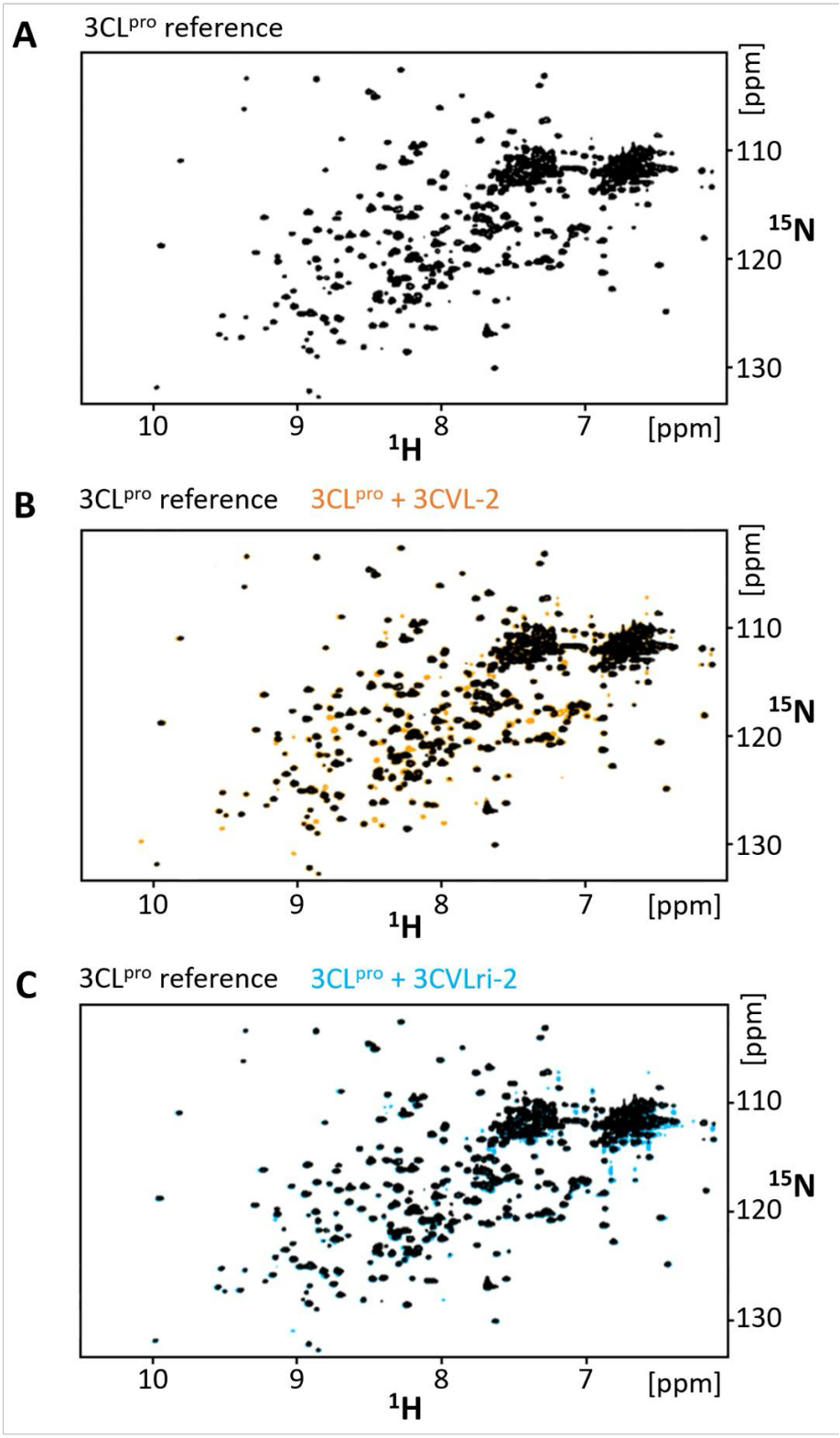


Figure S25. Two-dimensional 1H-15N correlation spectra of 100 µM 15N 3CL^{pro}, in the presence and absence of an equimolar amount (1:1 ratio) of 3CVL-2 or 3CVLri-2, respectively. (A) 3CL^{pro} reference spectrum. **(B)** 3CL^{pro} in the presence of an equimolar amount of 3CVL-2 (orange) overlayed with the reference spectrum (black). **(C)** same as **(B)** but with 3CVLri-2 (blue).

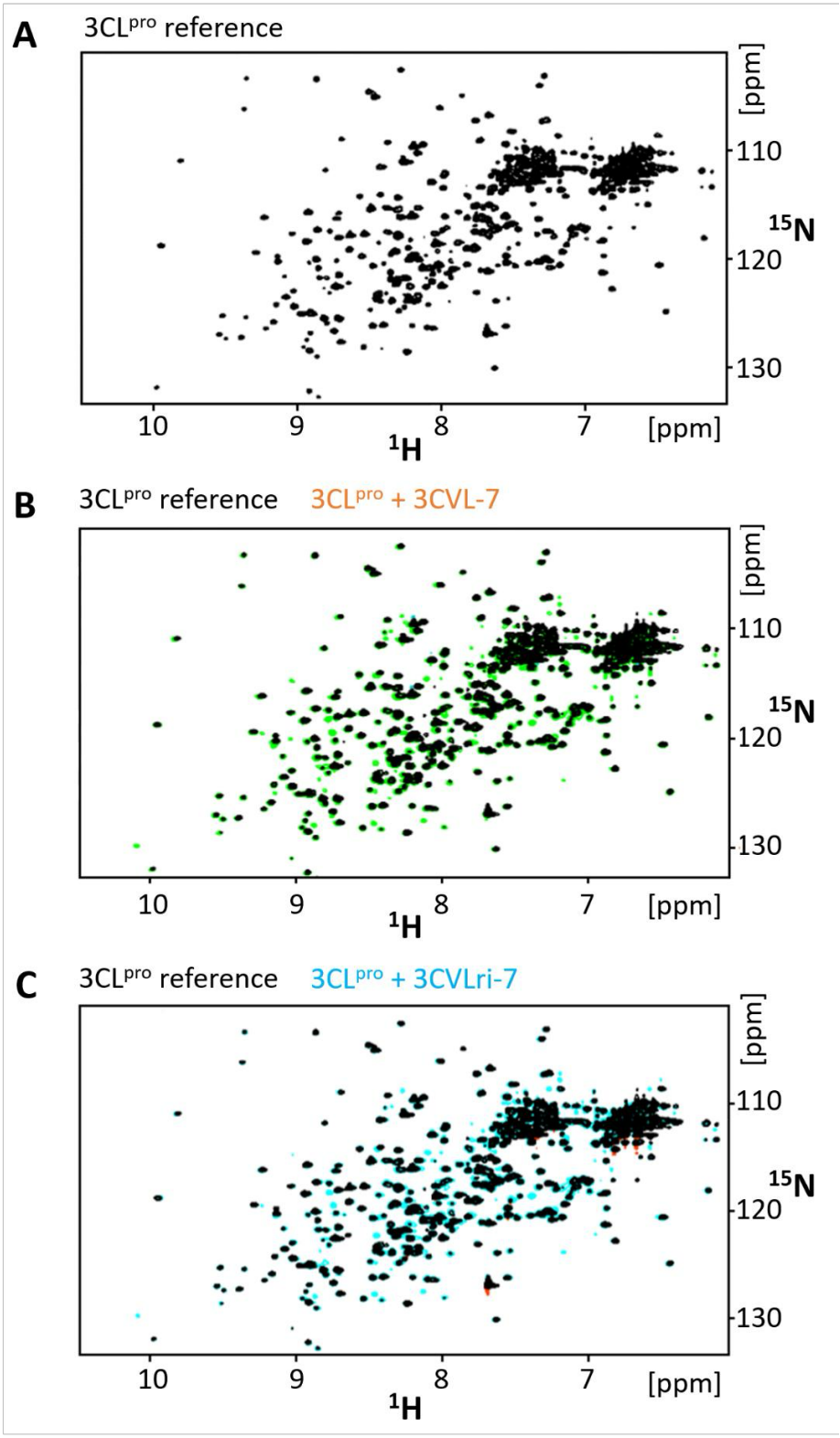


Figure S26. Two-dimensional 1H-15N correlation spectra of 100 µM 15N 3CL^{pro}, in the presence and absence of an equimolar amount (1:1 ratio) of 3CVL-7 or 3CVLri-7, respectively. (A) 3CL^{pro} reference spectrum. **(B)** 3CL^{pro} in the presence of an equimolar amount of 3CVL-7 (green) overlayed with the reference spectrum (black). **(C)** same as **(B)** but with 3CVLri-7 (blue).

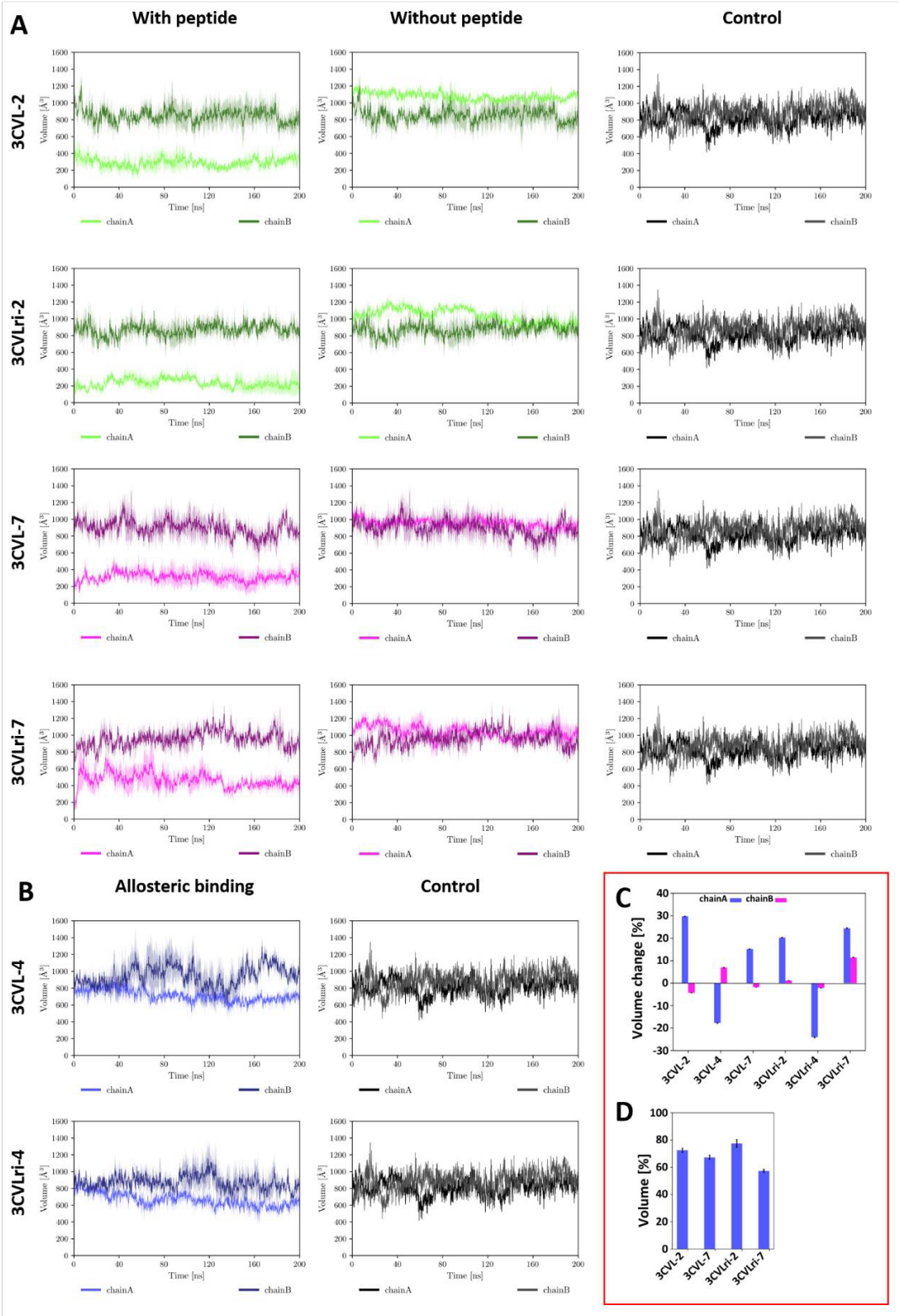


Figure S27. Volume of the SARS-CoV-2 3CL^{pro} active site with and without the D-ri-peptides. Changes of the 3CL^{pro} active sites in chain A and chain B over elongated 200 ns simulations, with and without peptide in the active site. **(A)** 3CVL-2, 3CVLri-2, 3CVL-7 and 3CVLri-7. **(B)** 3VL-4 and 3CVLri-4. **(C)** Overall volume changes of the active site with interacting peptide. **(D)** Volume change of the active site without interacting competitive peptide.

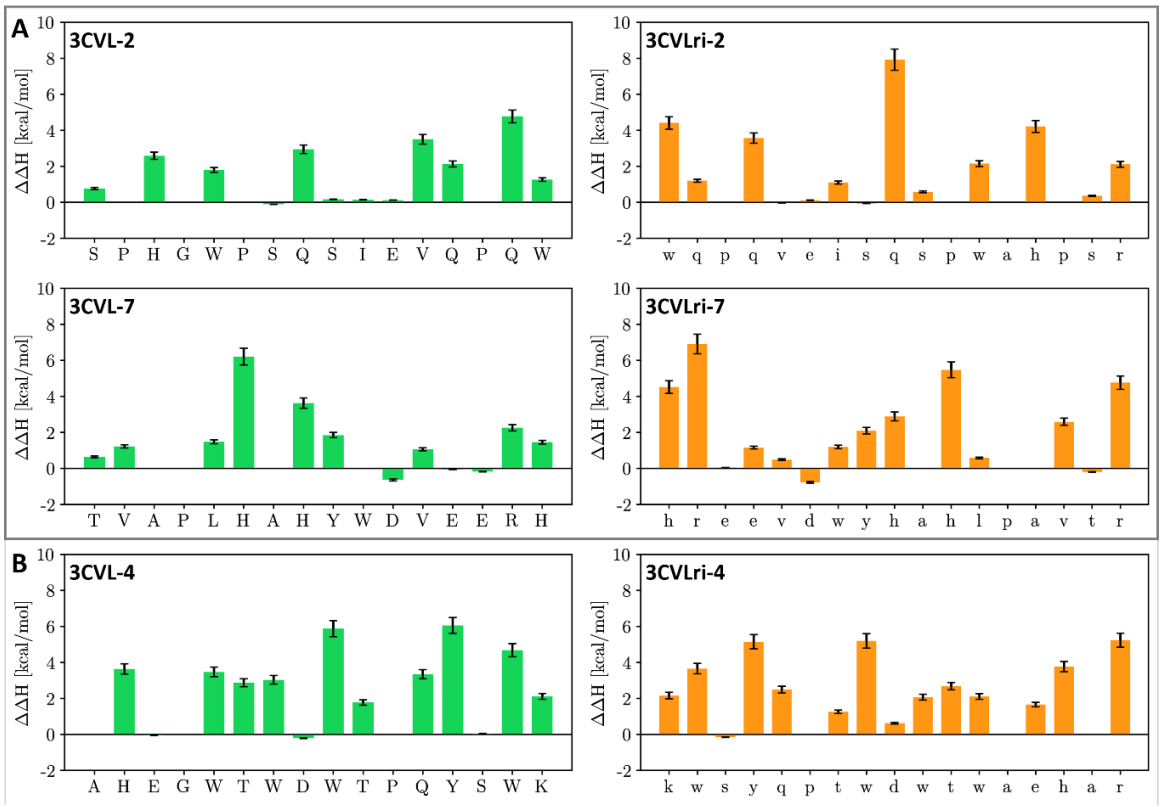


Figure S28. Computational alanine scanning of 3CVL and 3CVLri peptides to determine important residues for the interaction with 3CL^{pro}. Each residue of the 3CVL and 3CVL-ri peptides (Except Gly and Pro) was substituted by Ala and the enthalpy of binding difference $\Delta\Delta H^{X->A}$ upon alanine mutation was calculated. Alanine mutations that result in positive ΔΔH values relative to the parent structure suggest loss of binding. This outcome indicates the mutated amino acid was involved in an overall favorable network of electrostatic interactions in the parent structure. Mutations that result in negative ΔΔH suggest gain of binding. This outcome indicates the mutated amino acid was involved in an overall unfavorable network of electrostatic interactions in the parent structure. **(A)** Competitive inhibitors (3CVL-2, 3CVLri-2, 3CVL-7 and 3CVLri-7). **(B)** Noncompetitive inhibitors (3CVL-4 and 3CVLri-4).

