Site-specific fluorination of serine protease inhibitor BPTI for the investigation of protein-protein interactions

Inaugural-Dissertation
to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry, Pharmacy
of Freie Universität Berlin

by
Jakob Leppkes
from Singapore
May 2022

The research presented in this doctoral thesis was performed under the supervision of Prof. Dr. Beate Koksch from April 2017 until May 2022 at the Institute of Chemistry and Biochemistry in the Department of Biology, Chemistry, Pharmacy of Freie Universität Berlin.

1st reviewer:	Prof. Dr. Beate Koksch
2nd reviewer:	Prof. Dr. Bettina G. Keller
Date of defense:	28.06.2022

Declaration of Authorship

I hereby confirm that I have prepared this dissertation entitled “Site-specific fluorination of serine protease inhibitor BPTI for the investigation of protein-protein interactions” autonomously and without impermissible help. All external sources and resources have been specified and correctly cited or acknowledged. This thesis has not been submitted, accepted, rated as insufficient, or rejected in any other doctorate degree procedure.

Berlin, May 2022 Jakob Leppkes

Publications

- **J. Leppkes**, N. Dimos, B. Loll, T. Hohmann, M. Dyrks, A. Wieseke, B. G. Keller, B. Koksch, Fluorine-induced polarity increases inhibitory activity of BPTI towards chymotrypsin, *RSC Chem. Biol.* **2022**, *Accepted Manuscript* (DOI: doi.org/10.1039/D2CB00018K)
- A. A. Berger, **J. Leppkes**, B. Koksch, Investigation from the Belly of the Beast: N-Terminally Labeled Incretin Peptides That Are Both Potent Receptor Agonists and Stable to Protease Digestion, *ACS Cent. Sci.* **2021**, 7, 3, 400-402. (DOI: doi.org/10.1021/acscentsci.1c00265)
- **J. Leppkes**, T. Hohmann, B. Koksch, Improved enantioselective gram scale synthesis route to *N*-Fmoc-protected monofluoroethylglycine. *J. Fluorine Chem.* **2020**, 232, 109453. (DOI: doi.org/10.1016/j.jfluchem.2020.109453)
- J. Moschner, V. Stulberg, R. Fernandes, S. Huhmann, **J. Leppkes**, B. Koksch, Approaches to Obtaining Fluorinated α-Amino Acids. *Chem. Rev.* **2019**, 119, 10718-10801. (DOI: doi.org/10.1021/acs.chemrev.9b00024)

Oral presentations

- "Formation of networks between side-chain fluorinated amino acids and water molecules in protein binding pockets using the example of BPTI"; 2nd network meeting of the CRC 1349 "Fluorine-Specific Interactions: Fundamentals and Functions," Berlin, Germany, 28/08/**2020**.
- "Synthesis of fluorinated amino acids and their application for the fine-tuning of protein-protein interactions"; 1st Summerschool of the CRC 1349 "Fluorine-Specific Interactions: Fundamentals and Functions," Waren (Müritz), Germany, 09/09 – 10/09/**2019**.
- "Synthesis of fluorinated amino acids and their application for the fine-tuning of protein-protein interactions"; 19th European Symposium on Fluorine Chemistry, Warsaw, Poland, 25/08 – 31/08/**2019**.
- "Formation of networks between side-chain fluorinated amino acids and water molecules in protein binding pockets using the example of BPTI"; 1st network meeting of the CRC 1349 "Fluorine-Specific Interactions: Fundamentals and Functions," Berlin, Germany, 15/08/**2020**.

- "Formation of networks between side-chain fluorinated amino acids and water molecules in protein binding pockets using the example of BPTI" – speed talk; 8th Peptide Engineering Meeting (PEM-2018), Berlin, Germany, 08/11 – 10/11/**2018**.
- "Fluorinated Analogues of the bovine pancreatic trypsin inhibitor"; 9th Ph.D. workshop of the research training group 1582 "Fluorine as key element," Potsdam, Germany, 28/03 – 29/03/**2018.**

Posters

- "Synthesis of fluorinated amino acids and their application for the fine-tuning of protein-protein interactions"; Non-canonical amino acids (NcAAs): Tools for biological and biophysical investigations, Paris, France, 21/10 – 22/10/**2019**.
- "Formations of networks between side-chain fluorinated amino acids and water molecules in protein binding pockets using the example of BPTI"; 19th European Symposium on Fluorine Chemistry, Warsaw, Poland, 25/08 – 31/08/**2019**. (**Poster Prize**)
- "Formation of networks between side-chain fluorinated amino acids and water molecules in protein binding pockets using the example of BPTI"; 8th Peptide Engineering Meeting (PEM-2018), Berlin, Germany, 08/11 – 10/11/**2018**. (**Poster Prize**)
- "Formation of networks between side-chain fluorinated amino acids and water molecules in protein binding pockets using the example of BPTI"; 22nd International Symposium on Fluorine Chemistry, Oxford, UK, 22/07 – 27/07/**2018**. (**Poster Prize**)
- "Fluorinated Bovine Pancreatic Trypsin Inhibitor Analogues"; Red Hot MRI: Fluorine Imaging, Berlin, Germany, 15/10 – 16/10/**2017**.
- "Fluorinated Bovine Pancreatic Trypsin Inhibitor Analogues"; GDCh Wissenschaftsforum Chemie, Berlin, Germany, 10/09 – 14/09/**2017**.

Acknowledgments

I would like to express my gratitude and appreciation to Prof. Dr. Beate Koksch for her support and trust she has shown me throughout my time working for her. In times of doubt, she always managed to guide and inspire me. She encouraged me to pursue science with an open mind and gave me the freedom to approach my research self-determined. I am deeply grateful for her support to present the results of my research at numerous national and international research conferences and trusting me in representing her working group. Aside from that, I want to thank her for listening to me and valuing my opinion. Thank you Beate, I really enjoyed working for you the last few years!

I would also like to thank Prof. Dr. Bettina Keller for fruitful discussions in our meetings and for being the second supervisor of this dissertation.

I am grateful to Nicole Dimos and Dr. Bernhard Loll to be excellent and reliable cooperation partners. Without their work, the crystal structure analysis of all protease-inhibitor complexes would not have been possible. Thank you for the easy communication, your dedication, and your valuable input to this dissertation.

Furthermore, I would like to thank Leon Wehrhan and Prof. Dr. Bettina Keller for the great collaboration. Even though experimental and theoretical data sometimes seemed not to be in agreement with each other, they always had an open mind and came up with clever solutions.

Now, I would like to thank all my students, Thomas Hohmann, Michel Dyrks, and Ariane Wieseke for their valuable contributions to this dissertation. Thank you for doing a great job and being super easy to supervise. I definitely learned a lot from you!

It is hard to describe how much I want to thank Dr. Allison Berger for everything she has done in the past years. She was always there for critical discussion, had an incredible impact on this project and in the end proofread this thesis. Thank you, Allison, you're the best!

Especially, I would like to thank Dr. Ana Rita de Lima Fernandes, Dr. Susanne Huhmann, Dr. Dorian Mikolajczak, Dr. Johann Moschner, Suvrat Chowdhary and Thomas Hohmann for being fantastic colleagues and great lab partners with whom I could not only discuss science but also everything else.

I would like to thank everyone else from the Koksch group for the nice work atmosphere, stimulating discussions and their commitment to science and the working group.

I would like to thank the Deutsche Forschungsgemeinschaft in context of the research training group 1582 “Fluorine as Key Element” and the collaborative research center 1349 “Fluorine-specific interactions” for financial support.
I would like to thank the Core Facility BioSupraMol for the countless samples they measured and helping me solve numerous complex analytical issues.
Thank you to all my friends for being there for me in the times I needed them. I love you guys.
Thank you to all Rumgedoktore members, Jessi, Tony, Martin, Fenja, Caro, Felix, Julian, Dogus, Jonas, Anna, Frenio, Alex, Peter, Tugrul, Hannes, and Marius. I am not sure if I would have had such a great time without you.
I would like to thank my family for enabling this PhD by being supportive throughout my entire life and introducing me into the fascinating world of science and chemistry.
Lastly, I want to thank my partner Franzi. You were always there for me in the last years and went with me through high and low. I am incredible grateful to have you by my side. I love you!

Abstract

Proteins are important biomolecules for the regulation of quintessential mechanisms in living organisms. Through chemical modifications, their biophysical properties can be altered, leading to enhanced stability, higher bioactivity, and other beneficial alterations. In this doctoral thesis, the influence of site-specific fluorination on the activity of the bovine pancreatic trypsin inhibitor (BPTI), a well-studied serine protease inhibitor, was investigated. For this, new strategies for synthesizing a fluorinated amino acid and the chemical protein synthesis of BPTI were developed, followed by biological activity assays and structural investigations X-ray diffraction analysis.

In the first part of this project, a novel synthesis in a gram scale for monofluoroethylglycine (MfeGly) was developed. This eight-step strategy started with aspartic acid and obtained the Fmoc-protected product in yields of up to 50 %. A particular focus of this study laid in the investigation of various nucleophilic fluoride sources.