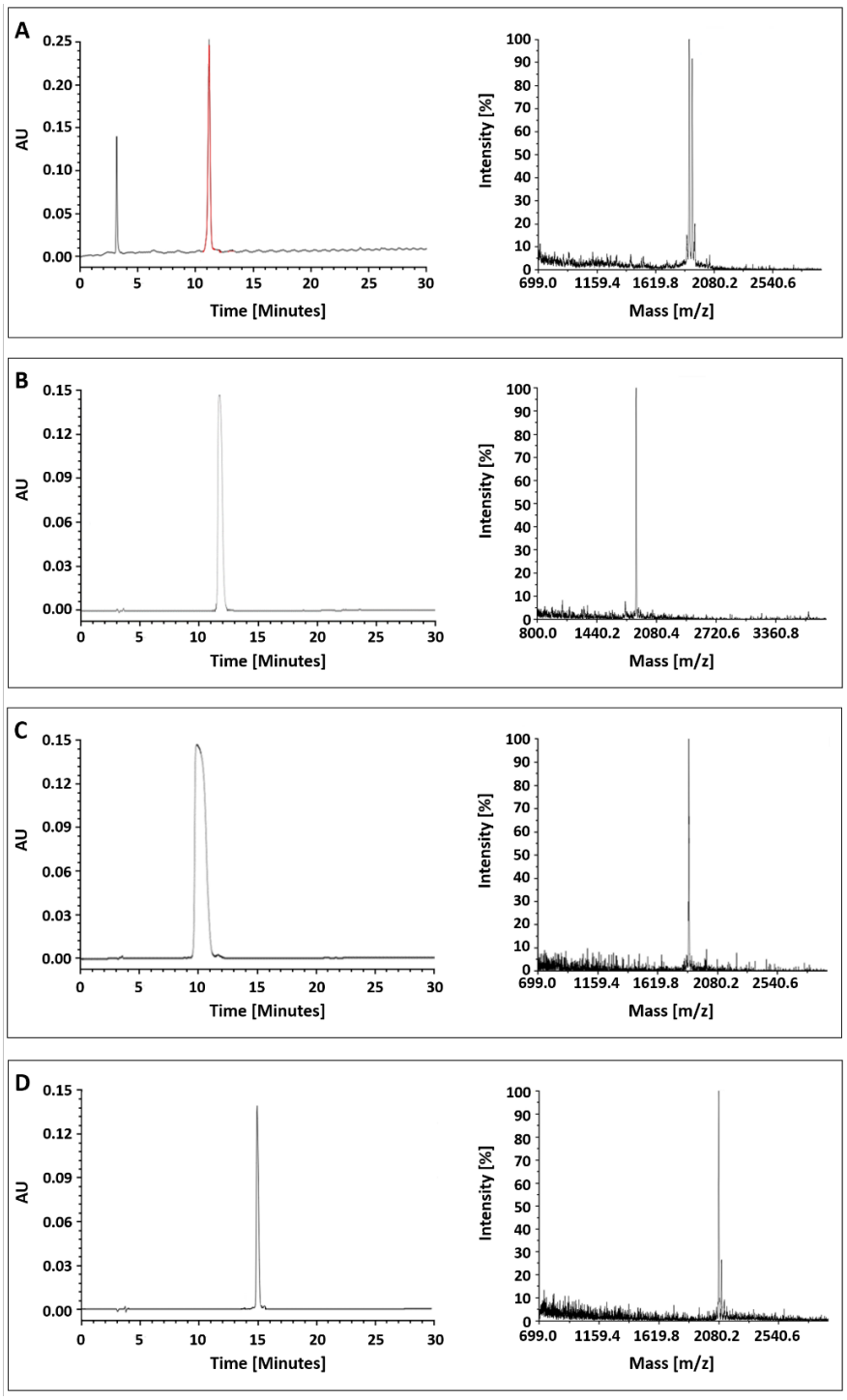


Figure S29. HPLC chromatogram and mass spectrometry of 3CVL peptides 1-4. Information obtained from the HPLC and mass spectrometry certificate from Caslo (Lyngby, Denmark). **(A)** 3CVL-1, **(B)** 2CVL-2, **(C)** 3CVL-3 and **(D)** 3CVL-4.

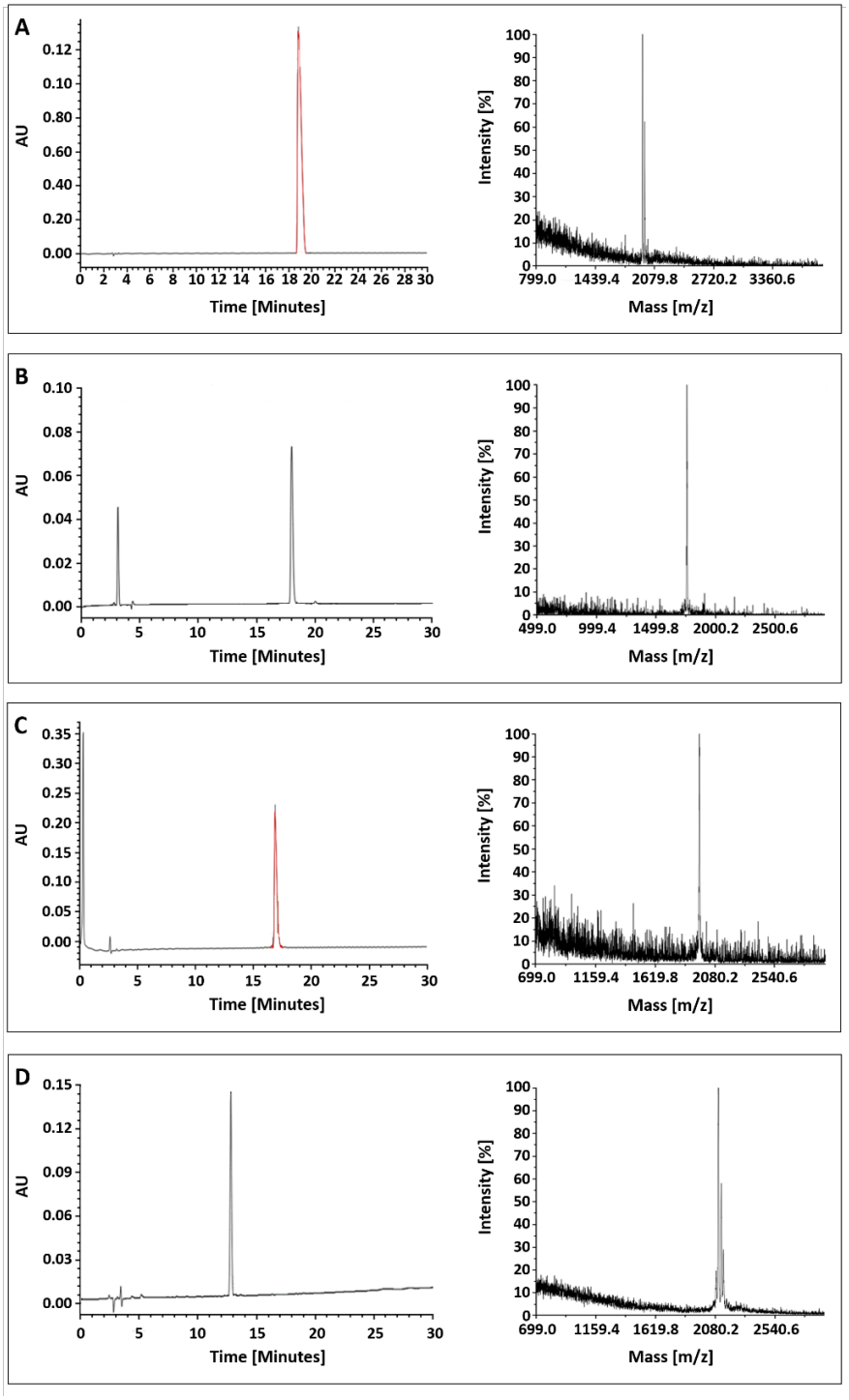


Figure S30. HPLC chromatogram and mass spectrometry of 3CVL peptides 5-8. Information obtained from the HPLC and mass spectrometry certificate from Caslo (Lyngby, Denmark). **(A)** 3CVL-5, **(B)** 2CVL-6, **(C)** 3CVL-7 and **(D)** 3CVL-8.

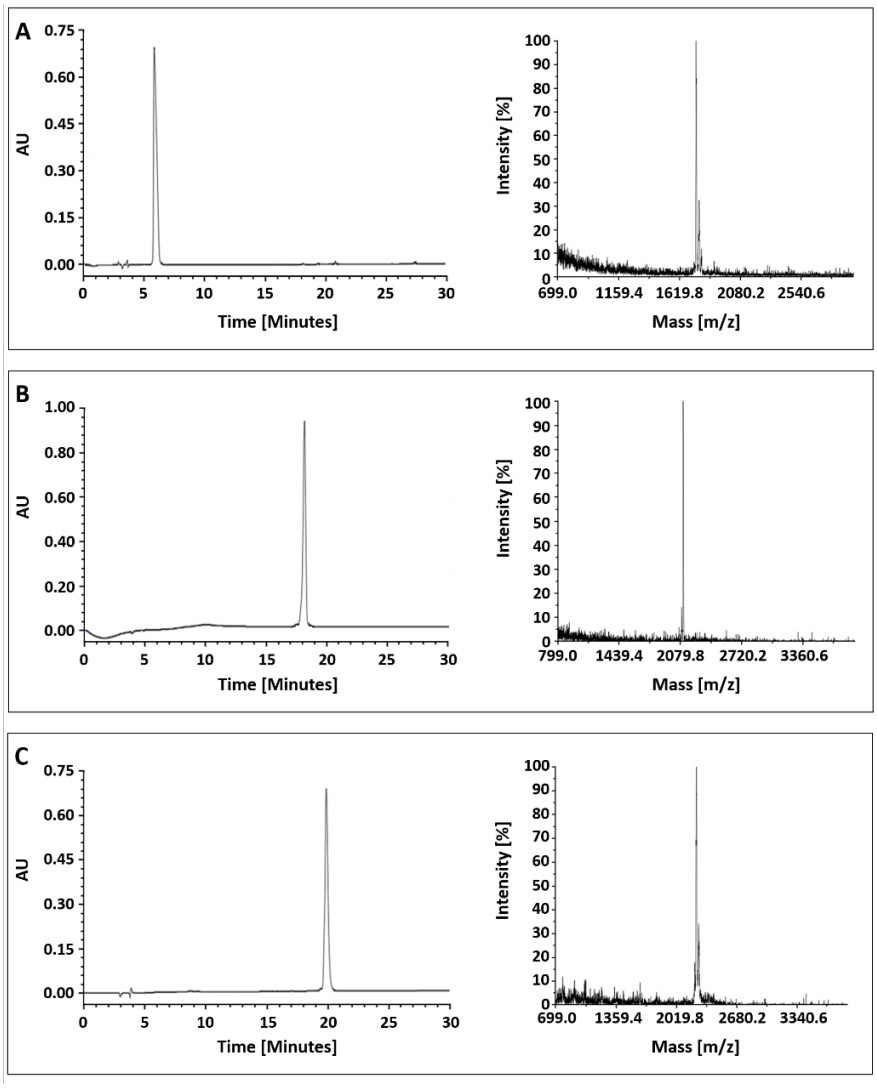


Figure S31. HPLC chromatogram and mass spectrometry of 3CVL peptides 9-11. Information obtained from the HPLC and mass spectrometry certificate from Caslo (Lyngby, Denmark). **(A)** 3CVL-9, **(B)** 2CVL-10 and **(C)** 3CVL-11.

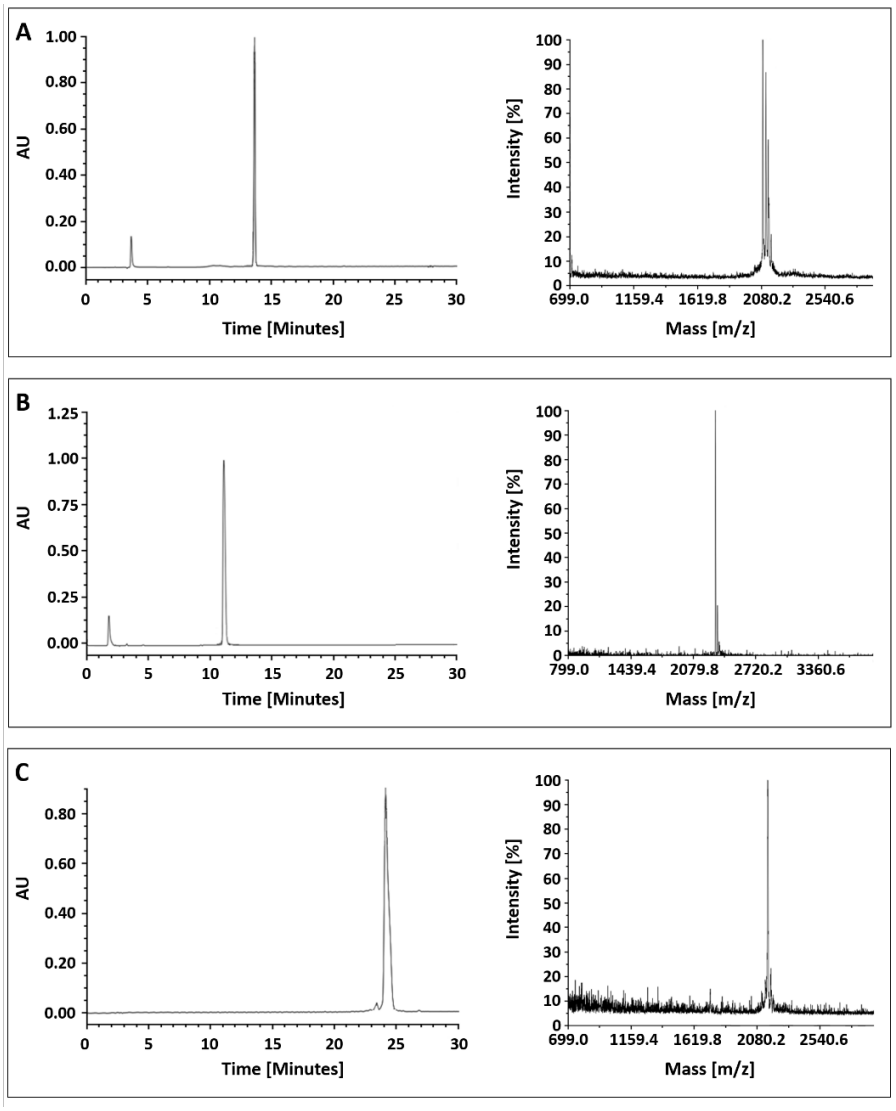


Figure S32. HPLC chromatogram and mass spectrometry of 3CVLri peptides. Information obtained from the HPLC and mass spectrometry certificate from Caslo (Lyngby, Denmark). **(A)** 3CVLri-2, **(B)** 2CVLri-4 and **(C)** 3CVLri-7.

Table S1. List of selected 3CL^{pro} phage display peptides.

Peptide	S[1]	Sequence	EmS[2]	EnF[3]	Cl ID[4]	Cl size total [5]	Cl size unique [6]
3CVL-1	A3	ANYKMTNNWDALTWAS	205.59	672.43	0	172	59
3CVL-2	A3	SPHGWPSQSIEVQPQW	79.84	65.28	4	145	69
3CVL-3	A3	SHNQEKSTYDHRMPE	75.85	41.35	6	73	18
3CVL-4	A3	AHEGWTWDWTPQYSWK	51.90	84.87	769	385	164
3CVL-5	A3	WADALNHVPKNNWPMR	47.91	52.23	14	74	25
3CVL-6	B2	INLSPKRIMPHPLFQH	984.03	107.14	0	1482	224
3CVL-7	B2	TVAPLHAHYWDVEERH	408.56	22.96	1	329	78
3CVL-8	B2	QPWEHRAQPQWQQPMR	225.87	56.21	8	761	259
3CVL-9	B2	HLSSASSPNANHHMQK	224.21	130.20	10	185	40
3CVL-10	B2	QPYMNGMHWDKTQWMR	207.60	60.28	196	720	192
3CVL-11	B2	DHLSWQQNWYWAWQYE	181.03	315.37	20	604	242

[1] Selection round
[2] Empty score value
[3] Enrichment factor value
[4] Cluster ID
[5] Total cluster size [AA]
[6] Unique cluster size [AA]

Table S2. Evaluation of 3CVL-2, -4 and -7 sequences with TSAT.

		TSAT analysis									
Peptide	Sequence	S[1]	EmS[2]	EmS-P.[3]	EnF[4]	EnF-P.[5]	TS3 [ppm]		ES3 [ppm]	DC3 [ppm]	Lib [ppm]
3CVL-2	SPHGWPSQSIEVQPQW	A	80	5	65	29	161	2	24	2	
3CVL-4	AHEGWTWDWTPQYSWK	B	105	14	85	20	105	2	32	1	
3CVL-7	TVAPLHAHYWDVEERH	B	409	2	23	428	877	2	712	38	

[1] Selection
[2] Empty score value
[3] Empty score position
[4] Enrichment factor value
[5] Enrichment factor position

Table S3. Evaluation of 3CVL-2, -4 and -7 sequences with Hammock.

		Hammock analysis					
Peptide	Sequence	S[1]	Cl size total[2]	Cl size unique[3]	unique S[4]	total S[5]	total A[6]
3CVL-2	SPHGWPSQSIEVQPQW	A	145	69	62649	100128	191
3CVL-4	AHEGWTWDWTPQYSWK	A	385	164	62649	100128	191
3CVL-7	TVAPLHAHYWDVEERH	B	329	78	54615	129860	162

[1] Selection
[2] Total cluster size [AA]
[3] Unique cluster size [AA]
[4] Unique sequences
[5] Total sequences
[6] Total alignments

Table S4. Contributing amino acids in the Substrate-binding site and the proposed allosteric site of SARS-CoV-2 3CL^{pro}.

Substrate subsite[2]		Allosteric site[3]
S1`	His41	Lys12
	Gly143	Lys97
	Ser144	Lys100
	Cys145	Tyr101
		Lys102
S1	Phe140	Phe103
	Leu141	Val104
	Asn142	Arg105
	His163	
	Glu166	
S2	His41	
	Met49	
	Tyr54	
	His164	
	Asp187	
	Arg188	
S3	Met165	
	Glu166	
	Leu167	
	Gln189	
	Thr190	
	Gln192	
Oxyanion hole	His41	
	Cys145	

Table S5. Purity and molecular weight of purchased L- and D-ri-peptides.

Peptide	Purity in [%]	Theoretical MW	MW [M+H$^+$] in main peak	MW
3CVL-1	98.44	1885.10	1886.27	1885.27
3CVL-2	99.73	1862.05	1862.67	1861.67
3CVL-3	99.53	1857.99	1859.33	1858.33
3CVL-4	98.15	2077.26	2078.73	2077.73
3CVL-5	99.95	1948.26	1949.32	1948.32
3CVL-6	95.24	1927.37	1928.39	1927.39
3CVL-7	98.92	1959.17	1960.43	1959.43
3CVL-8	99.02	2102.39	2104.33	2103.33
3CVL-9	98.53	1744.92	1745.83	1744.83
3CVL-10	98.45	2108.47	2109.08	2108.08
3CVL-11	99.22	2239.41	2240.27	2239.27
3CVLri-2	99.41	2090.30	2091.06	2090.06
3CVLri-4	98.95	2305.52	2306.44	2305.44
3CVLri-7	98.09	2157.40	2158.67	2157.67

Text S1

Selection of SARS-CoV-2 $3CL^{pro}$ binding peptides

To select binding peptides to the main protease of SARS-CoV-2, M13 phage display selection was performed on an amino plate in two independent selections (A and B) with three consecutive selection rounds, each other. The resulting output titer is depicted in Supplementary Fig. S2. While the titer of target selection (TS), empty selection (ES) and direct control (DC) of selection A did not show any constant increase during the consecutive selection rounds (Supplementary Fig. S2A), this effect was observed for all selections B (Supplementary Fig. S2B). The most significant output titer increases could be observed for ES, which enriches many non-target binding peptides. The increased titer within the TS and DC might indicate the enrichment of target binding peptides. Titer developments, as observed in selection A are not necessarily proof of failed selections as experienced in the past. We thus continued with the evaluation of both selections.

Enrichment ELISA has additionally been performed to estimate the selection success further. For selection A, decreasing absorption signal could be observed with increasing selection round of the TS (Supplementary Fig. S3A). The signal of the wells without a target is significantly smaller in all cases and points to the presence of $3CL^{pro}$ binding peptides. The signal distribution of selection B seems to be random during the selection rounds (Supplementary Fig. S3B). The higher signal intensity of the target-containing wells was consistently observed during selection. The signal intensity of the control wells without phages is higher in the wells with the target than those without the target. It indicates a low background signal due to the unspecific interaction of the antibody to 3CL protease.

Analysis of sequences derived from NGS sequencing

The most convincing and reliable way to evaluate the selection process is the analysis of the NGS results. The sequences filtered from the TSAT software[4] are ordered according to their empty score, which values the number of a sequence observed in the last round of TS compared to its abundance in the respective round of ES. We also judged each sequence with the help of the so-called enrichment factor, which values the number of a sequence observed in the last TS round compared to its abundance in the library. Hammock analysis[5] with the filtered sequences gave additional information about the cluster size and thus about the number of selected sequences that exhibit high similarity to a sequence of interest.

To choose peptides for further analysis, we considered the empty score and the enrichment factor and the cluster sizes. Considering the described parameters, 11 L-peptide sequences, named 3CVL-1 to 3CVL-11, were chosen for further inhibition characterisation (Supplementary table S1).

Supporting references

[1] 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.

[2] Tang, B.; He, F.; Liu, D.; Fang, M.; Wu, Z.; Xu, D. AI-aided design of novel targeted covalent inhibitors against SARS-CoV-2. *BioRxiv* **2020**, 2020.03.03.972133

[3] Eberle, R.J.; Olivier, D.S.; Amaral, M.S.; Gering, I.; Willbold, D.; Arni, R.K.; Coronado, M.A. The Repurposed Drugs Suramin and Quinacrine Cooperatively Inhibit SARS-CoV-2 $3CL^{pro}$ *In Vitro*. *Viruses.* **2021,** *13,* 873.

[4] Santur, K.; Reinartz, E.; Lien, Y.; Tusche, M.; Altendorf, T.; Sevenich, M.; Tamgüney, G.; Mohrlüder, J.; Willbold, D. Ligand-Induced Stabilisation of the Native Human Superoxide Dismutase 1. *ACS Chem. Neurosci.* **2021,** *12,* 2520-2528.

[5] Krejci, A.; Hupp, T.R.; Lexa, M.; Vojtesek, B.; Muller, P. Hammock: A hidden Markov model-based peptide clustering algorithm to identify protein-interaction consensus motifs in large datasets. *Bioinformatics.* **2016,** *32,* 9–16.

Molecular Imaging and Biology (2024) 26:81–89
https://doi.org/10.1007/s11307-023-01881-y

RESEARCH ARTICLE

The Theranostic Optimization of PSMA-GCK01 Does Not Compromise the Imaging Characteristics of [^{99m}Tc]Tc-PSMA-GCK01 Compared to Dedicated Diagnostic [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA in Prostate Cancer

Eduards Mamlins[1] · Lara Scharbert[2,3] · Jens Cardinale[1,4] · Maria Krotov[1] · Erik Winter[4] · Hendrik Rathke[4,5] · Birgit Strodel[2,3] · Alfred O. Ankrah[6] · Mike Sathekge[7] · Uwe Haberkorn[4,8,9] · Clemens Kratochwil[4] · Frederik L. Giesel[1,4]

Received: 6 October 2023 / Revised: 22 November 2023 / Accepted: 24 November 2023 / Published online: 8 December 2023
© The Author(s) 2023

Abstract

Purpose Radiolabeled PSMA-ligands play a major role in today's nuclear medicine. Since approval of [^{177}Lu]Lu-PSMA-617 for therapy of metastatic prostate cancer, availability of ^{177}Lu became bottleneck of supply due to the high demand. Recently, a theranostic PSMA-ligand, PSMA-GCK01, was developed which can be labeled either diagnostically with ^{99m}Tc or therapeutically with ^{188}Re with both nuclides available from well-known generator systems. This novel tracer might aid to overcome aforementioned supply limitations. In this investigation, the biodistribution and general imaging characteristics of [^{99m}Tc]Tc-PSMA-GCK01 were compared with the diagnostic reference compound [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA in patients with advanced stage prostate cancer. In addition, the binding of both ligands to PSMA was analyzed at the molecular level using molecular docking.

Procedures Two cohorts ($n = 19$ vs. $n = 21$) of patients with metastatic castration-resistant prostate cancer matched for age, tumor stage, and Gleason score underwent a planar gamma camera imaging with [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA or [^{99m}Tc]Tc-PSMA-GCK01 prior to PSMA-ligand therapy for PSMA-phenotyping. The imaging data were retrospective analyzed for salivary gland, kidney, liver, soft tissue, and tumor uptake on a semi-automated ROI-analysis using HERMES Medical Solutions AB (HMS, Sweden).

Results The data sets were semi-automated quantified on a ROI-based analysis. The tumor-to-background presented equal results of [^{99m}Tc]Tc-PSMA-GCK01 compared to [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA. The physiological PSMA-positive organs like salivary gland presented also equal uptake in counts/MBq (salivary gland median 9.48 [^{99m}Tc]Tc-PSMA-GCK01 vs. median 9.11 [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA), while liver-to-kidney ratio presented a slight shift to the liver parenchyma using [^{99m}Tc]Tc-PSMA-GCK01 (0.83) compared to [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA (0.55) with no statistical significance. This is in agreement with the results from the docking study revealing only a minor difference in the docking scores for both ligands.

Conclusions The novel theranostic tracer [^{99m}Tc]Tc/[^{188}Re]Re-PSMA-GCK01 demonstrates comparable general imaging characteristic with the reference compound [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA. These results pave the way for the PSMA-targeting imaging and theranostic agents for a broader, rather low-cost, generator applied radio-ligand therapy utilization.

Keywords PSMA-GCK01 · Technetium-99 m · SPECT · Theranostic · Prostate cancer

Introduction

The development of prostate-specific membrane antigen (PSMA)-inhibitor tracers has revolutionized the imaging of prostate cancer. PSMA PET tracers [^{68}Ga]Ga-PSMA-11, [^{18}F]PSMA-1007, and [^{18}F]DCFPyL demonstrate

✉ Eduards Mamlins
eduards.mamlins@med.uni-duesseldorf.de

Extended author information available on the last page of the article

high sensitivity of approx. 50% at PSA values as low as 0.2–0.5 ng/mL and reach their plateau sensitivity of approx. 90% from PSA > 1 ng/mL [1–3]. In contrast, ^{99m}Tc labeled SPECT imaging PSMA agents reach their highest sensitivity not before PSA 2–5 ng/mL [4–6]. Consequently, PSMA-PET is considered the standard of reference for primary staging of high-risk prostate cancer (PC) and biochemical recurrence. However, for response assessment of advanced stage patients [7] or tailoring metastatic castration-resistant prostate cancer (mCRPC) patients with or without tumor PSMA-expression to PSMA-targeted radioligand therapy (RLT) highest sensitivity is not that pivotal and cost-effectiveness and broad availability are more important [8]. It is worth mentioning in this context that a randomized phase 3 clinical study of ^{177}Lu-dotatate for midgut neuroendocrine tumors (NETTER-1 trial) used planar somatostatin receptor scintigraphy for the confirmation of somatostatin receptor positive tumor lesions [9]. Currently, most clinical experience in context of PSMA targeted scintigraphies is available for the ^{99m}Tc-tracers [^{99m}Tc]Tc-MIP1404 [10], [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA [11], and [^{99m}Tc]Tc-PSMA-I&S [12].

Recently, a novel theranostic PSMA-ligand GCK01, that can be labeled with diagnostic ^{99m}Tc and therapeutic ^{188}Re, has been introduced [13]. The generator-based beta-emitter ^{188}Re may aid to overcome supply limitations of ^{177}Lu and decrease the production costs for PSMA-RLT in comparison to ^{177}Lu. In contrast to the previous Tc-labeled ligands, GCK01 had to be refined with regard to labeling yield under more challenging rhenium conditions as well as in terms of tracer uptake in the dose-limiting organs. In addition, the *in vivo* stability has been improved, which represents a pivotal cornerstone for the therapeutic approach.

The aim of this work is to evaluate whether the optimizations of PSMA-GCK01 for the therapeutic application might have introduced an unfavorable trade-off regarding biodistribution in normal organs and general imaging characteristics.

Materials and Methods

Patient Population

We retrospectively included two cohorts of 21 consecutive patients with mCRPC who were considered to obtain systemic therapy with [^{177}Lu]Lu-PSMA-617 and underwent whole body scans with [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA or [^{99m}Tc]Tc-PSMA-GCK01 in the Department of Nuclear Medicine of University Hospital Heidelberg between 01/2021 and 02/2022. Two subjects from [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA cohort were censored due to ineligibility for further quantitative assessment because of improper image acquisition with motion artifacts. Informed consent was obtained from all the patients. The exams were conducted according to national regulatory on the basis of compassionate use, when a PSMA-PET scan was not available within reasonable time and the patient had urgently to be evaluated regarding PSMA therapy. Prospective clinical trial registration is not required for compassionate care that is performed under individual medical indication. The Ethical Committee of the University Hospital Heidelberg approved the retrospective Evaluation (permission S-732-18).

Radiopharmaceuticals and Image Acquisition

[^{99m}Tc]Tc-EDDA/HYNIC-iPSMA was produced from commercial kits according manufacturer instructions and [^{99m}Tc]Tc-PSMA-GCK01 was synthesized as described previously [13]. Both tracers were injected via an intravenous bolus ([^{99m}Tc]Tc-EDDA/HYNIC-iPSMA 669 ± 94 MBq; [^{99m}Tc]Tc-GCK01 705 ± 30 MBq). The acquisition of planar whole body images (skull to toes) started for both tracers 120–180 min after the injection on a Siemens gamma camera system (Symbia, Siemens Healthineers) with low-energy high-resolution collimation, a 128 × 128 matrix and 15 cm/min.

Image Analysis

Round regions of interest (ROIs) (1.6 cm^2) were placed in the areas with physiological tracer uptake (parotid gland (Gl. parotis) and kidneys pairwise, liver, and in the right thigh as a background) in conjugate view of planar anterior and posterior whole-body scans of the respective subjects and the geometric mean was calculated (HERMES Medical Imaging Suite v6.1, HERMES Medical Solutions AB, Strandbergsgatan 16, 112 51 Stockholm, Sweden). In addition to the measurement of physiological biodistribution in the healthy organs, three random metastases were visually identified and quantified in the same way.

Statistical Analysis

We performed the analysis of tracer uptake (measured counts divided by the injected activity in MBq), tumor-to-background, as well as liver-to-kidney ratios with descriptive statistics and box-plot analysis using SigmaPlot version 14.0 (Systat Software, Inc., San Jose, CA, USA).

Molecular Docking

The crystal structure of PSMA was retrieved from the Protein Data Bank (PDB #5O5T) and processed, i.e., hydrogen atoms were added to the protein, ligand, and water molecules, using the protein preparation wizard of Maestro, included in the Schrödinger 2022–3 package.

To simulate the environment of the binding site as realistically as possible, we kept the ions and all water molecules located in the buried part of the binding pocket. The compounds [^{99m}Tc]Tc-PSMA-GCK01 and [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA were drawn using Marvin Sketch version 21.9 (https://chemaxon.com/marvin) and exported in 3D to pdb format. The following preparation and docking steps were performed with the software of the ADFR suite version 1.0 (https://ccsb.scripps.edu/adfr/): First, Gasteiger charges were added to the protein, including the ions Zn^{2+}, Ca^{2+}, and Cl^{-} and the water molecules, using the prepare_receptor script. Then the docking grid was created with AGFR using the co-crystallized ligand 9OT to determine the size and position of the grid box. The prepare_ligand script from the ADFR suite was used to add Gasteiger charges and to define rotatable bonds of the compounds [^{99m}Tc]Tc-PSMA-GCK01 and [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA. Since Tc and Re are unfamiliar elements for Autodock, we changed the atom type Tc to Mn in the pdbqt files without affecting any atomic coordinates and assigned a charge of + 2 to Mn. Finally, we performed the docking run with 120 searches and a maximum number of 5 million evaluations per search using ADFR.

Results

Apart from the data of the two [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA patients that had to be removed, all other data sets were analyzed the way described above. Our subjects were matched in regard to age ([^{99m}Tc]Tc-EDDA/HYNIC-iPSMA 71 ± 10 years; range 55–94 years or [^{99m}Tc]Tc-PSMA-GCK01 73 ± 8 years; range 59–84 years), tumor stage, and total Gleason score so that we assume here a comparability of both cohorts (Table 1).

Both PSMA ligands demonstrated the highest physiological uptake in counts/MBq in the salivary glands (median 9.48 [^{99m}Tc]Tc-PSMA-GCK01 vs. median 9.11 [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA), followed by kidneys (median 8.64 [^{99m}Tc]Tc-PSMA-GCK01 vs. median 10.20 [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA), and the liver (median 7.14 [^{99m}Tc]Tc-PSMA-GCK01 vs. median 5.61 [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA) (Fig. 1).

The uptake was equal in the salivary glands (see above) and in background (counts/MBq median 0.47 [^{99m}Tc]Tc-PSMA-GCK01 vs. median 0.44 [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA). But the liver-to-kidney ratio presented a slight shift to the liver parenchyma for [^{99m}Tc]Tc-PSMA-GCK01 (0.83) compared to [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA (0.55), however, with no statistical significance. The uptake in tumor lesions as well as the tumor-to-background ratio presented comparable results for [^{99m}Tc]Tc-PSMA-GCK01 and [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA, however, with a slightly higher uptake for [^{99m}Tc]Tc-PSMA-GCK01 ([[^{99m}Tc]Tc-PSMA-GCK01 uptake in tumor lesions in counts/MBq for tumor lesion (TL) I: 4.45; TL II: 4.43; TL III: 3.82 with appropriate tumor-to-background ratio of TL I 9.47; TL II 9.43; TL III 8.12]; [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA uptake in tumor lesions in counts/MBq for TL I: 3.69; TL II: 3.74; TL III: 3.20 with appropriate tumor-to-background ratio of TL I 8.36; TL II 8.47; TL III 7.26]) (Figs. 2 and 3).