The second part of this project focused on incorporating the previously synthesized amino acid and several other derivatives into 58-amino acid long BPTI by solid-phase peptide synthesis (SPPS). Since fluorinated amino acids are of high value, special coupling cycles for their incorporation into amino acid sequence were developed, cutting usage of the material to 30 %. SPPS was followed by oxidative refolding the cysteine-rich protein to obtain the native conformation with correctly matched disulfide bridges. The inhibitory activity of fluorinated BPTI variants and their hydrocarbon derivatives was investigated towards the serine protease α-chymotrypsin. Difluoroethylglycine (DfeGly) showed enhanced inhibition, even better than wild-type BPTI. In contrast, inhibition activity drops by two orders of magnitude in non-fluorinated amino butanoic acid (Abu) containing BPTI. Lastly, BPTI variants containing Abu and its fluorinated derivatives MfeGly, DfeGly, and trifluoroethylglycine (TfeGly) were crystallized in complex with chymotrypsin and analyzed by X-ray diffraction.

The results of this thesis will contribute to a better understanding of how fluorinated amino acids impact protein-protein interactions.

Kurzzusammenfassung

Proteine sind wichtige Biomoleküle für die Regulierung grundlegender Mechanismen in lebenden Organismen. Durch chemische Modifikationen können ihre biophysikalischen Eigenschaften verändert werden, was zu verbesserter Stabilität, höherer Bioaktivität und anderen vorteilhaften Veränderungen führen kann.

In dieser Dissertation wurde der Einfluss der ortsspezifischen Fluorierung auf die Aktivität des bovinen pankreatischen Trypsin-Inhibitors (BPTI), eines gut untersuchten Serinprotease-Inhibitors, untersucht. Dazu wurden neue Strategien für die Synthese einer fluorierten Aminosäure und die chemische Proteinsynthese von BPTI entwickelt, gefolgt von biologischen Aktivitätstests und Strukturuntersuchungen mittels Röntgenbeugungsanalyse.

Im ersten Teil dieses Projekts wurde eine eine neuartige Synthese im Gramm-Maßstab für Monofluorethylglycin (MfeGly) entwickelt. In dieser achtstufigen Strategie, die mit Asparaginsäure begann, konnte das Fmoc-geschützte Produkt in Ausbeuten von bis zu 50 % erhalten werden. Ein besonderer Schwerpunkt dieser Studie lag auf der Untersuchung verschiedener nucleophiler Fluoridquellen.

Der zweite Teil dieses Projekts konzentrierte sich auf den Einbau der zuvor synthetisierten Aminosäure und mehrerer anderer Derivate in BPTI, einem aus 58 Aminosäuren bestehenden Protein, durch Festphasen-Peptidsynthese (SPPS). Aufgrund des limitierten Zugangs zu fluorierten Aminosäuren, wurden spezielle Kopplungszyklen für deren Einbau in die Aminosäuresequenz entwickelt, wodurch der Verbrauch des Materials auf 30 % gesenkt werden konnte. Nach der SPPS erfolgte eine oxidative Rückfaltung des cysteinreichen Proteins, um die native Konformation mit korrekt ausgebildeten Disulfidbrücken zu erhalten. Die fluorierten BPTI-Varianten und ihre Kohlenwasserstoffderivate wurden im Hinblick auf ihre inhibitorische Aktivität gegenüber der Serinprotease α-Chymotrypsin untersucht. Dabei wurde festgestellt, dass Difluorethylglycin (DfeGly) eine verstärkte Hemmung zeigte, die sogar besser war als die des Wildtyp-BPTIs, während die Hemmungsaktivität bei nicht fluoriertem, Aminobutansäure (Abu) enthaltendem BPTI um zwei Größenordnungen abnahm. Schließlich wurden BPTI-Varianten, die Abu und seine fluorierten Derivate MfeGly, DfeGly und Trifluoroethylglycin (TfeGly) enthalten, im Komplex mit Chymotrypsin kristallisiert und mittels Röntgenbeugung analysiert.

Die Ergebnisse dieser Arbeit werden dazu beitragen, besser zu verstehen, wie fluorierte Aminosäuren Protein-Protein-Wechselwirkungen beeinflussen können.

Table of Contents

List of Abbreviations

2-/3-fTyr	2-/3-fluorotyrosine
5'-FDA	5'-fluoro-5'-deoxyadenosine
AA	amino acid
Abu	L-2-aminobutanoic acid
AChE	acetylcholinesterase
AI	artificial intelligence
AMP	antimicrobial peptide
BA*p*NA	N_α-Benzoyl-L-arginine 4-nitroanilide
BPTI	bovine pancreatic trypsin inhibitor
BTEE	N_α-Benzoyl-L-tyrosine ethyl ester
calc.	calculated
CAT	chloramphenicol acetyltransferase
CD	circular dichroism
DAST	diethylaminosulfur trifluoride
DfeGly	difluorethylglycine / (S)-2-amino-4,4-difluorobutanoic acid
DfpGly	difluorpropylglycine / (S)-2-amino-4,4-difluoropentanoic acid
DIC	diisopropylcarbodiimide
DIPEA/DIEA	*N*, *N*-diisopropylethylamine (Hünig's base)
DMF	*N*, *N*-dimethylformamide
DNA	deoxyribonucleic acid
DPP4	dipeptidyl peptidase 4
ESI	electrospray ionization
FA	fluoroacetate
FAA	fluoroacetaldehyde
FDA	U. S. Food and Drug Administration
Fmoc	9-*N*-fluorenylmethyloxycarbonyl
GdmCl	guanidinium chloride
GFP	green fluorescent protein
GSH	glutathione (reduced)
GSSG	glutathione (oxidized)
HFA	hexafluoroacetone
HfLeu	5,5,5,5',5',5'-hexafluoroleucine

HIV	human immunodeficiency virus
HPLC	high-performance liquid chromatography
HRMS	high-resolution mass spectrometry
Hse	homoserine
IR	Infrared
ITC	isothermal titration calorimetry
LDA	lithium diisopropylamide
m/z	mass-to-charge ratio
MD	molecular dynamics
MeCN	acetonitrile
MfeGly	monofluorethylglycine / (S)-2-amino-4-fluorobutanoic acid
MWCO	molecular weight cut-off
ncAA	non-canonical amino acid
NCL	native chemical ligation
NMP	*N*-methyl-2-pyrrolidone
NMR	nuclear magnetic resonance
obs.	observed
PDB	Protein Data Bank
PDI	protein-disulfide isomerase
PET	positron emission tomography
PfpGly	pentafluorpropylglycine / (S)-2-amino-4,4,4-trifluoro-4,4-difluoropentanoic acid
PTFE	polytetrafluoroethylene
RP	reversed-phase
SAM	*S*-adenosyl-L-methionine
SPPS	Solid-phase peptide synthesis
TCEP	tris-(2-carboxyethyl) phosphine
TfeGly	trifluorethylglycine / (S)-2-amino-4,4,4-trifluorobutanoic acid
TfLeu	5',5',5'-tirfluoroleucine
TfMet	trifluoromethionine
TLC	thin-layer chromatography
ToF	time of flight
Ub	ubiquitin

UV	ultraviolet
vdW	van der Waals

1 Introduction

Life is based on a complex interplay of biological processes in which biomolecules specifically interact and react with each other. A central class of these biopolymers, nature has evolved, are proteins and peptides. These biopolymers are responsible for vital processes like DNA replication, bio-catalysis of metabolic reactions, providing structure to cells and organisms, and the transportation of molecules among an organism. Together with other biomolecules, they perform and control signal transduction. Other functions include the regulation of respiration, reproduction, and digestion, to name only a few.[1]

The basis of this variety of functions is attributed to the chemical diversity of functional groups in amino acids, ranging from aliphatic and aromatic hydrophobic, over polar uncharged to charged moieties. Even though over 300 natural amino acids are known, only 20 are encoded by the DNA, the canonical α-L-amino acids.[2] These amino acids form polypeptides, consisting of specific amino acid sequences, which develop into unique three-dimensional structures adopting particular biological activities. However, most proteins need cofactors or a post-translational modification to function fully. Due to that, protein engineering has become an essential part of biochemical research to develop novel chemical approaches to synthesize proteins with improved or unique bioactivities. Ever since Henri Moissan first isolated elemental fluorine in 1886,[3] this element had been of particular interest for a vast scene of scientists.[4] In the world of protein engineering, fluorinated amino acids have emerged as a powerful tool to modulate proteins' biochemical and biophysical properties.[5]

Approaches in this topic of research have led to the discovery of compounds with a broad range of applications, such as anticancer and antiviral agents, antibiotics, antidepressants, or enzyme inhibitors.[6] Fascinating is the fact that the integration of only a single fluorinated amino can significantly impact the properties of a protein.[7] Yet, there is no complete understanding of fluorine's impact on the structure and activity of proteins, which makes it difficult to predict the influence of hydrogen replacement by fluorine on protein stability, protein-protein interactions,[8] and other biochemical and pharmacological properties.[9]

2 Fluorine – an element with unique properties

2.1 The nature of the C-F bond

The carbon-fluorine bond is highly polarized,[10] due to the high electronegativity of fluorine. This results in an ionic rather than covalent character in the C-F bond, which leads to a significant dipole moment and dipolar/quadrupolar interactions with other dipoles in close proximity to the bond. This is why electrostatic interactions can explain conformations adopted by organofluorine compounds.