Table 1 Patient characteristics of cohort A ([^{99m}Tc]Tc-PSMA-GCK01) and cohort B ([^{99m}Tc]Tc-EDDA/HYNIC-iPSMA)

Parameters	Cohort A [^{99m}Tc]Tc-PSMA-GCK01	Cohort B [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA
Age (years)		
Mean ± SD	73 ± 8	71 ± 10
Max	84	94
Min	59	55
PSA (ng/mL)		
Mean ± SD	276.1 ± 613.3	279.2 ± 387.2
Max	2866.8	1399.0
Min	0.1	1.7
Gleason score		
Median (range)	7 (7–10)	9 (6–10)
LDH (U/L)		
Mean ± SD	385.2 ± 195.6	349.8 ± 231.3
Max	957.0	972.0
Min	192.0	185.0
AP (U/L)		
Mean ± SD	267.4 ± 370.5	399.0 ± 487.4
Max	1500.0	1799.0
Min	41.0	31.0
NAAD[a]		
None	1	1
1 cycle	10	6
2 cycle	9	11
Taxane		
None	3	4
1 cycle	6	6
2 cycle	11	8
PSMA positive tumor burden		
Low (< 10 lesions)	8	4
Medium (10–100 lesions)	4	6
High (> 100 lesions or bmc[b])	9	9

[a]Neoadjuvant androgen deprivation therapy

[b]Bone marrow carcinosis

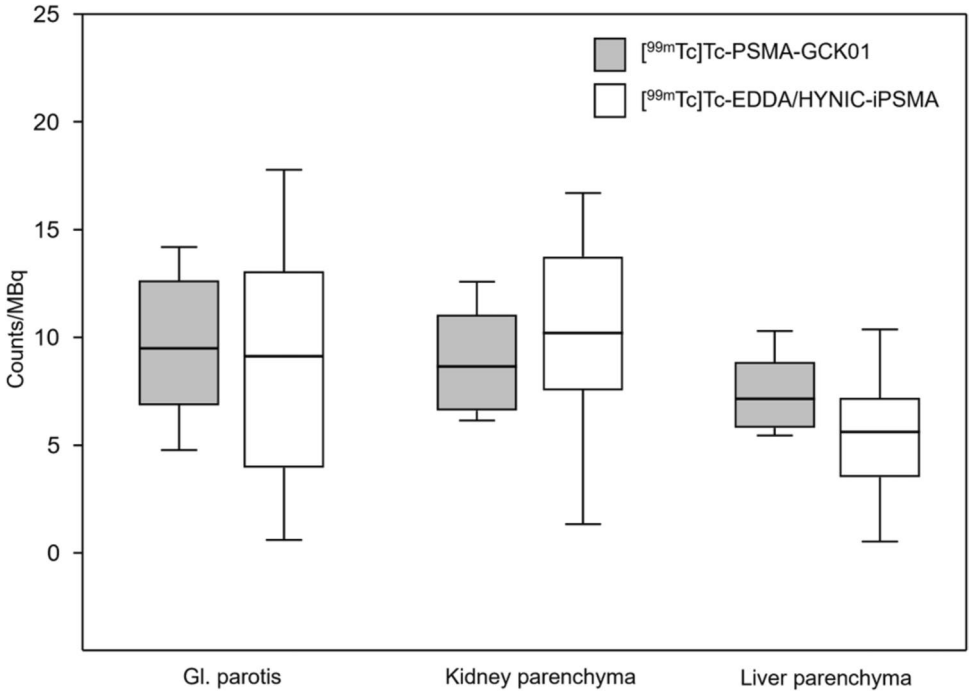


Fig. 1 Comparison of physiological biodistribution of [^{99m}Tc]Tc-PSMA-GCK01 and [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA. Box plots show 5th/95th IQR

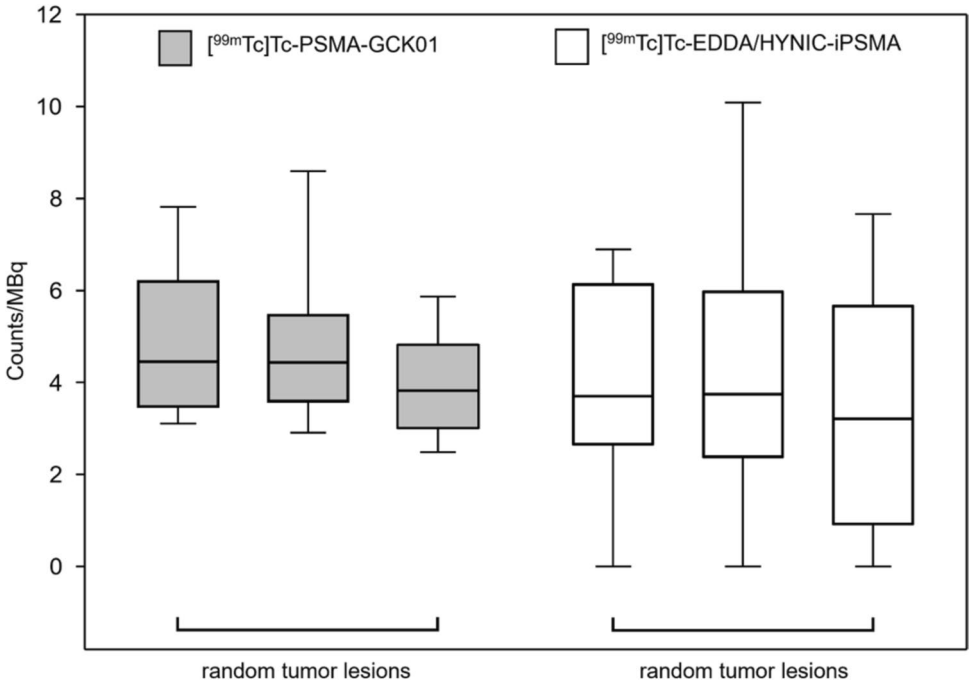


Fig. 2 Comparison of [^{99m}Tc]Tc-PSMA-GCK01 and [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA uptake in three random tumor lesions. Box plots show 5th/95th IQR

Molecular Docking Studies

In order to examine the binding of both ligands in the PSMA receptor pocket, we conducted molecular docking studies. As a measure for the binding affinity, docking scores of − 14.6 kcal/mol for [^{99m}Tc]Tc-PSMA-GCK01 and − 13.6 kcal/mol for [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA were obtained for the top-ranked conformations (Fig. 4): Both ligands are deeply embedded in the binding pocket, with the urea group interacting with the zinc(II) active site and the glutamate moiety pointing toward the glutamate pocket, while the 2-naphtylalanine occupies a lipophilic cavity and the chelators stick out of the entrance funnel (Fig. 4).

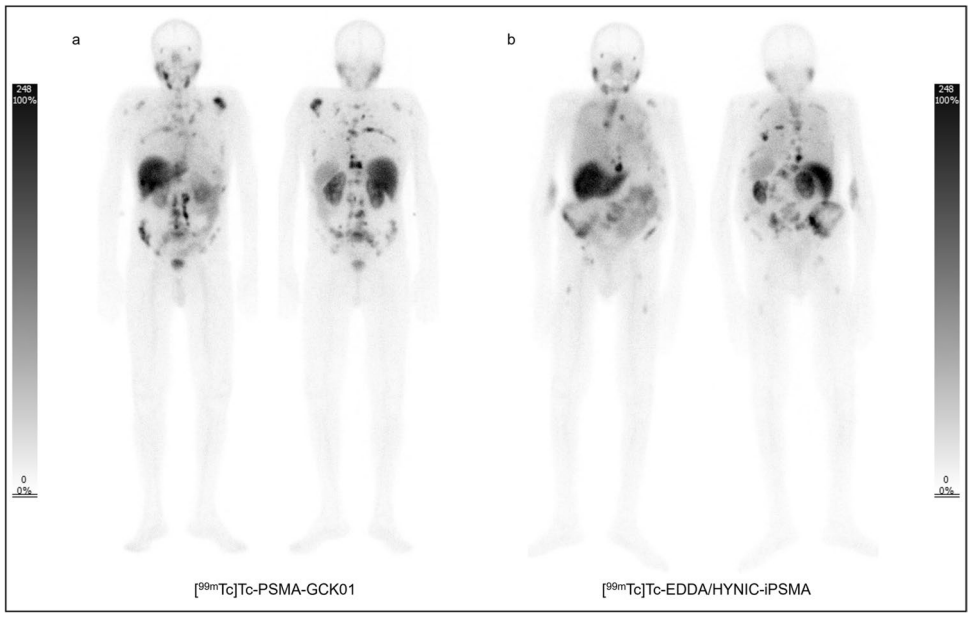


Fig. 3 Comparison of **a** [^{99m}Tc]Tc-PSMA-GCK01 and **b** [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA uptake in whole body projections in patients with high tumor mass. The uptake in metastasis of PSMA-GCK01 is not superior to that of its counterpart by less uptake in kidneys

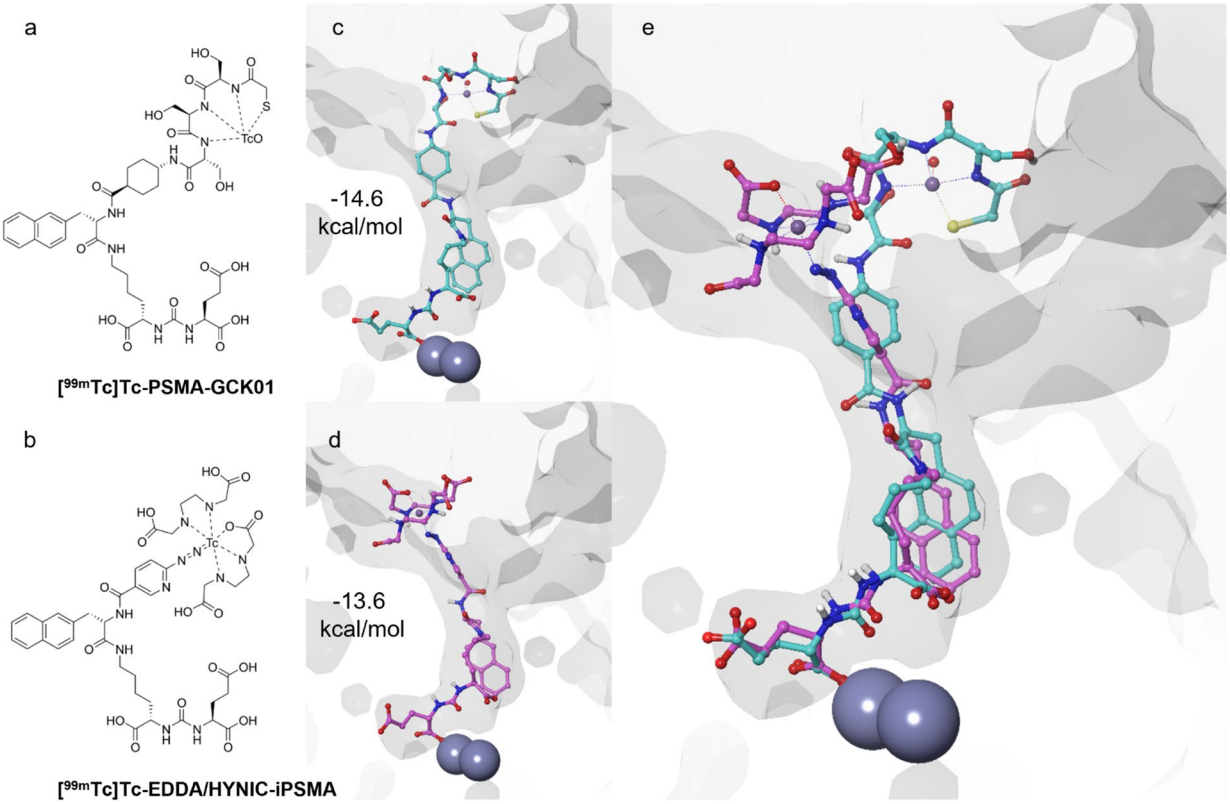


Fig. 4 Results from the docking study of [^{99m}Tc]Tc-PSMA-GCK01 and [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA with the PSMA protein; **a** chemical structure of [^{99m}Tc]Tc-PSMA-GCK01 and **b** chemical structure of [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA; **c** top-ranked docking pose and score of [^{99m}Tc]Tc-PSMA-GCK01 and **d** top-ranked docking pose and score of [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA; and **e** overlay of both top-ranked docking poses in PSMA

Discussion

Prostate cancer has an increasing incidence in the world so that there is a need of cost-effective PSMA imaging and therapy [4, 14]. Besides PSMA PET tracers, several [^{99m}Tc]Tc-labeled PSMA ligands have been developed in recent years [15]. Contrary to PET, gamma scanners are broadly available throughout the world suggesting the usefulness of compatible radioligands [6, 16].

The first human data with [^{99m}Tc]Tc-MIP-1404 and [^{99m}Tc]Tc-MIP-1405 showed very promising results [6, 17]. In 2017, Ferro-Flores et al. introduced [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA which offers a rapid blood clearance, the ability to detect prostate cancer, as well as its metastases, and has mainly kidney elimination with less liver uptake than [^{99m}Tc]Tc-MIP-1404 [18]. Further investigations with [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA demonstrated high detection rates for PSMA positive lesions, especially in the setting of biochemically recurrence of prostate cancer but a lower sensitivity compared with PET PSMA tracer [4, 19]. The main advantage of the new ^{99m}Tc-labeled tracer group is the detection of PSMA positive lesions for planning radio-ligand therapy, for instance, with [^{177}Lu]Lu-PSMA-617 [8] (Fig. 5). A specific radioligand therapy of prostate cancer can only be performed in presence of PSMA-positive phenotype of tumor lesions [20]. The sensitivity of PSMA imaging plays a subordinate role in this context, rather, the diagnostic and therapeutic compounds should be as similar as possible to predict kinetics and uptake of the latter [7, 8].

The novel PSMA ligand [^{99m}Tc]Tc-PSMA-GCK01 was optimized to provide a suitable basis for both ^{99m}Tc- and ^{188}Re-labeling, thus, providing a seamless PSMA-targeted theranostic tandem [13]. In this investigation we have analyzed the biodistribution of the diagnostic [^{99m}Tc]Tc-PSMA-GCK01 in organs with physiologic PSMA-uptake and in metastases to assess the appropriateness for imaging, especially in comparison with [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA. The ROI-based analysis showed equal tracer uptake in salivary glands and background with [^{99m}Tc]Tc-PSMA-GCK01 and [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA, respectively.

This observation is also in agreement with the results from our *in silico* docking study, revealing only minimal differences in the conformations of both molecules in the PSMA binding pocket, and, thus, only a minimal difference in the docking score of approx. 1.0 kcal/mol. Considering the high structural similarity of both PSMA ligands, differing only in their chelator moiety and slightly in the linker region, this explains our earlier findings of a slightly higher affinity of [^{99m}Tc]Tc-PSMA-GCK01 (Ki = 26 nm) toward PSMA when compared to [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA (Ki = 38 nm) [13].

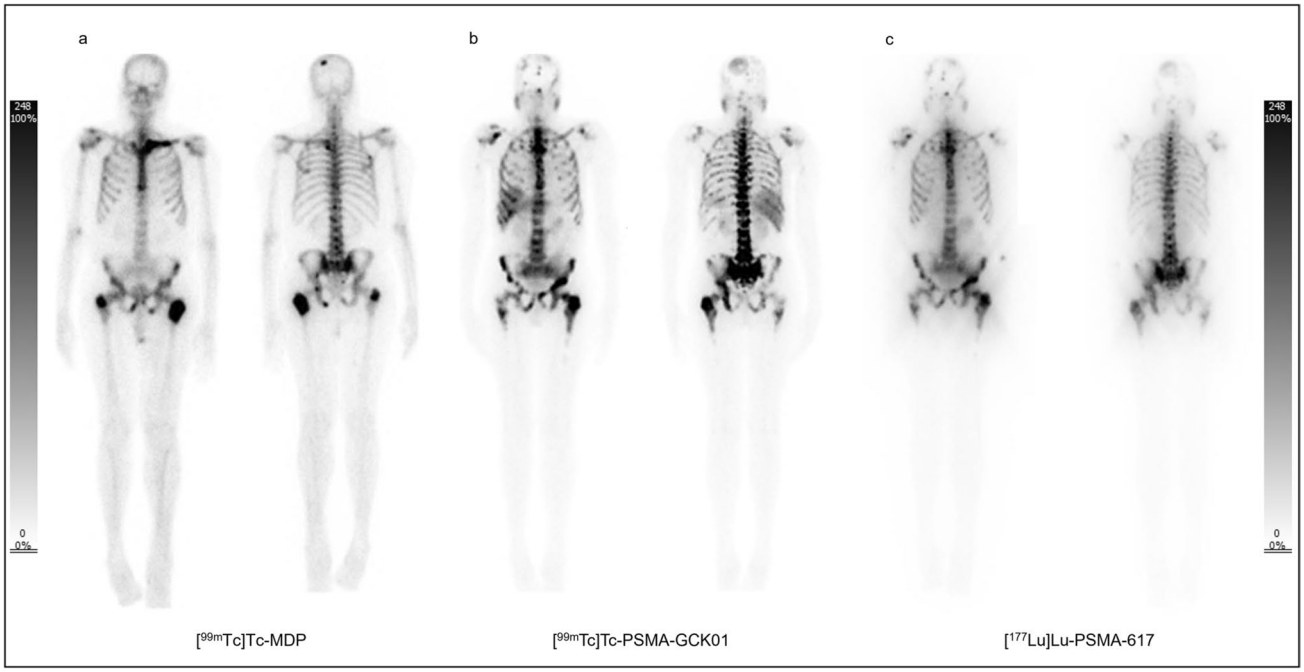


Fig. 5 Eighty-one-year-old male patient. This presents an extended prostate cancer undergoing a phenotyping for specific radio-ligand therapy with PSMA. **a** [^{99m}Tc]Tc-MDP uptake in whole body projections, **b** [^{99m}Tc]Tc-PSMA-GCK01 uptake in whole body projections, and **c** [^{177}Lu]Lu-PSMA-617 uptake in whole body projections. [^{99m}Tc]Tc-PSMA-GCK01 reveals more lesions compared with [^{99m}Tc]Tc-MDP. [^{99m}Tc]Tc-PSMA-GCK01 uptake correlates very well with therapeutic biodistribution

In the practical setting, both tracers had a relevant uptake in the kidneys and the liver. But in the case of [^{99m}Tc]Tc-PSMA-GCK01, a minor shift of the uptake to the liver parenchyma was observed. This behavior is particularly interesting because in mice only minimal liver uptake was previously described [13]. Differences in biodistribution between preclinical data and in humans have already been reported for PSMA tracers [21]. Vallabhajosula et al. found a very high liver uptake for [^{99m}Tc]Tc-MIP-1404 contrary to preclinical data [17]. [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA was optimized for rapid blood clearance; however, more activity is cleared by the kidneys to the bladder [18]. The rapid blood clearance of [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA is maintained with [^{99m}Tc]Tc-PSMA-GCK01 but the slight predominant hepatobiliary route of excretion of [^{99m}Tc]Tc-PSMA-GCK01 might allow a better diagnostic performance of local relapse or perivesical lymph node metastasis [22] since tracer retention in the bladder might complicate the proper discrimination of perivesical lesions. Further, the slight predominant hepatobiliary excretion could be particularly beneficial when the ligand is labeled with ^{188}Re in regard to radioligand therapy in terms of nephrotoxicity.

We have also compared the uptake of both tracers in the metastases. The analysis of tumor-to-background ratio presented equal results of [^{99m}Tc]Tc-PSMA-GCK01 and [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA on the planar whole body projections. Thus, the slight preference for hepatobiliary clearance of [^{99m}Tc]Tc-PSMA-GCK01 does not have a negative effect on the imaging properties of tumor lesions and rather offers an optimized biodistribution regarding radioligand therapy, as mentioned above.

The main advantage of [^{99m}Tc]Tc-PSMA-GCK01 over [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA is its utilization for theranostic purposes by ^{188}Re labeling [13, 15]. This ensures the correlation between pre-therapeutic and therapeutic distribution and is expected to have a direct impact on the clinical routine.

Although PSMA PET radioligands and Tc-labeled PSMA SPECT tracers use the same target structure, there are differences in sensitivity [19] and lesion-to-liver ratio. The latter is important in case of eligibility for [^{177}Lu]Lu-PSMA therapy and needs to be adjusted if ^{99m}Tc labeled tracers are used for the pre-therapeutic diagnostic [8, 23]. The tracer PSMA-GCK01 labeled with ^{99m}Tc-/^{188}Re-theranostic tandem could solve this problem [13] without negatively affecting the biodistribution and general imaging characteristics compared with [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA.

Conclusions

[^{99m}Tc]Tc-PSMA-GCK01 shows a similar uptake (counts/MBq) in normal organs and background, consequently leading to equivalent contrast ratios and general uptake in tumor lesions compared to [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA. [^{99m}Tc]Tc-PSMA-GCK01 presents a small shift from kidney to liver parenchyma in the clearance route, which is an appreciated characteristic regarding its therapeutic use but not a relevant factor for diagnostics. In sum, its additional appropriateness for use as a therapeutic agent does not compromise the general imaging characteristics of [^{99m}Tc]Tc-PSMA-GCK01 compared to the already well-established [^{99m}Tc]Tc-EDDA/HYNIC-iPSMA.

Acknowledgements We gratefully acknowledge HERMES Medical Solutions AB, Stockholm, Sweden. We gratefully acknowledge the computing time granted by the Jülich Aachen Research Alliance Vergabegremium and provided under project "3cldtp" on the Jülich Aachen Research Alliance Partition part of the supercomputer JURECA at Forschungszentrum Jülich.

Author Contribution EM and MK conducted the data analysis; LS, JC, and BS contributed the docking study; EW collected and constituted the patient characteristics; UH, CK, and FLG were in charge of patient application at University Hospital Heidelberg; AOA and MS conducted application of GCK01/Pluvicto at Korle Bu Teaching Hospital. The first draft of the manuscript was written by EM and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. This work was supported by Telix Pharmaceuticals Limited.

Data Availability The data used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethical Approval All procedures were carried out in accordance with the ethical standards of the institutional and/or national research committees and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The ethical Committee of the University Hospital Heidelberg approved the retrospective Evaluation (permission S-732-18).

Conflict of Interest FLG, UH, and JC have patent applications for PSMA-1007. FLG is also advisor at ABX Advanced Biochemical Compound, Telix Pharmaceuticals and SOFIE Biosciences. FLG, CK, JC, and UH are inventors of PSMA-GCK01 which is in the global patent applications process and owned by Telix Pharmaceuticals Limited. The other authors declare no conflict of interest regarding this manuscript.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Pienta KJ, Gorin MA, Rowe SP et al (2021) A phase 2/3 prospective multicenter study of the diagnostic accuracy of prostate specific membrane antigen PET/CT with 18F-DCFPyL in prostate cancer patients (OSPREY). J Urol 206:52–61. https://doi.org/10.1097/JU.0000000000001698
2. Morris MJ, Rowe SP, Gorin MA et al (2021) Diagnostic performance of 18F-DCFPyL-PET/CT in men with biochemically recurrent prostate cancer: results from the CONDOR phase III, multicenter study. Clin Cancer Res Off J Am Assoc Cancer Res 27:3674–3682. https://doi.org/10.1158/1078-0432.CCR-20-4573
3. Olivier P, Giraudet A-L, Skanjeti A et al (2023) Phase III study of 18F-PSMA-1007 versus 18F-fluorocholine PET/CT for localization of prostate cancer biochemical recurrence: a prospective, randomized, crossover multicenter study. J Nucl Med Off Publ Soc Nucl Med 64:579–585. https://doi.org/10.2967/jnumed.122.264743
4. Li B, Duan L, Shi J et al (2022) Diagnostic performance of 99mTc-HYNIC-PSMA SPECT/CT for biochemically recurrent prostate cancer after radical prostatectomy. Front Oncol 12:1072437. https://doi.org/10.3389/fonc.2022.1072437
5. Duncan I, Ingold N, Martinez-Marroquin E, Paterson C (2023) An Australian experience using Tc-PSMA SPECT/CT in the primary diagnosis of prostate cancer and for staging at biochemical recurrence after local therapy. Prostate. https://doi.org/10.1002/pros.24538
6. Schmidkonz C, Hollweg C, Beck M et al (2018) 99m Tc-MIP-1404-SPECT/CT for the detection of PSMA-positive lesions in 225 patients with biochemical recurrence of prostate cancer. Prostate 78:54–63. https://doi.org/10.1002/pros.23444
7. Schmidkonz C, Cordes M, Beck M et al (2018) Assessment of treatment response by 99mTc-MIP-1404 SPECT/CT: a pilot study in patients with metastatic prostate cancer. Clin Nucl Med 43:e250–e258. https://doi.org/10.1097/RLU.0000000000002162
8. Cook GJR, Wong W-L, Sanghera B et al (2023) Eligibility for 177Lu-PSMA therapy depends on the choice of companion diagnostic tracer: a comparison of 68Ga-PSMA-11 and 99mTc-MIP-1404 in metastatic castration-resistant prostate cancer. J Nucl Med Off Publ Soc Nucl Med 64:227–231. https://doi.org/10.2967/jnumed.122.264296
9. Strosberg J, El-Haddad G, Wolin E et al (2017) Phase 3 trial of 177Lu-dotatate for midgut neuroendocrine tumors. N Engl J Med 376:125–135. https://doi.org/10.1056/NEJMoa1607427
10. Goffin KE, Joniau S, Tenke P et al (2017) Phase 2 study of 99mTc-trofolastat SPECT/CT to identify and localize prostate cancer in intermediate- and high-risk patients undergoing radical prostatectomy and extended pelvic LN dissection. J Nucl Med Off Publ Soc Nucl Med 58:1408–1413. https://doi.org/10.2967/jnumed.116.187807
11. García-Pérez FO, Davanzo J, López-Buenrostro S et al (2018) Head to head comparison performance of 99mTc-EDDA/HYNIC-iPSMA SPECT/CT and 68Ga-PSMA-11 PET/CT a prospective study in biochemical recurrence prostate cancer patients. Am J Nucl Med Mol Imaging 8:332–340
12. Robu S, Schottelius M, Eiber M et al (2017) Preclinical evaluation and first patient application of 99mTc-PSMA-I&S for SPECT imaging and radioguided surgery in prostate cancer. J Nucl Med Off Publ Soc Nucl Med 58:235–242. https://doi.org/10.2967/jnumed.116.178939
13. Cardinale J, Giesel FL, Wensky C et al (2023) PSMA-GCK01: a generator-based 99mTc-/188Re theranostic ligand for the prostate-specific membrane antigen. J Nucl Med Off 64(7):1069–1075. https://doi.org/10.2967/jnumed.122.264944
14. Wang L, Lu B, He M et al (2022) Prostate cancer incidence and mortality: global status and temporal trends in 89 countries from 2000 to 2019. Front Public Health 10:811044. https://doi.org/10.3389/fpubh.2022.811044
15. Brunello S, Salvarese N, Carpanese D et al (2022) A review on the current state and future perspectives of [99mTc]Tc-housed PSMA-i in prostate cancer. Molecules 27:2617. https://doi.org/10.3390/molecules27092617
16. Hillier SM, Maresca KP, Lu G et al (2013) 99mTc-labeled small-molecule inhibitors of prostate-specific membrane antigen for molecular imaging of prostate cancer. J Nucl Med Off Publ Soc Nucl Med 54:1369–1376. https://doi.org/10.2967/jnumed.112.116624
17. Vallabhajosula S, Nikolopoulou A, Babich JW et al (2014) 99mTc-labeled small-molecule inhibitors of prostate-specific membrane antigen: pharmacokinetics and biodistribution studies in healthy subjects and patients with metastatic prostate cancer. J Nucl Med Off Publ Soc Nucl Med 55:1791–1798. https://doi.org/10.2967/jnumed.114.140426
18. Ferro-Flores G, Luna-Gutiérrez M, Ocampo-García B et al (2017) Clinical translation of a PSMA inhibitor for 99mTc-based SPECT. Nucl Med Biol 48:36–44. https://doi.org/10.1016/j.nucmedbio.2017.01.012
19. Lawal IO, Ankrah AO, Mokgoro NP et al (2017) Diagnostic sensitivity of Tc-99m HYNIC PSMA SPECT/CT in prostate carcinoma: a comparative analysis with Ga-68 PSMA PET/CT. Prostate 77:1205–1212. https://doi.org/10.1002/pros.23379
20. Kratochwil C, Giesel FL, Stefanova M et al (2016) PSMA-targeted radionuclide therapy of metastatic castration-resistant prostate cancer with 177Lu-labeled PSMA-617. J Nucl Med Off Publ Soc Nucl Med 57:1170–1176. https://doi.org/10.2967/jnumed.115.171397
21. Kabasakal L, AbuQbeitah M, Aygün A et al (2015) Pre-therapeutic dosimetry of normal organs and tissues of (177)Lu-PSMA-617 prostate-specific membrane antigen (PSMA) inhibitor in patients with castration-resistant prostate cancer. Eur J Nucl Med Mol Imaging 42:1976–1983. https://doi.org/10.1007/s00259-015-3125-3
22. Giesel FL, Will L, Lawal I et al (2018) Intraindividual comparison of 18F-PSMA-1007 and 18F-DCFPyL PET/CT in the prospective evaluation of patients with newly diagnosed prostate carcinoma: a pilot study. J Nucl Med Off Publ Soc Nucl Med 59:1076–1080. https://doi.org/10.2967/jnumed.117.204669
23. Sartor O, de Bono J, Chi KN et al (2021) Lutetium-177–PSMA-617 for metastatic castration-resistant prostate cancer. N Engl J Med 385:1091–1103. https://doi.org/10.1056/NEJMoa2107322

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Eduards Mamlins[1] · Lara Scharbert[2,3] · Jens Cardinale[1,4] · Maria Krotov[1] · Erik Winter[4] · Hendrik Rathke[4,5] · Birgit Strodel[2,3] · Alfred O. Ankrah[6] · Mike Sathekge[7] · Uwe Haberkorn[4,8,9] · Clemens Kratochwil[4] · Frederik L. Giesel[1,4]

1 Department of Nuclear Medicine, Medical Faculty, University Hospital Dusseldorf, Heinrich-Heine-University Dusseldorf, Moorenstrasse 5, 40225 Dusseldorf, Germany

2 Institute of Biological Information Processing: Structural Biochemistry (IBI-7), Forschungszentrum Juelich GmbH, Juelich, Germany

3 Institute of Theoretical and Computational Chemistry, Heinrich-Heine-University Dusseldorf, Dusseldorf, Germany

4 Department of Nuclear Medicine, University Hospital Heidelberg, Heidelberg, Germany

5 Department of Nuclear Medicine, University Hospital of Bern - Inselspital, Bern, Switzerland

6 Korle Bu Teaching Hospital, Accra, Ghana

7 Department of Nuclear Medicine, University of Pretoria and Steve Biko Academic Hospital, Pretoria, South Africa

8 Clinical Cooperation Unit Nuclear Medicine, DKFZ, Heidelberg, Germany

9 Translational Lung Research Center Heidelberg (TLRC), German Center for Lung Research (DZL), Heidelberg, Germany

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Innovative strategies for modeling peptide–protein interactions and rational peptide drug design

Lara Scharbert[1,2] and Birgit Strodel[1,2]

This review highlights cutting-edge techniques for modeling peptide–protein interactions and advancing computer-aided peptide–drug design. We examine significant progress in generating peptide poses through docking and artificial intelligence (AI), assessing peptide flexibility via enhanced molecular dynamics simulations, and analyzing binding interactions through free energy calculations. Additionally, we discuss how these insights can inform the rational design of therapeutic peptides by utilizing free energy metrics and strategic modifications to enhance their binding affinity and therapeutic potential. Looking forward, further integrating AI will be crucial for optimizing peptide design and enhancing drug development efforts.

Addresses
[1] Institute of Theoretical and Computational Chemistry, Heinrich-Heine-University Düsseldorf, Universitaetsstr. 1, Düsseldorf, 40225, Germany
[2] Institute of Biological Information Processing, Structural Biochemistry (IBI-7), Forschungszentrum Jülich, Wilhelm-Johnen-Straße, Jülich, 52428, Germany

Corresponding author: Strodel, Birgit (birgit.strodel@uni-duesseldorf.de)

Current Opinion in Structural Biology 2025, **93**:103083

This review comes from a themed issue on **Biophysical Methods (2025)**

Edited by **Hoi Sung Chung** and **Francesca Marassi**

For complete overview of the section, please refer the article collection - Biophysical Methods (2025)

Available online 25 June 2025

https://doi.org/10.1016/j.sbi.2025.103083

0959-440X/© 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Introduction

Peptides are estimated to be involved in 15–40 % of all protein-related interactions [1]. These peptide–protein interactions (pepPIs) play crucial roles in processes such as signal transduction, immune regulation, and proteolytic processing, making them attractive targets for therapeutic intervention. Currently, more than 80 approved peptide drugs are available on the global market. Although therapeutic peptides account for only 5 % of the pharmaceutical industry, the market is expected to grow with a projected compound annual growth rate of 10.8 % from 2025 to 2030 [2,3]. Small molecules still dominate the global pharmaceutical market with 75 % of the shares as they offer several advantages, including low production costs, oral bioavailability, and good membrane permeability. However, their small size limits their ability to inhibit interactions involving large surface areas, such as protein–protein interactions (PPIs) [4]. Due to their larger size and inherent flexibility, peptide drugs, in contrast, have a greater potential to target and inhibit PPIs that were previously considered 'undruggable' [5].