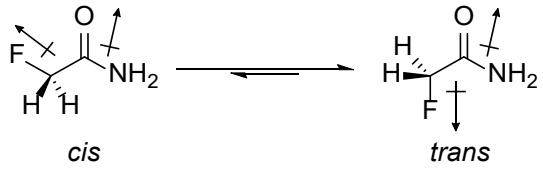
When looking at the strength of the C-F bond, the high difference in electronegativity between the fluorine and the carbon atom is of importance (Table 2.1). The resulting dipole moment in this bond leads to increased localization of electrons around the fluorine atom, leading to a high partial positive charge at the carbon atom. The partial charges of $F^{\delta-}$ and $C^{\delta+}$ are attractive, which strengthens the bond. With regard to the van der Waals (vdW) radius of fluorine (1.47 Å), it is slightly larger than hydrogen (1.20 Å) and almost as big as oxygen (1.52 Å).[11] Also, in terms of bond length, the C-F bond (1.35 Å) is longer than the C-H (1.09 Å) and similar in size to the C-O single bond (1.43 Å).[10]

Table 2.1. Physical properties of fluorine atom and C-F bond compared to biological relevant elements and other halogens[11]

Element	Electro-negativity [χ_P]	vdW radius [Å]	C-X bond strength [kJ mol^{-1}]	Polariza-bility [Å^3]	Ionization potential [kJ mol^{-1}]	Electron affinity [kJ mol^{-1}]
H	2.20	1.20	413.4	0.67	1312.1	74.1
C	2.55	1.70	347.7	1.76	1006.3	121.3
N	3.40	1.55	291.6	1.10	1402.1	-25.9
O	3.44	1.52	351.5	0.82	1313.8	141.4
F	**3.98**	**1.47**	**441.0**	**0.56**	**1681.1**	**332.6**
Cl	3.16	1.75	328.4	2.18	1251.0	348.5
Br	2.96	1.85	275.7	3.05	1139.7	303.8
I	2.66	1.98	240.2	4.70	1009.2	295.4

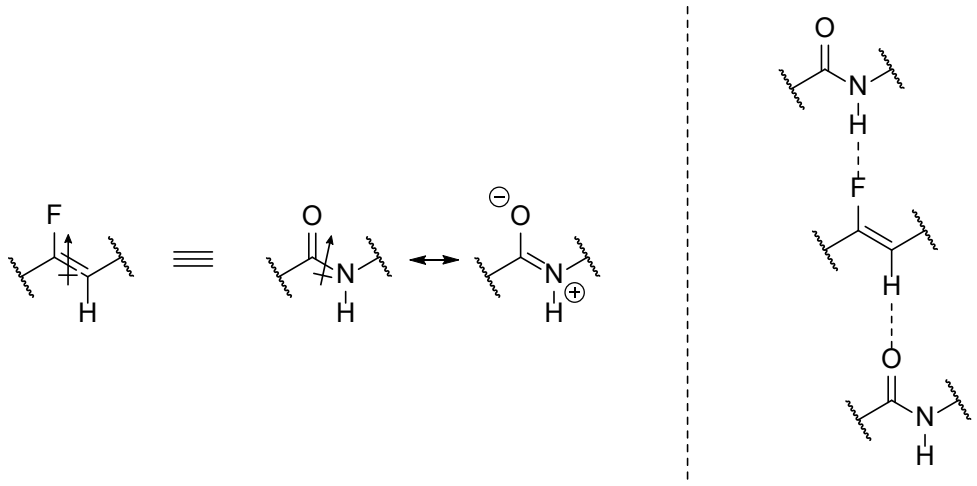
In organofluorine chemistry, several electronic and steric effects are essential to notice to better understand specific systems' behavior. Due to its ionic character, the C-F bond has a large dipole moment (μ). When looking at different fluorination degrees in methane, μ is 1.85 D fluoromethane, increasing to 1.97 D in difluoromethane and 1.65 D in trifluoromethane. In general, the dipole moment of the C-F bond is significant for

determining the conformation in organofluorine compounds. An example of a dipole-dipole interaction can be found in α-fluoroamide shown in Scheme 2.1, the *trans*- is preferred over the *cis*-conformer with a calculated energy difference of 31.4 kJ mol^{-1}, with the dipoles of the amide carbonyl and the C-F bond opposing each other.[12]



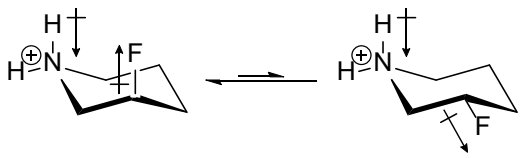
Scheme 2.1. Different conformers of α-fluoroamide with dipoles indicated as arrows.

Despite the absence of hydrogen bonding capacity, fluoroolefins have emerged as promising steric and polar hydrophobic mimetics of amide bonds. [13-15] The vinylfluoride dipole is oriented similar to the amide dipole, and thus the moieties have proven to be successful mimics in medicinal chemistry studies. [16] Oishi *et al.* showed that after incorporating fluoroalkenes into an α-helical peptide, its structure and anti-HIV activity were retained (Scheme 2.2).[17]



Scheme 2.2. Fluoroolefin as amide bond mimetic (left) and its participation in hydrogen bonding.

The strongest intramolecular interactions of the C-F bond are found in charge-dipole interactions between C-F and a formally charged heteroatom. Sun *et al.* discussed the observation of a significant axial preference (22.7 kJ mol^{-1}) over the equatorial position in 3-fluoropiperidinium ion.[18]



Scheme 2.3. Charge-dipole interactions between C-F and formally charged heteroatom.

The *gauche* effect is another interesting phenomenon in organofluorine chemistry. In 1,2-difluoroethane, the *gauche* is preferred over the *anti* conformation (Figure 2.1),[19] by 2.1 to 3.7 kJ mol^{-1} based on gas-phase calculations and experimental observations.[20-22] These findings seem unreasonable as one would predict repulsive forces between the two fluorine atoms, favoring the *anti* conformation as is the case for in 1,2-dichloroethane and 1,2-dibromoethane. A model for explaining the *gauche* effect is the hyperconjugation model: two stabilizing $\sigma_{CH}\cdots\sigma^*_{CF}$ interactions are supported in the *gauche* conformer, whereas in the *anti* conformer, a relative electron-deficient C-F bond is now anti to the σ^*_{CF} orbital, not allowing the exact extent of hyperconjugation to happen.

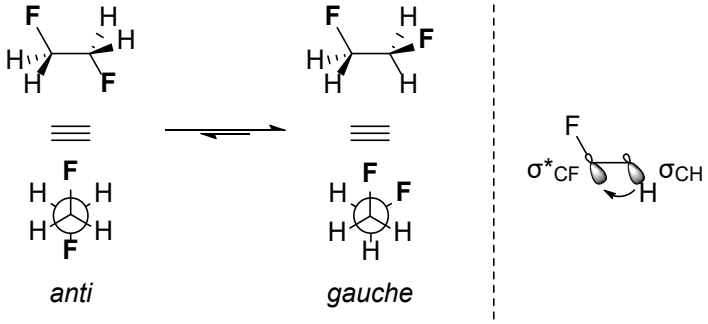
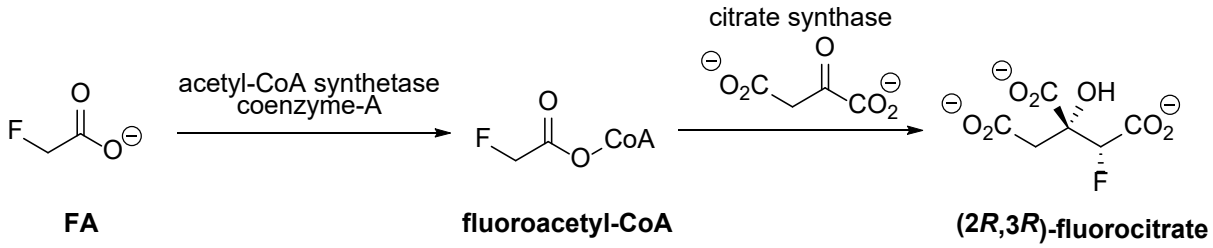


Figure 2.1. 1,2-Difluoroethane in Natta and Newman projection, *gauche* conformation is preferred over *anti*.

The three electron lone pairs at the fluorine atom are in close proximity to the core due to the atom's high electronegativity, which leads to reduced participation of fluorine as hydrogen bond acceptor, compared to elements like oxygen or nitrogen. Even though stated by Dunitz in the late 1990s, "Organic fluorine hardly ever accepts hydrogen bonds,"[23] by now, several groups have given experimental and theoretical evidence of the existence of these interactions.[24-31] Hydrogen bonds, in which organofluorine is functioning as an acceptor, are significantly weaker (8 – 13 kJ mol^{-1}) than conventional hydrogen bonds (21 – 42 kJ mol^{-1}). [10, 32] A reasonable explanation for this might be the low polarizability and high electron affinity of fluorine, resulting in a reluctance of donating electrons in a hydrogen bond. This is why in literature, it is proposed to speak about "weak dipolar interactions" rather than "hydrogen bonds".[23, 24, 33]

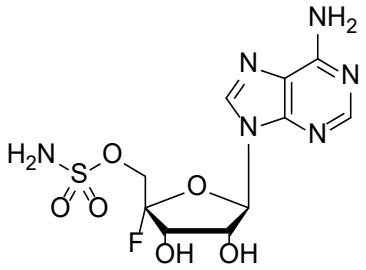
2.2 Fluorine in natural organisms

In natural organisms, like plants, fungi, and micro-organisms, more than 3000 halogenated compounds have been identified to date.[34] Only 13 of these structures carry a fluorine atom, which is surprising since this element is the most abundant in the earth's crust among all halogens.[35] Fluoroacetate was first isolated in 1943 from *Dichapetalum cymosum*, a poisonous shrub found in southern Africa, with fresh leaves containing concentrations of up to 250 ppm.[36] Until today, FA was found in more than 40 other organisms, all growing in tropical or subtropical regions of Africa, Australia, and South America. The toxicity of FA is caused by its metabolism. FA is transformed into fluoroacetyl-CoA, followed by a reaction with oxaloacetate and citrate synthase forming (2*R*,3*R*)-fluorocitrate. This molecule inhibits citrate transport, and its metabolite, (*R*)-4-hydroxy-*trans*-aconitate, inhibits the citric acid cycle.[35]



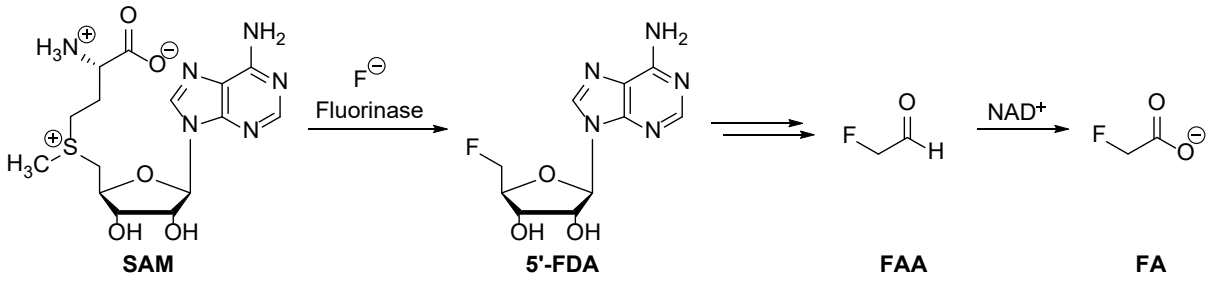
Scheme 2.4. Transformation of fluoroacetate to fluorocitrate

Nucleocidin is a fluorine-containing natural metabolite with broad-spectrum antibiotic activity; however, dit is not suitable for clinical use due to its toxicity.[35] It was first detected in 1956 in culture extracts from *Streptomyces calvus*, followed by purification from large-scale culture and structural characterization by NMR spectroscopy and mass spectrometry in the late 1960s.[37] Six decades after the first discovery of nucleocidin, Zhu *et al.* unrevealed the dependence of its production on a gene named *bldA*.[38] Previous attempts by multiple research groups to reisolate nucleocidin from *S. calvus* cultures were unsuccessful due to a mutation in *bldA*, making it non-functional. Only providing a functional *bldA* copy to the strain restored nucleocidin synthesis.[39]



Scheme 2.5. Chemical structure of fluorine-containing nucleocidin

In 2002, O'Hagan and coworkers first described the discovery of a fluorinase enzyme.[40] The enzyme, 5'-fluorodeoxyadenosine synthase, was first isolated a year later from *Streptomyces cattleya; it* was found out that, in the presence of potassium fluoride, it catalyzes the conversion of *S*-adenosyl-L-methionine (SAM) to 5'-fluoro-5'-deoxyadenosine (5'-FDA) by an S_N2-type nucleophilic substitution,[41] as well as the synthesis of 5'-FDA into FA (Scheme 2.6).[42]



Scheme 2.6. Conversion of SAM to 5'-FDA and FA by 5'-fluoro-5'-deoxyadenosine synthase from *S. cattleya*

Numerous studies investigated the exact mechanism of the 5'-FDA synthesis catalyzed by the fluorinase from *S. cattleya,* as reviewed by O'Hagan and Deng.[43] Based on computational studies, it was shown that the fluoride ion is engaging three hydrogen bonds with the enzyme, with the hydroxy functionality in the side chain and the amide NH in the backbone of Ser180, and the hydroxy functionality of Thr80. Furthermore, it engages an electrostatic interaction with partially positive charged 5'-carbon of SAM (Figure 2.2).

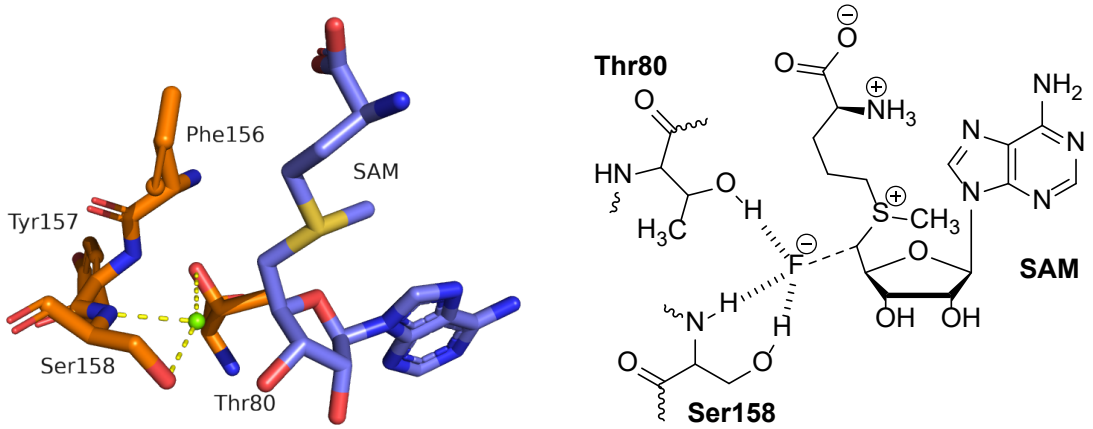


Figure 2.2. Activation of fluoride ion by fluorinase prior to S_N2 reaction with SAM represented in two schemes. Fluoride ion is modeled into SAM-fluorinase complex, derived from X-ray structure (PDB code: 1RQP) (**left**); Scheme of complex, with selected hydrogen bonds shown (**right**).

The enzyme desolvates the fluoride anion through these interactions, enabling a nucleophilic attack even from an aqueous environment. The desolvation of the anion is crucial since F^- has the highest heat of hydration among all halides.[43, 44] In other halogenation reactions in biological systems, halide ions (X^-) are being oxidized to

halonium ions (X^+) or halide radicals ($X^\cdot$),[45]. Still, due to the high electronegativity of fluorine, oxidation is precluded.

In the meantime, genes encoding for 5'-FDA synthase (*flA*) have been discovered in four more organisms, *Streptomyces xinghaiensis*, *Streptomyces* sp. MA37, *Actinoplanes* sp. N902-109, and *Nocardia brasiliensis*. However, fluorinase activity has not yet been detected in the case of *A.* sp. N902-109 and *N. brasiliensis* since either no efforts of cultivating the strain have been described so far, or the amount of detectible fluorine-bearing metabolites has not been sufficient for further analyses.[46] The discovery of fluorinases has resulted in multiple applications in drug development or diagnostic probing. For instance, Eustáquio *et al.* replaced the chlorinase gene *salL* in *Salinispora tropica* with the fluorinase gene *flA*, resulting in the production of fluorosalinosporamide instead of the chlorinated anti-cancer drug salinosporamide.[47] Martarello *et al.* showed initial studies in which fluorinase mediated the C-^{18}F bond formation as a potential tool for generating PET tracers.[48] This method has since further been investigated in numerous studies.[49-59]

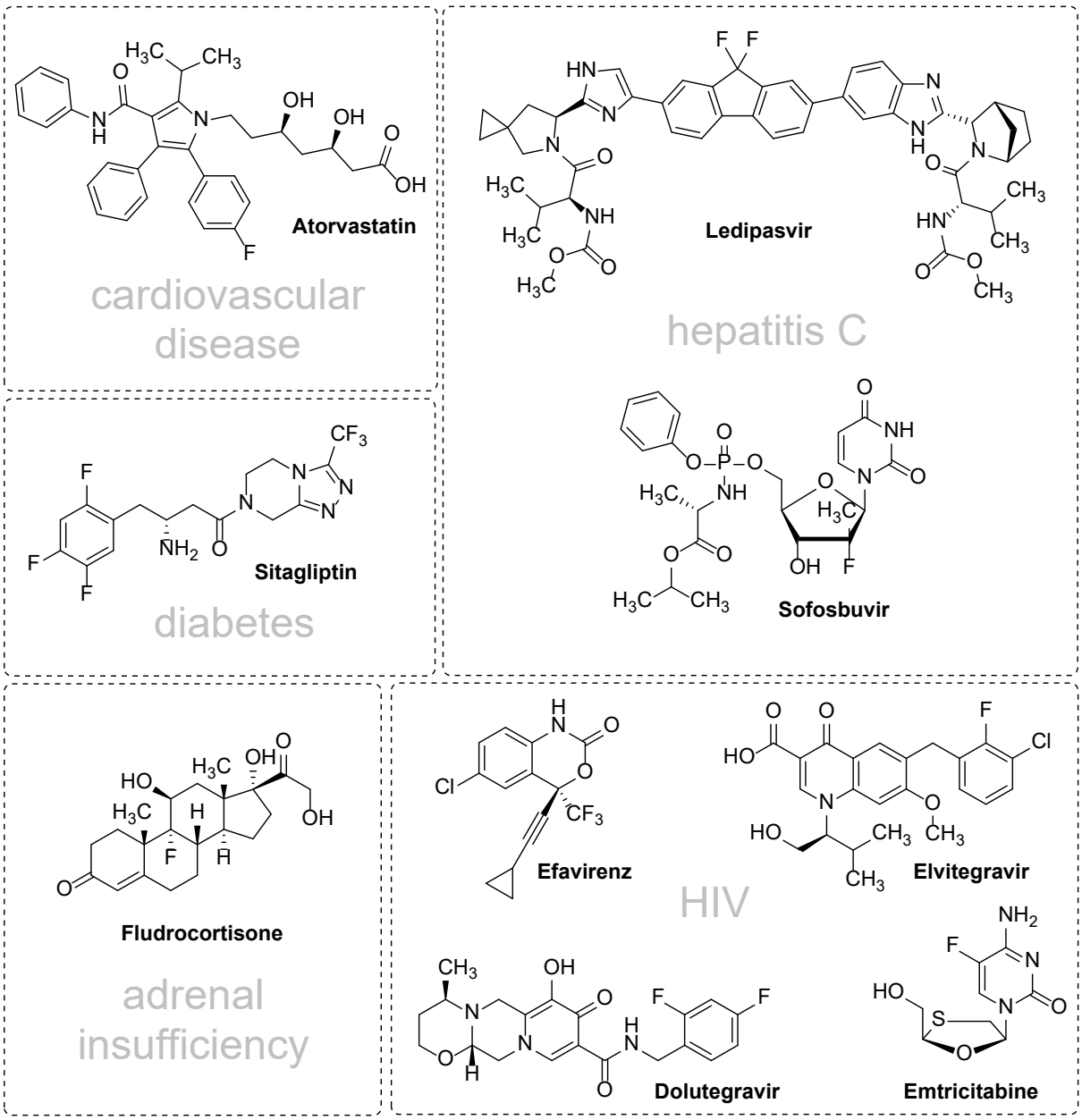
2.3 Fluorine in agrochemicals and drug design