In the early days of peptide–drug discovery, scientists were restricted to natural sources. For instance, the first peptide–drug, insulin, was derived from animals, often leading to allergic reactions. With the advent of recombinant technology in the 1980s, clean and selective production of peptides and proteins in cell cultures became feasible. Another major milestone was the invention of a target-based search for novel peptide–drug candidates; the phage display technique enabled the screening of large peptide libraries (up to 10^{10}) against a target protein using bacteriophages [4,6].

With advances in computational power and knowledge, computer-aided drug design (CADD) has become an essential tool in drug discovery. CADD accelerates drug development while reducing costs and resources by enabling virtual screening of vast compound databases and using a knowledge-driven approach to understanding protein–ligand interactions [7]. Building on the principles established in CADD, computer-aided peptide–drug design focuses specifically on optimizing peptide interactions with their targets. A rational peptide design approach necessitates a comprehensive understanding of the underlying pepPIs; therefore, reliable structural information about the involved binding partners is essential. For targets, existing structures can often be utilized as experimental high-resolution methods such as X-ray crystallography and nuclear magnetic resonance spectroscopy have determined over 200,000 protein structures available in the Protein Data Bank (PDB) to date. However, these techniques are complex, time-consuming, costly, and may result in non-native conformations [8,9]. AlphaFold developed by Google DeepMind revolutionized structural biology by leveraging artificial intelligence (AI) to predict protein structures solely from the amino acid sequence, now achieving results in minutes using the AlphaFold server [10,11].

While protein structures for most target proteins can typically be found in databases, establishing the structures of peptides and their interactions with proteins poses greater challenges. This difficulty applies to both experimental and simulation approaches, largely due to the inherent flexibility of peptides, which results in a vast number of possible conformations. Moreover, while small-molecule binding pockets are usually deeply buried and well-defined, peptides rather bind to larger, disordered binding sites on the surface of the protein (Figure 1a vs. b). These issues can lead to uncertainty in predicting binding poses using docking techniques. To address this, molecular dynamics (MD) simulations enable the sampling of the dynamics of pepPIs, facilitating a more holistic analysis of the conformational landscape [7,12].

This review aims to highlight some of the latest and most innovative methods for modeling pepPIs and computer-aided peptide drug design. While comprehensive reviews have previously covered individual aspects of these topics [12,13], we focus on the most recent advancements in the field, including peptide pose generation, evaluation of peptide flexibility and interactions at the binding site, and the rational design of new peptide drugs. These various approaches can be combined into a workflow as shown in Figure 2.

Generation of peptide–protein binding poses

A classic structure-based drug design journey typically starts with the identification of a therapeutic target. If a high-resolution structure of the target protein is not yet available in the PDB, computational methods such as homology modeling—though this approach is becoming increasingly less relevant—and AI approaches like AlphaFold can be employed, to predict its three-dimensional conformation [9]. With the protein structure by hand, the next step is to determine the area where the ligand can bind.

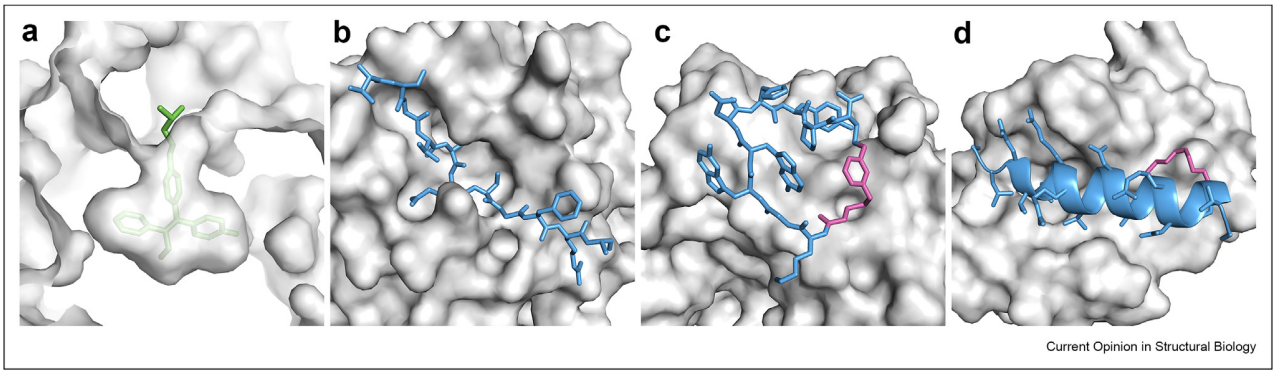
Pocket detection

If no potential binding site is known, there are several open source algorithms for locating and evaluating surface pockets and cavities. Two prominent tools are the web server CASTp 3.0 [14] and Fpocket [15], which utilize a geometry-based method that relies on spheres to detect grooves on the protein surface. In peptide drug discovery, the large, relatively flat interface of pepPIs contrasts sharply with the well-defined and often deeply buried binding pockets commonly found for small molecules. Standard algorithms could therefore be misleading in predicting binding sites for peptides. To address this problem, methods for the identification of pepPI sites were developed. The well-established web server PepSite utilizes a scoring method based on residue binding preferences derived from known peptide–protein structures to predict potential peptide binding sites on protein surfaces [16]. The recently introduced SiteFerret is a novel method for automatic pocket detection that effectively identifies both small-molecule binding pockets and peptide binding sites, utilizing an algorithm that combines hierarchical clustering of virtual probe spheres with the Isolation Forest method to rank putative pockets [17].

Global docking

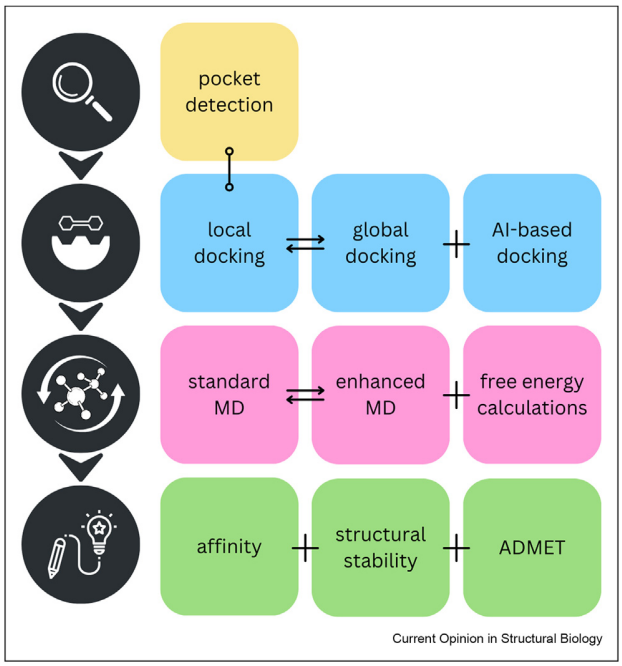
A fast-forward method for determining the structures of peptide–protein complexes—skipping the step of binding pocket detection—is to employ global docking. To this end, the user can choose from a vast range of docking programs that can be classified into protein–protein, peptide–protein, and small-molecule–protein docking. Agrawal et al. compared a diverse set of docking methods for peptide–protein docking using

Figure 1



Representative examples of different protein binding sites occupied by (a) a small-molecule inhibitor (PDB code 3ERT), (b) an extended linear peptide (PDB code 7T70), (c) a cyclic peptide (PDB code 6XIB), and (d) a stapled helical peptide (PDB code 3MK8). The proteins are shown as gray surface and the ligands as sticks, with green for the small molecule, blue for the peptides, purple for peptide modifications, and a blue cartoon for the helix. PDB, Protein Data Bank.

Figure 2



Workflow of a typical peptide–drug design study. The process begins with identifying the peptide binding site on the protein surface if local docking is applied in the next step. Global docking involves the concurrent identification of both the binding site and binding pose. In a consensus docking approach, local or global docking is combined with AI-based pepPI prediction. The peptide pose is then validated and evaluated through standard or enhanced MD simulations, which also quantify peptide flexibility at the binding site and allow for the calculation of binding free energies. Ultimately, the insights gained from pepPI modeling inform peptide design, enhancing the affinity, structural stability, and pharmacological activity of the peptide–drug candidate. AI, artificial intelligence; MD, molecular dynamics; pepPIs, peptide–protein interactions.

133 peptide–protein complexes, focusing on peptides that are 9–15 residues long [18]. They found that FRODOCK [19], a fast global protein–protein docking web server, performed best in blind docking. In the work of Weng et al. a similar, but more comprehensive approach was followed, where they compared 14 docking programs—three for small-molecule–protein docking, three for protein–protein docking, and eight for peptide–protein docking—on a set of 185 peptide–protein complexes, comprising peptides of 5–20 residues [20]. In contrast to Agrawal et al. they found that peptide–protein docking algorithms generally performed better than protein–protein or small-molecule docking algorithms. Notably, in global docking, the web server HPEPDOCK [21], which employs rapid peptide conformation sampling and ensemble docking techniques, produced the best predictions for the entire data set.

Local docking

When the binding site is known, local docking is the preferred and more efficient method. In this category, AutoDock CrankPep (ADCP) [22] has demonstrated the best overall results [20]. The high number of rotatable bonds in peptides contributes to their conformational versatility, which distinguishes them not only from the more rigid small molecules but also from folded proteins. This enormous flexibility is a big challenge for standard docking programs [12]. The peptide docking program ADCP combines the peptide backbone conformation sampling method of CRANKITE [23] with the grid-based receptor representation of AutoDock [24]. The extensive conformational space is efficiently navigated by ADCP using a Metropolis Monte Carlo search strategy, sampling peptide positions and orientations within the receptor's binding site, thereby producing docking poses [22]. Moreover, ADCP is also capable of docking cyclic peptides that are either linked through their backbone or side chain disulfide bonds [25]. Besides peptide flexibility, some docking programs—such as HADDOCK [26], CABS-Dock [27], and FlexPepDock [28]—also account for receptor flexibility by allowing movement of selected side chains and in some cases also refinement of the backbone.

Defining the appropriate size of the binding site can be a challenge in local docking, as, in particular, long peptides can vary dramatically in their full size depending on how expanded or compact they are. Weng et al. studied the impact of the binding site size on the performance of AutoDock or Vina-based Docking programs, like ADCP, by comparing docking boxes generated based on the co-crystallized ligand (ligBS) with those generated based on the peptide length (pepBS) [20]. Compared to the ligBS, the success rates for pepBS were considerably lower for most docking programs, especially in the case of ADCP, and dropped even more with increasing peptide length. Therefore, when using AutoDock or Vina-based docking programs, it is advisable to define the binding site more precisely – preferably by using the co-crystallized ligand as template if available or restricting the docking box to the relevant binding site residues.

AI-based docking

In recent years, AI-based methods for protein structure prediction have emerged, with AlphaFold [10] consistently leading this field and often serving as the benchmark for state-of-the-art comparisons. The release of AlphaFold-Multimer (AF2multi) [29], an update to AlphaFold 2, enabled the prediction of protein–protein complexes, including homomeric and heteromeric protein complexes. While AF2multi was primarily designed for protein–protein complexes, it has also been shown to be effective for peptide–protein docking problems [30,31]. Shanker and Sanner compared AF2multi with

two other deep learning models, OmegaFold and AlphaFold 2 Monomer, as well as their peptide-docking software ADCP [31]. The dataset used in this study contains 99 nonredundant peptide−protein complexes, with the vast majority of peptides having a length of 5−25 amino acids and containing different secondary structures (48 % coil, 34 % helix, and 18 % strand). AF2multi outperformed all other methods with a success rate of 53 %, whereas ADCP achieved only 23 % for the top-ranked solutions. However, when considering all generated solutions beyond the top-ranked poses, ADCP's success rate rises to 62 %, indicating that there is room for improvement in the scoring function. The authors also noted that AF2multi and ADCP performed complementarily, each succeeding on different peptide−protein complexes. AF2multi is suitable for docking long, linear peptides composed of standard amino acids, accounting for both peptide and receptor flexibility. In contrast, ADCP allows peptide flexibility while docking into rigid receptors [22], supporting peptides up to 30 residues, including cyclic structures and noncanonical amino acids. For peptides compatible with both methods, a consensus docking strategy combining ADCP and AF2multi, along with an effective selection mechanism, achieved a 60 % success rate for top-ranking poses and 66 % within the top five [31].

With the release of the AlphaFold Server powered by AlphaFold 3 (AF3), predicting peptide−protein complexes has become easily and rapidly accessible to everyone [11]. Nevertheless, in the case of peptide−protein docking, latest benchmark studies have found that AF2multi performs slightly better than AF3 [32]. The deep learning model ESMFold is a protein structure prediction tool, that, unlike AlphaFold, does not rely on external databases, template searches, or multiple sequence alignments. ESMFold predicts protein structures directly from amino acid sequences, while being up to 60 times faster than AlphaFold [33]. In recent studies, ESMFold's performance in peptide−protein docking was evaluated and compared with the latest AlphaFold versions. Despite generally lower accuracy, ESMFold occasionally outperformed AF3. Given its computational efficiency, ESMFold can be an alternative or a beneficial complement to other methods in pepPI studies [34].

Limitations of docking

A limitation of docking approaches and AI-based methods for predicting peptide−protein complexes is that they provide only static snapshots of the peptides and proteins, excluding their dynamic nature and, in particular, the flexibility of peptides and thus pepPIs, which can be addressed through MD simulations (Figure 2).

Evaluation of poses, interactions, and peptide flexibility

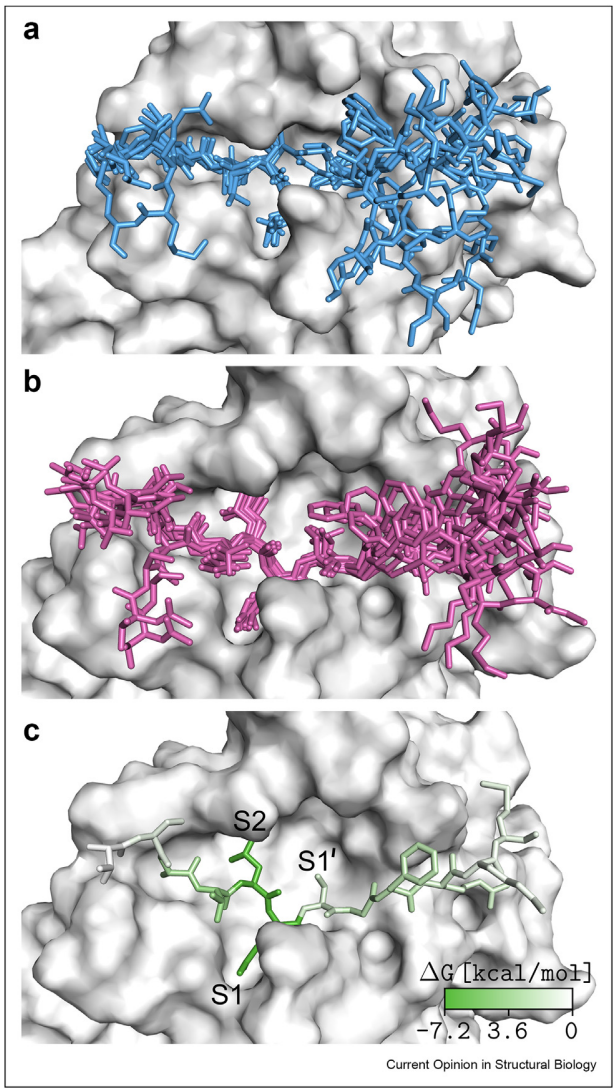
Standard MD

Molecular dynamics simulations have long been a cornerstone in understanding the dynamic nature of biological macromolecules, providing atomistic insights into conformational flexibility and interaction mechanisms that often elude static structural determination methods. Given the limitations of scoring functions in docking methods, evaluating multiple top-ranked poses (e.g. the top 3 or 5; see Figure 3a for a presentation of different docking poses) is advisable rather than relying solely on the highest-ranked pose. Furthermore, when combining physics-based and AI-based methods in a consensus approach, it is crucial to identify which predicted structures are more reliable [31]. MD simulations serve as valuable tools for assessing these docking predictions as they allow for the exploration of the stability and conformational dynamics of peptide−protein complexes over time (see Figure 3b for the peptide flexibility sampled by MD). This dynamic perspective is essential for understanding the interaction mode and affinity of pepPIs, forming the basis for subsequent drug design studies [7]. Recent examples of combining docking and MD simulations for the development of therapeutic peptides include designing peptides to target multiple over-expressed receptors in tumor cells [35], creating peptide inhibitors for the interaction between the Omicron SARS-CoV-2 spike protein and its receptor ACE2 [36], and optimizing a potent macrocyclic peptide that interferes with the protein−protein interaction between the Programmed Cell Death Protein 1 and its Ligand 1 (PD-1/PD-L1) through rational design [37].

Enhanced MD

However, traditional MD simulations can be computationally intensive and struggle to efficiently sample rare events, such as transitions between the conformational states of peptides in peptide−protein complexes, where pepPIs may raise energy barriers. To address these limitations, enhanced MD sampling techniques, such as replica exchange MD, accelerated MD, or Gaussian accelerated MD (GaMD), have been refined or developed to improve the sampling efficiency of MD simulations of PPIs or pepPIs [38,39]. With peptide Gaussian accelerated molecular dynamics (Pep-GaMD), Wang et al. introduced a GaMD technique that is specifically tailored for modeling the high flexibility of peptides [40]. Its dual-boost algorithm applies two distinct boost potentials: one specifically targeting the peptide's essential potential energy and another for the remaining potential energy of the entire system. The same group has published a PepBinding

Figure 3



The flexibility of pepPIs exemplified by docking and MD simulations of the SARS-CoV-2 main protease (PDB code 7T70) with its substrate. (a) The top 10 ranked structures after peptide docking with ADCP. (b) The 10 most frequently sampled peptide conformations during a 100 ns MD simulation. (c) Per-residue ΔG_{bind} decomposition of the peptide, from the highest (green) to the lowest (white) contribution to the affinity (measured as negative energies, see color scale in the bottom right corner). Subsites S1 and S2 are occupied by key residues, while S1′ has potential for peptide modifications. *ΔG_{bind}, binding free energy; ADCP, AutoDock CrankPep; AI, artificial intelligence; MD, molecular dynamics; PDB, Protein Data Bank; pepPIs, peptide–protein interactions.*

workflow for predicting peptide–protein structures combining peptide docking, enhanced sampling using Pep-GaMD, and structural clustering [41]. They demonstrated their workflow on seven peptide–protein models generated via docking with HPEPDOCK, initially classified as inaccurate to medium-quality based on the Critical Assessment of PRediction of Interactions criteria. Each model underwent a 200 ns Pep-GaMD simulation, successfully refining them into five medium-quality and two acceptable-quality structures.

Free energy calculations

While docking methods use fast empirical or force field-based scoring functions to estimate the affinity between protein and peptide, free energy calculations combined with MD simulations can achieve a more accurate estimation of binding affinity [42]. To this end, the most commonly used free energy calculation methods are molecular mechanics/Poisson–Boltzmann surface area (MM/PBSA) or molecular mechanics/generalized Born surface area (MM/GBSA), free energy perturbation (FEP), and thermodynamic integration (TI). While end-point methods, like MM/PBSA and MM/GBSA, are more computationally efficient, requiring only a single trajectory, FEP and TI are both pathway-based methods that are considered more accurate but involve much higher computational costs. Therefore, MM/PB(GB)SA has become a popular technique to calculate the binding free energy (ΔG_{bind}) between a protein and ligand [43]. Weng et al. suggested that the choice between MM/PBSA and MM/GBSA may depend on peptide length and other system-specific factors. Their study showed that MM/PBSA yielded higher accuracy for short peptides (5–12 residues), while MM/GBSA performed better for medium-length peptides (20–25 residues) [44]. Furthermore, both methods are sensitive to parameters such as the solute dielectric constant and the choice of force field, and are challenged by the high computational cost and limited accuracy of entropy calculations, typically failing to (correctly) capture conformational entropy [45].

Among the available tools for MM/PB(GB)SA calculations, Amber includes the widely used MMPBSA.py script within the AmberTools package, which enables free energy calculations based on MD snapshots [46]. To enhance accessibility for GROMACS users, Valdés-Tresanco developed gmx_mmpbsa, an adaptation of MMPBSA.py for processing trajectories and topology files generated with GROMACS [47]. Compared with other MM/PB(GB)SA implementations for GROMACS, such as g_mmpbsa and GMXPBSA2.1, gmx_mmpbsa offers several advantages [48]. It is compatible with all GROMACS versions and provides enhanced calculation and analysis features, including ΔG_{bind} calculations with different solvation models, stability assessments, computational alanine scanning, ΔG_{bind} decomposition, and entropy corrections. Furthermore, it supports the Amber, OPLS, and CHARMM force fields and is easy to install using Conda or AmberTools. These advantages have contributed to the increasing use of gmx_mmpbsa for analyzing pepPI sampled by MD simulation in peptide–drug discovery studies [49–52].

Peptide–drug design

Following the identification of promising binding peptides, further refinements are needed to optimize their affinity, stability, and absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties for effective and safe therapeutic development (Figure 2) [53].

Optimizing affinity

Rational design of a peptide drug requires reliable and efficient computational tools for analyzing the affinity of pepPIs. Computational alanine scanning (CAS) is a ΔG_{bind}-based method that systematically replaces amino acids with alanine to identify key residues critical for the affinity and stability of pepPIs. CAS is not only a great method to identify these 'hot spot' residues in pepPIs, it has also helped to elucidate the structure–activity relationship of different peptide drug candidates, by giving insights into which residues are essential for activity and which can be modified to enhance affinity, specificity, or stability [54–56]. Another frequently used method in computational peptide design is the per-residue ΔG_{bind} decomposition, which analyses the energetic contributions of individual amino acids to the overall ΔG_{bind} of a peptide–protein complex (Figure 3c). Weakly contributing residues can be modified, with the effects evaluated through subsequent ΔG_{bind} calculations after mutation. Using this method, researchers were able to propose peptide residue substitutions that led to improved binding affinity [57,58].

There are several tools for ΔG_{bind}-based analyses of pepPIs at the residues level. The BAlaS web application offers a fast and user-friendly platform for performing CAS, requiring only a structure file in PDB format as input [59]. For the analysis of MD trajectories using MM/PB(GB)SA calculations, tools such as CompASM [60], an Amber-based plug-in for Visual Molecular Dynamics (VMD) [61] for conducting CAS, and gmx_MMPBSA [48], which enables both per-residue ΔG_{bind} decomposition and CAS, have been developed. While these methods can identify residues with modification potential, they do not directly suggest specific mutations. In a peptide design approach based on modeling the SARS-CoV-2 main protease with its substrates, Chan et al. [62] utilized the computational saturation mutagenesis algorithm BUDE_SM [63] from the Bristol University Docking Engine (BUDE) [64] to substitute each peptide residue with the corresponding 19 alternative proteinogenic amino acids. *In vitro* inhibition assays confirmed that they successfully translated the BUDE_SM predictions into potent peptide inhibitors.

Structural stability

To enhance a peptide's binding affinity and selectivity, one approach is to maintain its target-bound conformation, which can be achieved through strategies like cyclization. It also improves its stability and pharmacokinetic properties, as cyclization protects peptides from enzymatic degradation and hydrolysis, and reduces its immunogenicity. Peptides can be linked to themselves in different ways, such as head to tail (Figure 1c), side chain to side chain, or head to side chain, using nonpeptidic linkers or groups to facilitate these connections [65,66]. Cyclization is often applied to peptides derived from so-called 'hot segments', an extension of the 'hot spots' concept that defines a continuous binding region essential for a PPI [5]. This was demonstrated by Lopez et al. who utilized MD simulations with enhanced sampling methods to assess the degree of preorganization—the structural conformation that most accurately mimics the 'hot loop'—in a series of cyclic peptides differing in a few point mutations and linker types [67]. They ultimately obtained an inhibitor with the highest degree of preorganization and a 280-fold improvement in binding affinity. In a recent *de novo* peptide design study, the computational generation of novel macrocyclic peptides without relying on preexisting sequences was realized using AI [68]. To this end, Rettie et al. developed RFpeptides, a deep learning framework using denoising diffusion models to design macrocyclic peptides with high affinity for selected protein targets.

Another popular method to stabilize secondary structure elements like helices is by using staples, which involve cross-linking two or more side chains to provide additional structural support (Figure 1d). The most common form of staples in peptides is the introduction of disulfide bonds. However, synthetic staples can also include various linkers or chemical groups that can react with amino acid side chains to stabilize the desired conformation. Like cyclic peptides, stapled peptides often possess further advantages like enhanced cell penetration and proteolytic stability [69,70].

Absorption, distribution, metabolism, excretion, and toxicity

In addition to high affinity and selectivity, the development of effective and safe drugs requires careful consideration of their ADMET properties [53]. Especially in the context of peptide-based therapeutics, proteolytic cleavage, rapid clearance, and low membrane permeability represent a major challenge [6,71]. To overcome these limitations, strategies like chemical modifications (e.g. acetylation/amidation and nonnatural or d-amino acids), conjugation strategies (e.g.

lipidation, PEGylation, and fusion proteins), and structural stabilization (e.g. stapling or cross-linking) have been developed [72]. Two web servers for ADMET prediction have recently been introduced enabling the assessment of key drug-like properties, such as solubility, bioavailability, blood–brain barrier penetration, and toxicity. ADMET-AI allows rapid screening of up to 1000 molecules (and one million molecules with a local version) at once, making it ideal for high-throughput evaluation [73]. For a more comprehensive approach, ADMETlab 3.0 offers a wider range of ADMET properties, detailed toxicity risk assessments, and an uncertainty estimate module to aid in the candidate selection process [74]. However, it should be noted that most ADMET prediction tools are trained on small molecules and therefore face challenges to accurately predict the pharmacokinetics of therapeutic peptides. Developing a dedicated platform to address these specific ADMET challenges for peptide drugs would be highly beneficial.

Conclusion and outlook

This review emphasizes some of the latest and most innovative methods for modeling peptide–protein interactions (pepPIs) and computer-aided peptide drug design. We addressed recent advancements for the peptide pose generation using docking and AI techniques, the evaluation of peptide flexibility using MD simulations, interactions at the binding site using binding free energy decompositions, and the rational design of new peptide drugs. We recommend employing a consensus approach that integrates physics-based and AI-based docking methods to increase the likelihood of obtaining accurate structural predictions. Enhanced sampling MD simulations can further improve structural reliability, while techniques like MM/PBSA for binding free energy decomposition and computational alanine scanning facilitate the peptide design process. Despite these advancements, a significant challenge remains: the development of a comprehensive workflow as illustrated in Figure 2 that consolidates the best available tools in one platform would greatly benefit the field. Looking ahead, the integration of AI for producing structures of peptide–protein complexes as well as for assessing the ADMET properties of peptides will play an increasingly vital role in peptide design, further enhancing our capabilities in drug development. The ultimate goal is to accurately predict pepPIs so that *in silico* results match the reliability of protein structure predictions with AlphaFold and yield binding energies that correlate with experimentally determined K_D values from binding assays. Achieving this would enable peptide–protein docking to replace experiments, which are challenged by the often transient pepPIs that produce noisy data.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank the VolkswagenStiftung for their financial support of this work (grant no. 99761).

Data availability

No data were used for the research described in the article.

References

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Petsalaki E, Russell RB: **Peptide-mediated interactions in biological systems: new discoveries and applications**. *Curr Opin Biotechnol* 2008, **19**:344–350, https://doi.org/10.1016/j.copbio.2008.06.004.

2. Rossino G, Marchese E, Galli G, Verde F, Finizio M, Serra M, Linciano P, Collina S: **Peptides as therapeutic agents: challenges and opportunities in the green transition era**. *Molecules (Basel)* 2023, **28**:7165, https://doi.org/10.3390/molecules28207165.

3. *Global peptide therapeutics market size & outlook*. 2025 [cited Monday, 10. March 2025]. URL, https://www.grandviewresearch.com/horizon/outlook/peptide-therapeutics-market-size/global.

4. Wang L, Wang N, Zhang W, Cheng X, Yan Z, Shao G, Wang X, Wang R, Fu C: **Therapeutic peptides: current applications and future directions**. *Signal Transduct Targeted Ther* Feb. 2022, **7**, https://doi.org/10.1038/s41392-022-00904-4.

5. London N, Raveh B, Schueler-Furman O: **Druggable protein–protein interactions – from hot spots to hot segments**. *Curr Opin Chem Biol* 2013, **17**:952–959, https://doi.org/10.1016/j.cbpa.2013.10.011.

6. Muttenthaler M, King GF, Adams DJ, Alewood PF: **Trends in peptide drug discovery**. *Nat Rev Drug Discov* 2021, **20**:309–325, https://doi.org/10.1038/s41573-020-00135-8.

7. Sabe VT, Ntombela T, Jhamba LA, Maguire GE, Govender T, Naicker T, Kruger HG: **Current trends in computer aided drug design and a highlight of drugs discovered via computational techniques: a review**. *Eur J Med Chem* 2021, **224**, 113705, https://doi.org/10.1016/j.ejmech.2021.113705.

8. RCSB: ***RCSB protein Data Bank*** [cited Tuesday, 11. March 2025]. URL, https://www.rcsb.org.

9. Bertoline LMF, Lima AN, Krieger JE, Teixeira SK: **Before and after alphafold2: an overview of protein structure prediction**. *Front Bioinform* 2023, **3**, 1120370, https://doi.org/10.3389/fbinf.2023.1120370.

10. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, *et al.*: **Highly accurate protein structure prediction with alphafold**. *Nature* 2021, **596**:583–589.

11. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, Bodenstein SW, Evans DA, Hung C-C, O'Neill M, Reiman D, Tunyasuvunakool K, Wu Z, Žemgulytė A, Arvaniti E, Beattie C, Bertolli O, Bridgland A, Cherepanov A, Congreve M, Cowen-Rivers AI, Cowie A, Figurnov M, Fuchs FB, Gladman H, Jain R, Khan YA, Low CMR, Perlin K, Potapenko A, Savy P, Singh S, Stecula A, Thillaisundaram A, Tong C, Yakneen S, Zhong ED, Zielinski M, Žídek A, Bapst V, Kohli P, Jaderberg M, Hassabis D, Jumper JM: **Accurate structure prediction of biomolecular interactions with alphafold 3**. *Nature* 2024, **630**:493–500, https://doi.org/10.1038/s41586-024-07487-w.

12. Mondal A, Chang L, Perez A: **Modelling peptide–protein complexes: docking, simulations and machine learning**. *QRB Discov* 2022, **3**:e17, https://doi.org/10.1017/qrd.2022.14.

13. Wang X, Ni D, Liu Y, Lu S: **Rational design of peptide-based inhibitors disrupting protein-protein interactions**. *Front Chem* May 2021, **9**, https://doi.org/10.3389/fchem.2021.682675.

14. Tian W, Chen C, Lei X, Zhao J, Liang J: **Castp 3.0: computed atlas of surface topography of proteins**. *Nucleic Acids Res* 2018, **46**:W363–W367, https://doi.org/10.1093/nar/gky473.

15. Le Guilloux V, Schmidtke P, Tuffery P: **Fpocket: an open source platform for ligand pocket detection**. *BMC Bioinf* 2009, **10**:168, https://doi.org/10.1186/1471-2105-10-168.

16. Trabuco LG, Lise S, Petsalaki E, Russell RB: **Pepsite: prediction of peptide-binding sites from protein surfaces**. *Nucleic Acids Res* 2012, **40**:W423–W427, https://doi.org/10.1093/nar/gks398.

17. •• Gagliardi L, Rocchia W: **Siteferret: beyond simple pocket identification in proteins**. *J Chem Theor Comput* 2023, **19**: 5242–5259, https://doi.org/10.1021/acs.jctc.2c01306.

SiteFerret is a powerful and versatile tool for identifying all types of binding pockets on protein surfaces, including those for peptides. Using AI techniques, it accurately predicts and ranks binding sites, offering a reliable method for subsequent local docking.

18. Agrawal P, Singh H, Srivastava HK, Singh S, Kishore G, Raghava GPS: **Benchmarking of different molecular docking methods for protein-peptide docking**. *BMC Bioinf* Feb. 2019, **19**, https://doi.org/10.1186/s12859-018-2449-y.

19. Ramírez-Aportela E, López-Blanco JR, Chacón P: **Frodock 2.0: fast protein–protein docking server**. *Bioinformatics* 2016, **32**: 2386–2388, https://doi.org/10.1093/bioinformatics/btw141.

20. Weng G, Gao J, Wang Z, Wang E, Hu X, Yao X, Cao D, Hou T: **Comprehensive evaluation of fourteen docking programs on protein–peptide complexes**. *J Chem Theor Comput* 2020, **16**: 3959–3969, https://doi.org/10.1021/acs.jctc.9b01208.

21. Zhou P, Jin B, Li H, Huang S-Y: **HPEPDOCK: a web server for blind peptide–protein docking based on a hierarchical algorithm**. *Nucleic Acids Res* 2018, **46**:W443–W450, https://doi.org/10.1093/nar/gky357.

22. Zhang Y, Sanner MF: **AutoDock CrankPep: combining folding and docking to predict protein-peptide complexes**. *Bioinformatics* 2019, **35**:5121–5127, https://doi.org/10.1093/bioinformatics/btz459.

23. Podtelezhnikov AA, Wild DL: **CRANKITE: a fast polypeptide backbone conformation sampler**. *Source Code Biol Med* 2008, **3**:12, https://doi.org/10.1186/1751-0473-3-12.

24. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ: **AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility**. *J Comput Chem* 2009, **30**:2785–2791, https://doi.org/10.1002/jcc.21256.

25. Zhang Y, Sanner MF: **Docking flexible cyclic peptides with AutoDock CrankPep**. *J Chem Theor Comput* 2019, **15**: 5161–5168, https://doi.org/10.1021/acs.jctc.9b00557.