Chemists have harnessed the power of fluorine for decades by introducing it into organic molecules. Fluorine can play a crucial role when employed in agrochemicals or pharmaceuticals. This is why, today, 53 % of newly registered agrochemicals[60] and 23 % of newly registered pharmaceuticals[61] contain at least one fluorine atom and a multitude of today's blockbuster drugs contain fluorine in some form. [62] By introducing a specific fluorinated functional group, the properties of the molecules, like their acidity, hydrophobicity, binding affinity, metabolic stability, and conformation, can be altered.

Fluorinated functionalities can be found in drugs administered for a broad range of diseases; some selected examples are shown in Scheme 2.7. Prominent among these drugs is atorvastatin, marketed as Lipitor by Pfizer. This statin is used to prevent cardiovascular disease. It is considered the most profitable drug of all time, with sales revenues of over $ 120 billion.[63] Since it came off patent in 2011, it is not a leading revenue generator anymore; however, it remains amongst the most prescribed drugs worldwide.[62]

Besides in atorvastatin, fluorine can be found in sitagliptin, treating diabetes as a dipeptidyl peptidase-4 (DPP4) inhibitor, in ledipasvir and sofosbuvir employed as antiviral agents in hepatitis C therapy, in the corticosteroid fludrocortisone to treat

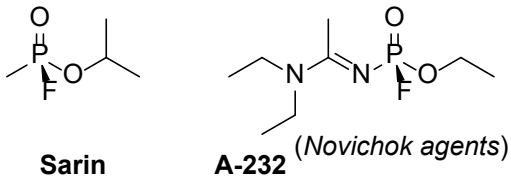
adrenal insufficiency and in efavirenz, elvitegravir, dolutegravir and emtricitabine as anti-HIV medication.[62]



Scheme 2.7. Selected fluorine-containing pharmaceuticals grouped by their activity against diseases.

There are numerous reasons why fluorine is introduced into drugs. Substituting a C-H with a C-F bond can increase the metabolic stability of a drug since the compound is not as rapidly oxidatively metabolized by liver enzymes, thus increasing the bioavailability of the molecule by blocking metabolically labile sites.[64, 65] Fluorination of a drug can alter the pK_a values of neighboring functional groups, which can impact the binding affinity of the drug to its target receptor.[65] an exciting aspect of the impact of fluorination is altered hydrophobicity. It is important to note that fluorination of organic molecules does not always lead to increased hydrophobicities compared to their hydrocarbon analogue.

While fluorinated aromatic and olefinic compounds show increased hydrophobicity, especially fluorination of aliphatic moieties can impart decreased hydrophobicity; one example of this is terminally fluorinated alkanes due to fluorine's strong electron-withdrawing effect.[66] These features are employed to fine-tune the hydrophobicity of drugs to pass through cell membranes without getting stuck in the bilayer's core.[64]



Scheme 2.8. Chemical structures of nerve agents Sarin and A-232.

Fluorine's ability to act as a leaving group has not only been used in applications beneficial for human health. The nerve agent sarin, developed by Gerhard Schrader in 1938, developed initially as organophosphate insecticide, is arguably one of the deadliest nerve agents that have been used.[67] More recently, A-232, a *Novichok agent* (Scheme 2.8), has been spotlighted after the attempted murder of a former Russian spy in Great Britain.[68] The organophosphorus compounds act as irreversible acetylcholinesterase (AChE) inhibitors, Ser200 of AChE attacks the phosphorus atom, forming a stable phosphonate, while fluoride acts as a leaving group.

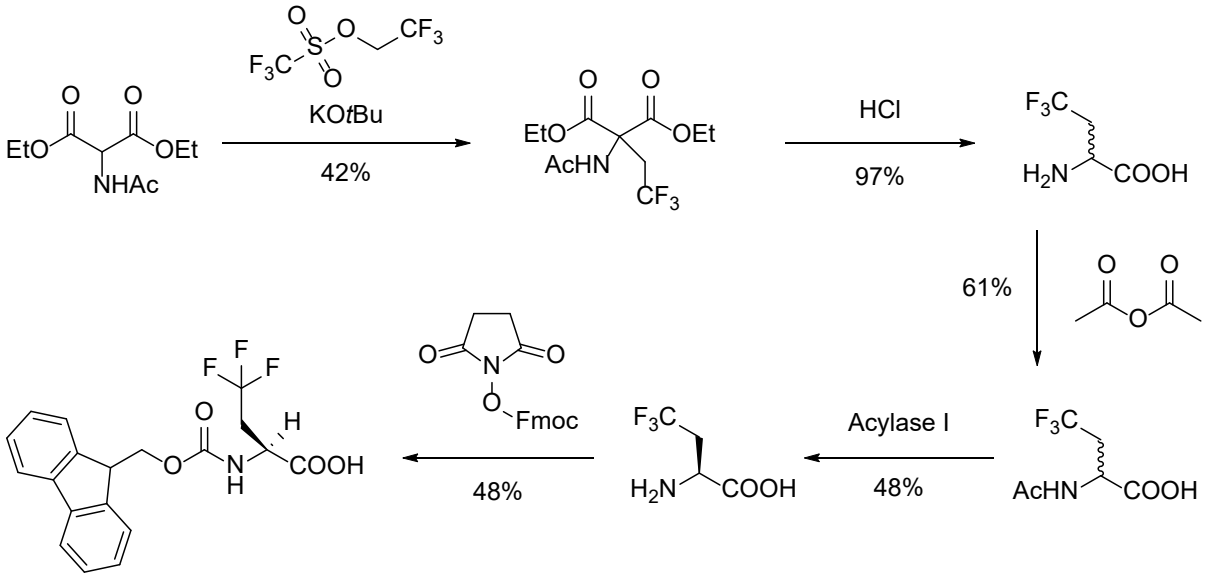
2.4 Fluorinated amino acids as tools in peptide engineering

The fascinating and often beneficial impact fluorine can have on small molecules' biological, chemical, and physical properties is also observed when modifying peptides and proteins with this element. In these cases, fluorinated derivatives of amino acids are incorporated into these biological systems by a wide variety of techniques.