26. Trellet M, Melquiond AS, Bonvin AM: **A unified conformational selection and induced fit approach to protein-peptide docking**. *PLoS One* 2013, **8**, e58769.

27. Kurcinski M, Jamroz M, Blaszczyk M, Kolinski A, Kmiecik S: **Cabs-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site**. *Nucleic Acids Res* 2015, **43**:W419–W424.

28. London N, Raveh B, Cohen E, Fathi G, Schueler-Furman O: **Rosetta FlexPepDock web server—high resolution modeling of peptide–protein interactions**. *Nucleic Acids Res* 2011, **39**: W249–W253, https://doi.org/10.1093/nar/gkr431.

29. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, Žídek A, Bates R, Blackwell S, Yim J, Ronneberger O, Bodenstein S, Zielinski M, Bridgland A, Potapenko A, Cowie A, Tunyasuvunakool K, Jain R, Clancy E, Kohli P, Jumper J, Hassabis D: **Protein complex prediction with alphafold-multimer**. *bioRxiv* 2022, https://doi.org/10.1101/2021.10.04.463034.

30. Johansson-Åkhe I, Wallner B: **Improving peptide-protein docking with AlphaFold-Multimer using forced sampling**. *Front Bioinform* 2022, **2**, 959160, https://doi.org/10.3389/fbinf.2022.959160.

31. •• Shanker S, Sanner MF: **Predicting protein–peptide interactions: benchmarking deep learning techniques and a comparison with focused docking**. *J Chem Inf Model* 2023, **63**: 3158–3170, https://doi.org/10.1021/acs.jcim.3c00602.

This study evaluates the efficacy of deep learning models, including AlphaFold2, in predicting pepPIs and compares their performance to the docking method AutoDock CrankPep. AlphaFold2 achieves a docking success rate of 53%, while AutoDock CrankPep samples correct solutions for about 62% of complexes when considering all generated solutions. Notably, a consensus approach combining both methods yields a 60% success rate for the top-ranking results, providing valuable insights for advancing computational strategies in modeling pepPIs.

32. Manshour N, Ren J, Esmaili F, Bergstrom E, Xu D: **Comprehensive evaluation of alphafold-multimer, alphafold3 and colabfold, and scoring functions in predicting protein-peptide complex structures**. *bioRxiv* 2024, https://doi.org/10.1101/2024.11.11.622992.

33. Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, Smetanin N, Verkuil R, Kabeli O, Shmueli Y, dos Santos Costa A, Fazel-Zarandi M, Sercu T, Candido S, Rives A: **Evolutionary-scale prediction of atomic-level protein structure with a language model**. *Sci Technol Humanit* 2023, **379**:1123–1130, https://doi.org/10.1126/science.ade2574.

34. • Zalewski M, Wallner B, Kmiecik S: **Protein–peptide docking with ESMFold language model**. *J Chem Theor Comput* Mar. 2025, https://doi.org/10.1021/acs.jctc.4c01585.

In this study, the ESMFold language model for peptide–protein docking is compared to state of the art docking programs like AlphaFold 3. While ESMFold's accuracy is generally lower than that of AlphaFold, it offers excellent computational efficiency. This makes ESMFold a valuable tool for high-throughput pose prediction studies and a useful complement to other methods in a consensus approach.

35. Rico MI, Lebedenko CG, Mitchell SM, Banerjee IA: **Molecular dynamics simulations, docking and mmgbsa studies of newly designed peptide-conjugated glucosyloxy stilbene derivatives with tumor cell receptors**. *Mol Divers* 2022, **26**: 2717–2743, https://doi.org/10.1007/s11030-021-10354-9.

36. Jin J, Wang N, Wang S, Alsaikhan F, Ahmad I: **Targeting the main sars-cov-2 pathways with peptide inhibitors by molecular docking and molecular simulation approaches**. *J Mol Liq* 2024, **398**, 124193, https://doi.org/10.1016/j.molliq.2024.124193.

37. Miao Q, Zhang W, Zhang K, Li H, Zhu J, Jiang S: **Rational design of a potent macrocyclic peptide inhibitor targeting the pd-1/pd-l1 protein–protein interaction**. *RSC Adv* 2021, **11**: 23270–23279, https://doi.org/10.1039/d1ra03118j.

38. Siebenmorgen T, Zacharias M: **Efficient refinement and free energy scoring of predicted protein–protein complexes using replica exchange with repulsive scaling**. *J Chem Inf Model* 2020, **60**:5552–5562, https://doi.org/10.1021/acs.jcim.0c00853.

39. Wang J, Alekseenko A, Kozakov D, Miao Y: **Improved modeling of peptide-protein binding through global docking and accelerated molecular dynamics simulations**. *Front Mol Biosci* 2019, **6**, https://doi.org/10.3389/fmolb.2019.00112.

40. Wang J, Miao Y: **Peptide Gaussian accelerated molecular dynamics (Pep-GaMD): enhanced sampling and free energy and kinetics calculations of peptide binding**. *J Chem Phys* 2020, **153**, 154109, https://doi.org/10.1063/5.0021399.

41. •• Wang J, Koirala K, Do HN, Miao Y: **PepBinding: a workflow for predicting peptide binding structures by combining peptide docking and peptide Gaussian accelerated molecular dynamics simulations**. *J Phys Chem B* 2024, **128**:7332–7340, https://doi.org/10.1021/acs.jpcb.4c02047.

PepBinding is a workflow that combines peptide docking and Gaussian accelerated molecular dynamics (Pep-GaMD) simulations to predict peptide binding structures. The authors showed that while peptide docking alone provided medium-to-poor quality models, the addition of Pep-GaMD significantly improved the results, offering a more accurate prediction of pepPIs. This approach serves as a valuable component of an efficient peptide drug discovery workflow.

42. Malmstrom RD, Watowich SJ: **Using free energy of binding calculations to improve the accuracy of virtual screening predictions**. *J Chem Inf Model* 2011, **51**:1648–1655, https://doi.org/10.1021/ci200126v.

43. Wang S, Sun X, Cui W, Yuan S: **MM/PB(GB)SA benchmarks on soluble proteins and membrane proteins**. *Front Pharmacol* 2022, **13**, 1018351, https://doi.org/10.3389/fphar.2022.1018351.

44. Weng G, Wang E, Chen F, Sun H, Wang Z, Hou T: **Assessing the performance of mm/pbsa and mm/gbsa methods. 9. prediction reliability of binding affinities and binding poses for protein–peptide complexes**. *Phys Chem Chem Phys* 2019, **21**: 10135–10145, https://doi.org/10.1039/c9cp01674k.

45. Wang S, Sun X, Cui W, Yuan S: **MM/PB(GB)SA benchmarks on soluble proteins and membrane proteins**. *Front Pharmacol* 2022, **13**, https://doi.org/10.3389/fphar.2022.1018351.

46. Miller BR, McGee TD, Swails JM, Homeyer N, Gohlke H, Roitberg AE: **MMPBSA.py: an efficient program for end-state free energy calculations**. *J Chem Theor Comput* 2012, **8**: 3314–3321, https://doi.org/10.1021/ct300418h.

47. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E: **Gromacs: high performance molecular simulations through multi-level parallelism from laptops to supercomputers**. *SoftwareX* 2015, **1**:19–25, https://doi.org/10.1016/j.softx.2015.06.001.

48. Valdés-Tresanco MS, Valdés-Tresanco ME, Valiente PA, Moreno E: **gmx_MMPBSA: a new tool to perform end-state free energy calculations with GROMACS**. *J Chem Theor Comput* 2021, **17**:6281–6291, https://doi.org/10.1021/acs.jctc.1c00645.

49. Chen S, Lin T, Basu R, Ritchey J, Wang S, Luo Y, Li X, Pei D, Kara LB, Cheng X: **Design of target specific peptide inhibitors using generative deep learning and molecular dynamics simulations**. *Nat Commun* 2024, **15**:1611, https://doi.org/10.1038/s41467-024-45766-2.

50. • Wu Y, Deng S, Wang X, Thunders M, Qiu J, *et al.*: **Discovery and mechanism of action of a novel antimicrobial peptide from an earthworm**. *Microbiol Spectr* 2023, **11**:e03206–e03222, https://doi.org/10.1128/spectrum.03206-22.

In this comprehensive approach, extensive wet-laboratory experimental techniques and computational modeling methods are combined to identify and characterize a novel antimicrobial peptide (AMP) from earthworms. This combination enhances the understanding of AMP mechanisms and supports the design of new peptide-based antibiotics.

51. Niu P, Tao Y, Meng Q, Huang Y, Li S, Ding K, Ma D, Ye Z, Fan M: **Discovery of novel macrocyclic derivatives as potent and selective cyclin-dependent kinase 2 inhibitors**. *Bioorg Med Chem* 2024, **104**, 117711, https://doi.org/10.1016/j.bmc.2024.117711.

52. Eberle RJ, Sevenich M, Gering I, Scharbert L, Strodel B, Lakomek NA, Santur K, Mohrlüder J, Coronado MA, Willbold D: **Discovery of all-D-peptide inhibitors of SARS-CoV-2 3C-like protease**. *ACS Chem Biol* 2023, **18**:315–330, https://doi.org/10.1021/acschembio.2c00735.

53. Ferreira LL, Andricopulo AD: **ADMET modeling approaches in drug discovery**. *Drug Discov Today* 2019, **24**:1157–1165, https://doi.org/10.1016/j.drudis.2019.03.015.

54. Valiente PA, Wen H, Nim S, Lee J, Kim HJ, Kim J, Perez-Riba A, Paudel YP, Hwang I, Kim K-D, Kim S, Kim PM: **Computational design of potent D-peptide inhibitors of SARS-CoV-2**. *J Med Chem* 2021, **64**:14955–14967, https://doi.org/10.1021/acs.jmedchem.1c00655.

55. Amarasinghe KN, De Maria L, Tyrchan C, Eriksson LA, Sadowski J, Petrović D: **Virtual screening expands the non-natural amino acid palette for peptide optimization**. *J Chem Inf Model* 2022, **62**:2999–3007, https://doi.org/10.1021/acs.jcim.2c00193.

56. Shahab M, Khan SS, Zulfat M, Bin Jardan YA, Mekonnen AB, Bourhia M, Zheng G: **In silico mutagenesis-based designing of oncogenic SHP2 peptide to inhibit cancer progression**. *Sci Rep* 2023, **13**, https://doi.org/10.1038/s41598-023-37020-4.

57. Khan A, Umbreen S, Hameed A, Fatima R, Zahoor U, Babar Z, Waseem M, Hussain Z, Rizwan M, Zaman N, Ali S, Suleman M, Shah A, Ali L, Ali SS, Wei D-Q: **In silico mutagenesis-based remodelling of SARS-CoV-1 peptide (ATLQAIAS) to inhibit SARS-CoV-2: structural-dynamics and free energy calculations**. *Interdiscipl Sci Comput Life Sci* 2021, **13**:521–534, https://doi.org/10.1007/s12539-021-00447-2.

58. Ghobadi Z, Mahnam K, Shakhsi-Niaei M: **In-silico design of peptides for inhibition of HLA-A*03-KLIETYFSK complex as a new drug design for treatment of multiples sclerosis disease**. *J Mol Graph Model* 2022, **111**, 108079, https://doi.org/10.1016/j.jmgm.2021.108079.

59. Wood CW, Ibarra AA, Bartlett GJ, Wilson AJ, Woolfson DN, Sessions RB: **BAlaS: fast, interactive and accessible computational alanine-scanning using BudeAlaScan**. *Bioinformatics* 2020, **36**:2917–2919, https://doi.org/10.1093/bioinformatics/btaa026.

60. Ribeiro JV, Cerqueira NM, Moreira IS, Fernandes PA, Ramos MJ: **CompASM: an Amber-VMD alanine scanning mutagenesis plug**. *Theor Chem Acc* 2012, **131**:1271, https://doi.org/10.1007/s00214-012-1271-2.

61. Humphrey W, Dalke A, Schulten K: **VMD: visual molecular dynamics**. *J Mol Graph* 1996, **14**:33–38, https://doi.org/10.1016/0263-7855(96)00018-5.

62. Chan HTH, Moesser MA, Walters RK, Malla TR, Twidale RM, John T, Deeks HM, Johnston-Wood T, Mikhailov V, Sessions RB, Dawson W, Salah E, Lukacik P, Strain-Damerell C, Owen CD, Nakajima T, Świderek K, Lodola A, Moliner V, Glowacki DR, Spencer J, Walsh MA, Schofield CJ, Genovese L, Shoemark DK, Mulholland AJ, Duarte F, Morris GM: **Discovery of SARS-CoV-2 Mpro peptide inhibitors from modelling substrate and ligand binding**. *Chem Sci* 2021, **12**:13686–13703, https://doi.org/10.1039/d1sc03628a.

63. Hetherington K, Dutt S, Ibarra AA, Cawood EE, Hobor F, Woolfson DN, Edwards TA, Nelson A, Sessions RB, Wilson AJ: **Towards optimizing peptide-based inhibitors of protein–protein interactions: predictive saturation variation scanning (PreSaVS)**. *RSC Chem Biol* 2021, **2**:1474–1478, https://doi.org/10.1039/d1cb00137j.

64. McIntosh-Smith S, Price J, Sessions RB, Ibarra AA: **High performance in silico virtual drug screening on many-core processors**. *Int J High Perform Comput Appl* 2014, **29**:119–134, https://doi.org/10.1177/1094342014528252.

65. Vu QN, Young R, Sudhakar HK, Gao T, Huang T, Tan YS, Lau YH: **Cyclisation strategies for stabilising peptides with irregular conformations**. *RSC Med Chem* 2021, **12**:887–901, https://doi.org/10.1039/d1md00098e.

66. Marsault E, Peterson ML: **Macrocycles are great cycles: applications, opportunities, and challenges of synthetic macrocycles in drug discovery**. *J Med Chem* 2011, **54**: 1961–2004, https://doi.org/10.1021/jm1012374.

67. • Fonseca Lopez F, Miao J, Damjanovic J, Bischof L, Braun MB, *et al.*: **Computational prediction of cyclic peptide structural ensembles and application to the design of keap1 binders**. *J Chem Inf Model* 2023, **63**:6925–6937, https://doi.org/10.1021/acs.jcim.3c01337.

This MD simulation study explored the structural conformations of cyclic peptides. Aiming to mimic the binding interface of a PPI, the authors designed peptides that effectively inhibit Keap1, a protein involved in the cellular stress response, providing valuable insights into the development of potential cyclic peptide-based drugs.

68. Rettie SA, Juergens D, Adebomi V, Bueso YF, Zhao Q, Leveille AN, Liu A, Bera AK, Wilms JA, Üffing A, Kang A, Brackenbrough E, Lamb M, Gerben SR, Murray A, Levine PM, Schneider M, Vasireddy V, Ovchinnikov S, Weiergräber OH, Willbold D, Kritzer JA, Mougous JD, Baker D, DiMaio F, Bhardwaj G: **Accuratede novodesign of high-affinity protein binding macrocycles using deep learning**. *bioRxiv* 2024, https://doi.org/10.1101/2024.11.18.622547.

69. Bluntzer MTJ, O'Connell J, Baker TS, Michel J, Hulme AN: **Designing stapled peptides to inhibit protein-protein interactions: an analysis of successes in a rapidly changing field**. *Pept Sci* 2021, **113**(1), e24191, https://doi.org/10.1002/pep2.24191.

70. Li Y, Wu M, Fu Y, Xue J, Yuan F, Qu T, Rissanou AN, Wang Y, Li X, Hu H: **Therapeutic stapled peptides: efficacy and molecular targets**. *Pharmacol Res* 2024, **203**, 107137, https://doi.org/10.1016/j.phrs.2024.107137.

71. Barman P, Joshi S, Sharma S, Preet S, Sharma S, Saini A: **Strategic approaches to improvise peptide drugs as next generation therapeutics**. *Int J Pept Res Therapeut* 2023, **29**(4): 61, https://doi.org/10.1007/s10989-023-10524-3.

72. Akbarian M, Chen S-H: **Instability challenges and stabilization strategies of pharmaceutical proteins**. *Pharmaceutics* 2022, **14**:2533, https://doi.org/10.3390/pharmaceutics14112533.

73. Swanson K, Walther P, Leitz J, Mukherjee S, Wu JC, Shivnaraine RV, Zou J: **ADMET-AI: a machine learning ADMET platform for evaluation of large-scale chemical libraries**. *Bioinformatics* 2024, **40**, btae416, https://doi.org/10.1093/bioinformatics/btae416.

74. Fu L, Shi S, Yi J, Wang N, He Y, Wu Z, Peng J, Deng Y, Wang W, Wu C, Lyu A, Zeng X, Zhao W, Hou T, Cao D: **ADMETlab 3.0: an updated comprehensive online ADMET prediction platform enhanced with broader coverage, improved performance, API functionality and decision support**. *Nucleic Acids Res* 2024, **52**:W422–W431, https://doi.org/10.1093/nar/gkae236.