2.4.1 Synthesis of fluorinated amino acids

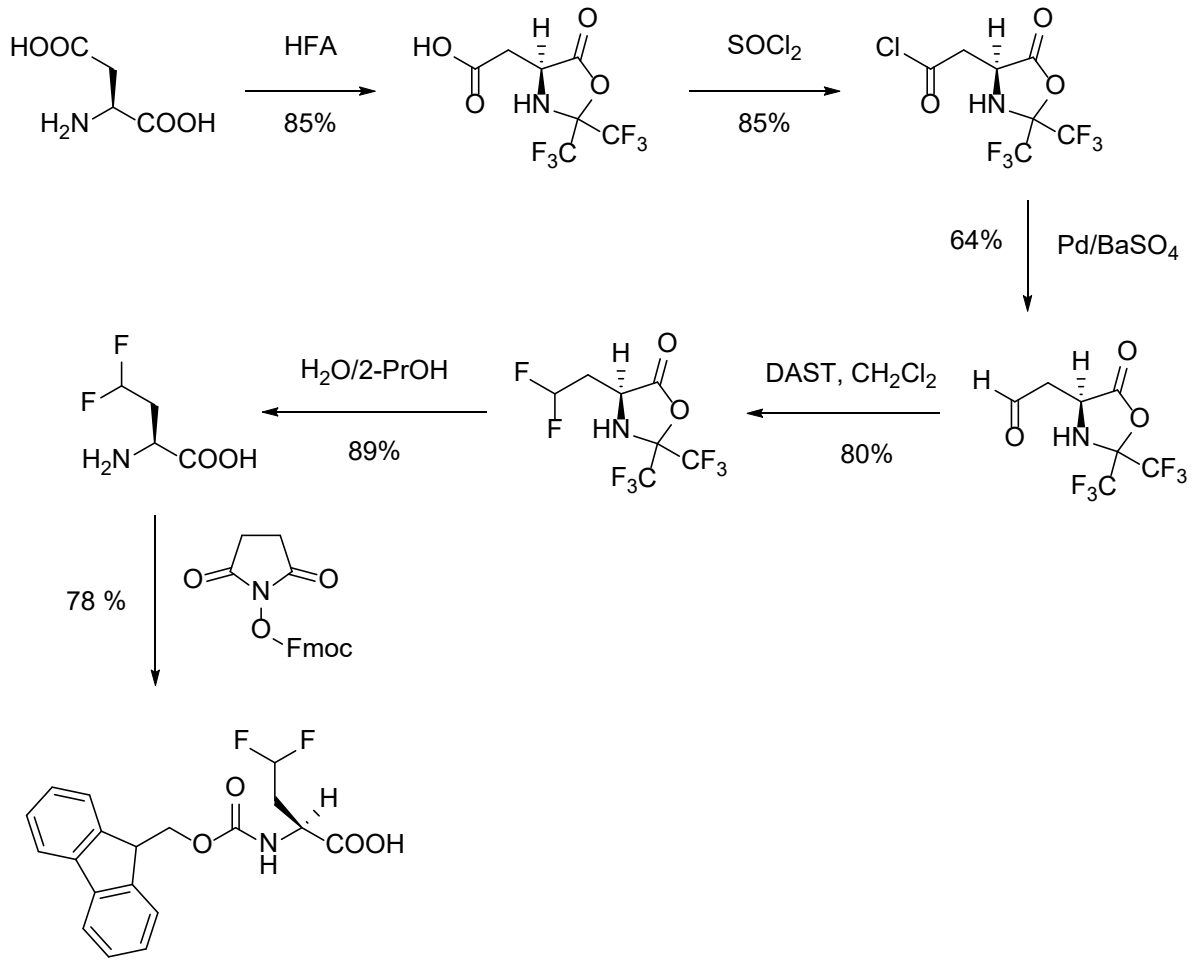
Before incorporating these amino acids into peptides and proteins, access to these non-natural building blocks is often necessary. In many cases fluorinated amino acids are commercially not available and need to be synthesized. This happens in two different approaches, either by electrophilic, nucleophilic or radical fluorination reagents with the corresponding precursors or by the introduction of fluorinated building blocks under the formation of new carbon-carbon bonds. In general, these syntheses are special for every desired fluorinated amino acid and require high synthetic efforts to be generated. In 2019, we reviewed approaches to obtain fluorinated α-amino acids.[69] In the following section,

the synthesis pathways of fluorinated derivatives of Abu, an amino acid of importance for this thesis, are briefly discussed.

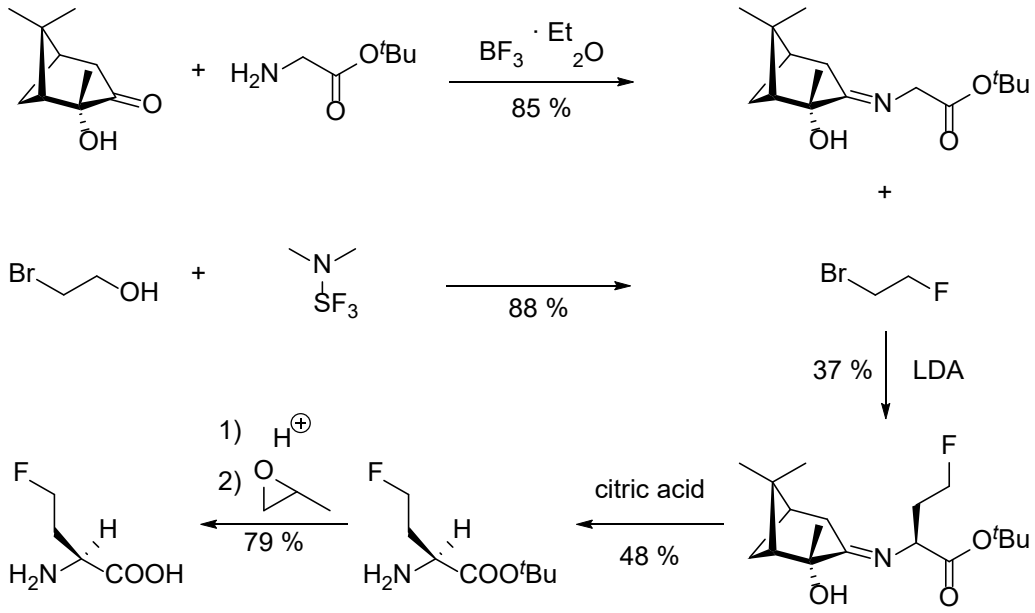


Scheme 2.9. Synthesis of Fmoc-TfeGly by Tsushima *et al.*.[70]

To synthesize Fmoc-protected TfeGly, Tsushima *et al.* described a strategy starting from diethyl *N*-acetamidomalonate (Scheme 2.9). After deprotonation with potassium tert-butoxide, α-(*N*-acetylamino)-α-(2,2,2-trifluoroethyl)malonate is formed by the treatment with 2,2,2-trifluoroethyl trifluoromethanesulfonate.[70] Racemic TfeGly is afforded by hydrolysis under acid conditions. The separation of the two enantiomers is accomplished by acylase I. This metalloenzyme can selectively deacetylate the L-enantiomer,[71] which can be separated from D-Ac-TfeGly by ion-exchange chromatography. In a final step, the enantiopure L-amino acid is Fmoc-protected for further application in Fmoc-based SPPS. For the difluorinated variant of Abu, DfeGly, Winkler, and Burger presented an approach starting from aspartic acid (Scheme 2.10).[72] The acid and the amino functionality of Asp are protected for synthesis by treatment with hexafluoroacetone (HFA). Treatment of the HFA-protected amino acid with thionyl chloride affords the corresponding acid chloride, which is reacted to the aldehyde in a Rosenmund reaction with Pd/$BaSO_4$. HFA-DfeGly is formed in nucleophilic fluorination with diethylaminosulfur trifluoride (DAST), followed by hydrolysis of the HFA-protecting group and Fmoc-protection.



Scheme 2.10. Synthesis of Fmoc-DfeGly by Winkler and Burger. [72]



Scheme 2.11. Synthesis of MfeGly by Haufe *et al.*.[73]

Haufe and coworkers first described the synthesis of enantiopure MfeGly.[73] Starting from (+)-2-hydroxy-3-pinanone, a Schiff base is formed by reaction with glycine *tert*-butyl ester. Following this, alkylation of the Schiff base with 1-bromo-2-fluoroethane, generated from the reaction of 2-bromoethanol with dimethylaminosulfur trifluoride[74] after deprotonation with lithium diisopropylamide (LDA), proceeds with high diastereoselectivity. The enantiopure L-amino acid is obtained after mild acidic deprotection. A Fmoc-protected MfeGly has not been described in the literature so far.

2.4.2 Physical and chemical properties of fluorinated amino acids

The three amino acids described above have proven to be well-suited model systems when studying the impact of varying degrees of fluorination on amino acids and subsequently incorporated in peptides and proteins. As previously mentioned in chapter 2.3, hydrophobicity is an important property when looking at fluorinated molecules, primarily in amino acids. The assumption that fluorination leads to increased hydrophobicity cannot generally be applied. Koksch and coworkers investigated the hydrophobic properties of Abu and its fluorinated derivatives in an RP-HPLC-based assay (Figure 2.3, B). It was shown that single fluorination of Abu decreases the hydrophobicity of the amino acid, whereas di- and trifluorination increases hydrophobicity. Thus, the trend in polarity among the Abu variants is the following MfeGly > Abu > DfeGly > TfeGly. The difference of MfeGly can be explained by the polarized geminal C-H bonds, introducing a dipole moment.[75]

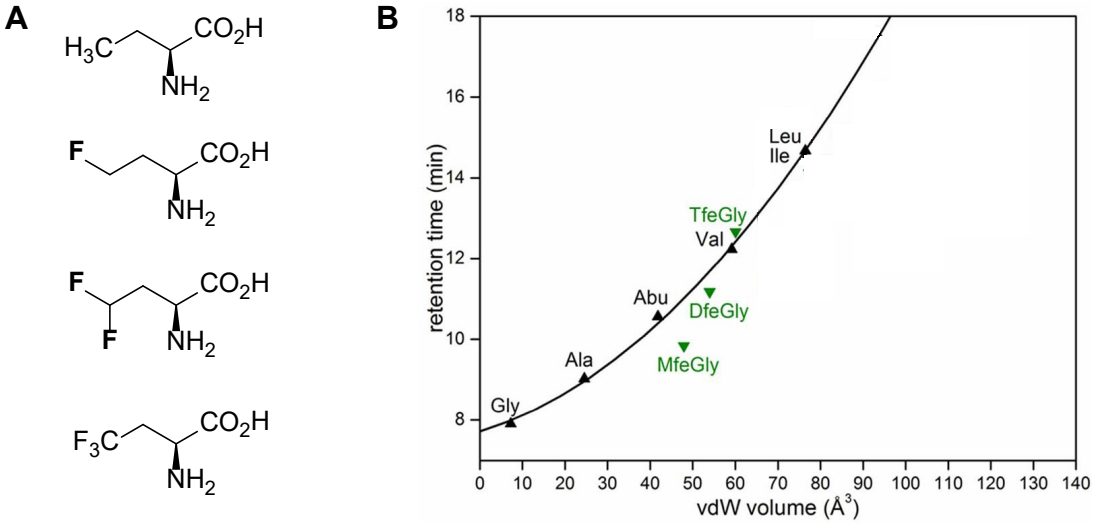


Figure 2.3. Fluorinated derivatives of (2*S*)-2-aminobutanoic acid Abu (**A**); hydrophobicity plot of fluorinated Abu variants and canonical aliphatic amino acids based on RP-HPLC assay (**B**, adapted from Koksch et al. [75]).

In addition to a liquid chromatography-based approach of determining the hydrophobicity of fluorinated amino acids, Pagel *et al.* presented a method based on ion mobility mass spectrometry, in which ions are separated according to their mass-to-charge ratio (*m*/*z*) as well as their size and shape.[76] Furthermore, Vila Verde *et al.* presented theoretical data regarding hydrophobicities based on molecular dynamics simulations combined with the atomistic force field.[77, 78]

2.4.3 Modulating properties of peptides and proteins through incorporation of fluorinated amino acids

Fine-tuning the biological, chemical, and physical properties of peptides and proteins by fluorination has been of great interest in developing pharmaceuticals[79] and bio-inspired materials[80] in the past decades. It was shown that the introduction of fluorine-containing functional groups can enhance metabolic or thermal stability or lead to higher bioavailability or bioactivity of these biomolecules.[5] Fluorine can be incorporated into proteins by various synthetic and biological methods.[9] In a synthetic chemical approach, non-canonical amino acids are mainly introduced into peptides and proteins by solid-phase peptide synthesis (SPPS). This is a technique in which the biomolecule is synthesized from the *C*- to *N*-terminus in a step-growth procedure using orthogonal protection groups to prevent side reactions.[81] Ever since Merrifield first described this method in 1963,[81], an entire field of research developed around it, and it has been optimized and automated.[82] Even though the synthesis of homogeneous peptides through SPPS has had a limitation in chain length of approximately 50 amino acids, novel ligation techniques like native chemical ligation (NCL),[83] microwave heating while coupling the amino acids,[84] and flow-based SPPS [85] have helped expand its utility to longer chain lengths. For example, Brik and coworkers recently described the synthesis of a 53 kDa tetra-Ub-α-globin protein composed of 472 amino acids by combining these techniques with other chemical protein synthesis approaches.[86]

When it comes to fluorine in the context of peptides and proteins, Koksch and coworkers have described various systems in which this fascinating element has been incorporated.[75] Fluorinated analogues of Abu were systematically investigated in peptide model systems, mimicking natural protein environments. As one example, depending on the degree of fluorination in the hydrophobic core of α-helical coiled-coil peptides, an overall increase in thermal stability could be observed by Buer *et al.* when a global substitution with fluorinated amino acids occurred.[87] In contrast, Koksch and coworkers

described that single substitution could lead to destabilization. Depending on how efficiently the fluorinated amino acids packed against adjacent side chains, the thermal stability varied measurably[88-91]

Gottler *et al.* investigated the impact of fluorination on the antimicrobial peptide (AMP) MSI-78.[92] While showing an undefined secondary structure in aqueous solution, MSI-78 assembles into a dimeric antiparallel coiled-coil in the presence of lipid bilayers. The substitution of Leu and Ile with hexafluoroleucine (HfLeu) led to an increase in proteolytic resistance of the AMP, while antimicrobial activity was retained. They assumed that the higher proteolytic stability could be attributed to the enhanced hydrophobic interactions in the dimer induced by fluorine. The substitution of two Val with Leu and HfLeu in β-hairpin AMP protegrin-1 showed increased antimicrobial activity in the case of Leu substitution. In contrast, a loss of activity was observed in the case of HfLeu substitution.[93] This concludes that fluorination does not always have to be beneficial; each model has to be investigated individually since precise impacts are hard to generalize.

Besides small model peptides, several fluorinated amino acids have been incorporated in naturally occurring globular proteins to gather more insights on fluorine's impact on their biological and physicochemical properties. In these studies, the incentive lies in investigating the alterations in bioavailability, conformational changes, thermal and proteolytic stability, biological activity, and protein-protein interactions. The substitution of Leu with diastereomeric 5',5',5'-trifluoroleucine (TfLeu) in chloramphenicol acetyltransferase (CAT) lead to maintenance of enzymatic activity under ambient temperatures as well as an enhancement in secondary structure, but a loss in stability against heat and chemical denaturants.[94] CD studies revealed that fluorination leads to increased secondary structure intensity; however, misfolded structures were observed at elevated temperatures. This led to the conclusion that TfLeu incorporation does not necessarily lead to an overall improvement in stability. Similar observations were made when incorporating TfLeu into green fluorescent protein (GFP).[95] When the two fluorinated derivatives of Tyr, 2-fluorotyrosine (2-fTyr) and 3-fluorotyrosine (3-fTyr), were incorporated into GFP, no significant changes in the overall protein stability could be observed. Yet, the spectral properties of the protein showed substantial alterations due to electronic differences in the chromophore of GFP caused by the substitution of Tyr66 with 2-fTyr and 3-fTyr (Figure 2.4).[96] Similar effects were observed when fluorinated Trp analogues were incorporated into GFP.[97] These studies show that fluorination can alter

the microenvironment of protein chromophores, enabling fine-tuning of the spectral properties of the fluorinated species compared to their wild-type analogues.

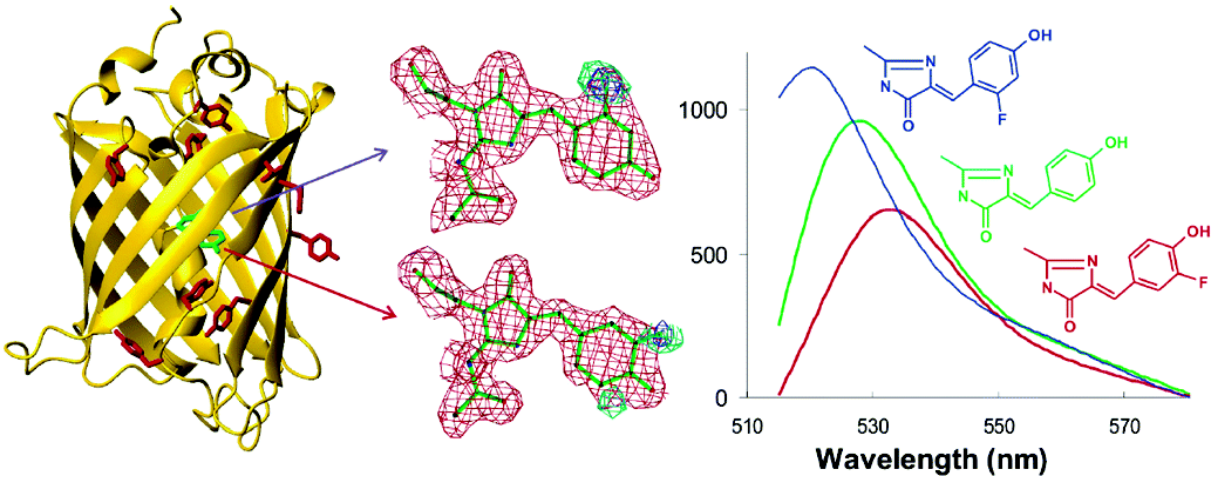


Figure 2.4. Schematic representation of probing the spectroscopic properties of GFP by incorporation of fluorinated analogues of tyrosine; general structure of GFP, which contains 11 Tyr-residues; Try66 is an essential part of chromophore; structure of 2-fTry and 3-fTry (**left**); spectral shifts as a result of fluorination (**right**). (Adapted from Pal *et al.*,[96] Copyright ® 2005, American Chemical Society)

Aside from fluorinated aromatic amino acids, global substitutions of Pro with (4*R*)-fluoroproline[98] and an additional study of Met with trifluoromethionine (TfMet)[99] in DNA polymerase I from *Thermus aquaticus* were investigated. The catalytic activity and stability were preserved in the fluorinated variants in both cases. Since no significant alterations in activity upon fluorination were observed, fluorine could be utilized as a label in ^{19}F NMR spectroscopy to monitor the enzyme's catalytic activity.

The incorporation of fluorinated derivatives of Abu into BPTI by Ye *et al.* showed differences in inhibition properties towards the serine protease β-trypsin.[100] As substitution of Lys15 with Abu showed a significant decrease in inhibition activity compared to wild-type BPTI; the inhibition properties are restored upon the increasing degree of fluorination. These results demonstrate what the authors describe as “chemical complementation”; partially fluorinated amino acids and structural water molecules interact in the binding pocket of β-trypsin due to fluorine’s influence on the otherwise hydrophobic aliphatic side chain.

Such a role for the bioorthogonal element fluorine was not described before, and the observations lend support not only to the view that fluorinated amino acids constitute a truly unique family of building blocks for protein engineering but provide important information that will find applications in peptide and protein engineering, peptide drug design, organocatalysis, and biomaterial development.[100] Despite the studies described above, it remains difficult to precisely predict the impact of fluorine on the overall

properties of proteins.[101] To further optimize protein engineering by fluorination, additional systematic studies are needed to elucidate further the contributions of non-covalent interactions on specific properties of proteins. To date, most of the studies described in the literature focus on the influence of fluorinated aliphatic and aromatic amino acids on proteins. This spectrum should be widened to charged and polar amino acids as they are essential factors in protein properties such as in stability or biological activity. For instance, charged amino acids are present in many active centers of enzymes, and their fluorination may lead to fascinating alterations in the activity of these proteins.

3 Bovine Pancreatic Trypsin inhibitor (BPTI)

3.1 Highly investigated model system

The naturally occurring BPTI, also known as aprotinin, is a protein belonging to the class of Kunitz-type serine protease inhibitors.[102, 103] It is one of the most in-depth studied globular proteins.[104] First discovered by Kraut *et al.* in 1930 in bovine lymph nodes,[105] it was identified as a trypsin inhibitor by Kunitz and Northrop in 1936.[106] This protein inhibitor has been of particular interest for structural biologists, as it was one of the first proteins for which a crystal structure was solved.[107] In addition, it was one of the first biomacromolecules to which molecular dynamics simulations were applied,[108] and one of the first proteins to have its structure confirmed by NMR spectroscopy.[109] BPTI has been accessed by multiple approaches over the past decades.[110] In literature, the synthesis of BPTI has been described by using SPPS and NCL. [111, 112] Furthermore, the inhibitor was expressed from organisms like *saccharomyces cerevisiae*.[113] The exact amino acid sequence composition, called the primary structure (Figure 3.1), was first resolved by Kassell and Laskowski in 1965.[114] The protein consists of a linear chain comprised of 58 amino acids, the primary sequence is depicted in Figure 3.2.

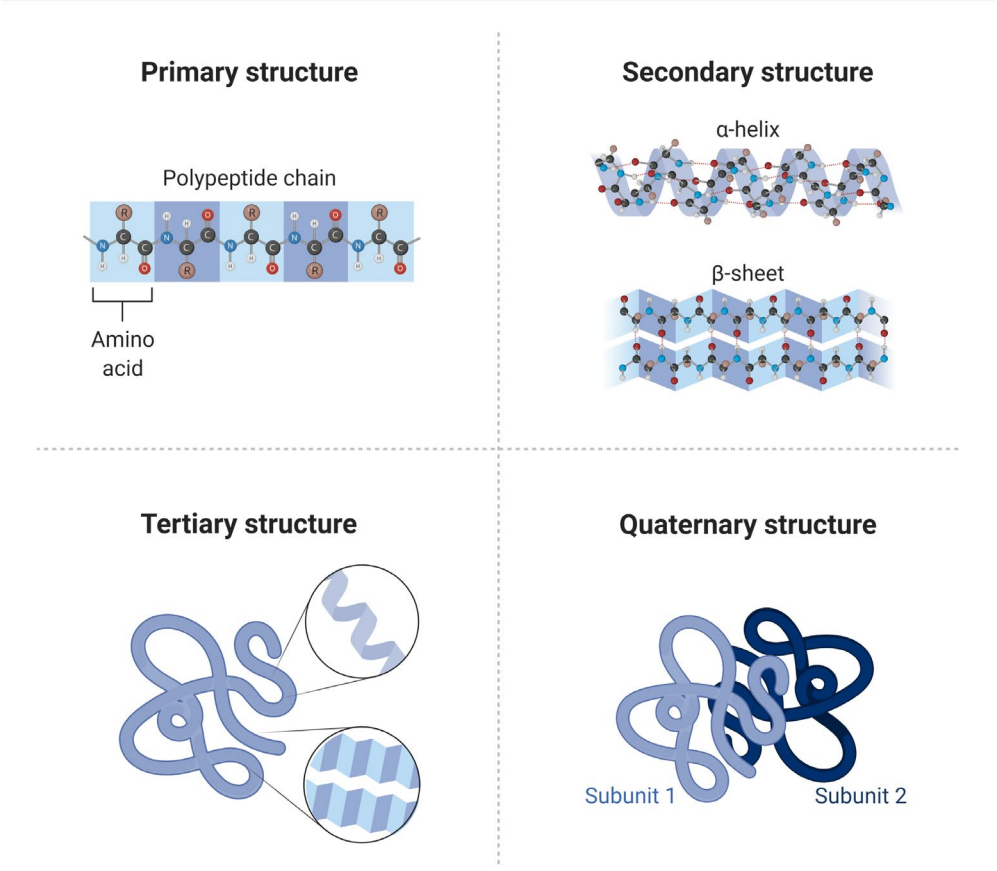


Figure 3.1. Structural hierarchy in proteins. Created with BioRender.com

Structural similarities in the primary structure can be found throughout numerous protease inhibitors of living organisms (Figure 3.2). Lys/Arg15 and Cys residues 5, 14, 30, 38, 51, and 55 are highly conserved throughout all species, as they are essential residues in the activity and overall stabilization of the protein. Most of the variant amino acid residues can be explained by single-base mutations in the third position of the respective codons.[104]

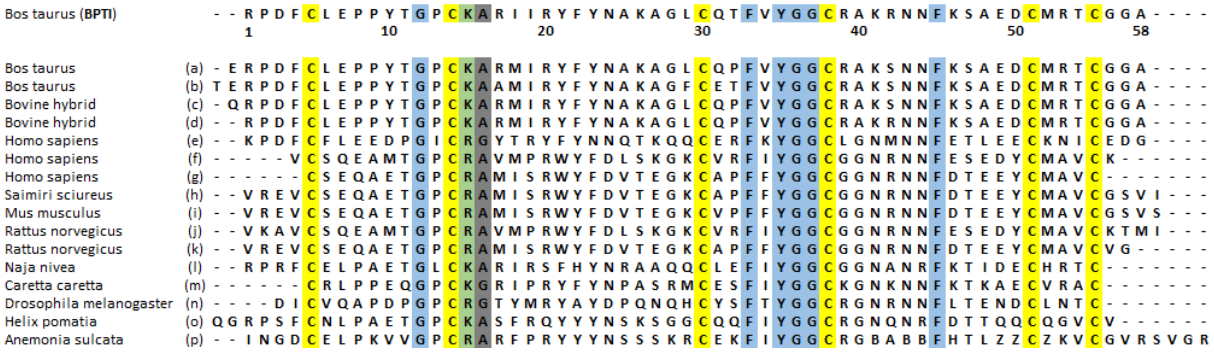


Figure 3.2. Sequence alignment of BPTI and BPTI-like proteins (a) to (p). P_1 and P_1' residues forming the potentially scissile reactive site bond are highlighted green and grey, respectively. Amino acid numbering of BPTI is shown. Residues conserved in all proteins are highlighted in blue. Cysteine residues are highlighted in yellow; Aprotinin analog (a),[115] serum basic protease inhibitor (b),[116] isoaprotinin G1 (c),[115] isoaprotinin G2 (d),[115] second Kunitz domain of tissue factor pathway inhibitor-β (e),[117] amyloid-β (A4) precursor-like protein 2 (f),[118] Amyloid-β peptide (g),[119] Alzheimer's disease amyloid A4 (h),[120] Alzheimer's disease amyloid A4 amyloidogenic glycoprotein (i),[121] amyloid-like protein precursor (j),[122] Alzheimer's disease amyloid A4 protein homolog precursor (k),[123] venom basic protease inhibitor II (l),[124] chelonianim (m),[125] fat-spondin-P1 (n),[126] isoinhibitor K (o),[127] snake-locks sea anemone (p).[128]

About secondary structure motives (Figure 3.1), BPTI consists of a twisted anti-parallel β-sheet (Ile18 to Asn24 and Leu29 to Tyr35) and two helical regions, an α- (Ser47 to Gly56) and a 3_{10}-helix (Pro2 to Glu7). Two β-turns can be found from Ala25 to Gly28 and Lys41 to Asn44. The overall pear-shaped conformation of the protein is stabilized by three disulfide bridges, linking Cys5-Cys55, Cys14-Cys38, and Cys30-Cys51 ().

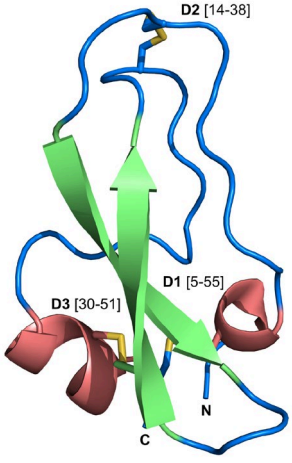


Figure 3.3. The tertiary structure of BPTI, protein chain depicted in blue, helices in red, β-strands in green. Disulfide bridges are shown as D1-D3 in yellow (PDB code: 4Y0Y).[129]

The interacting loop of the protein is centered at the residue Lys15. This site can bind to serine proteases and other proteolytic enzymes. In the past, crystal structures of BPTI were analyzed in complexes with numerous serine proteases,[130] like β-trypsin,[131] α-chymotrypsin,[132] anhydrotrypsin,[133] trypsinogen,[134] pancreatic kallikrein-A,[135] to give a few examples. X-ray crystallography and neutron diffraction have studied multiple crystal forms in the following years. [102, 130, 136-140] The protease-BPTI complexes form a mushroom shape, in which the inhibitor is the stipe and the protease the cap.

The BPTI reactive site from Lys15(P_1)-Ala16(P_1') is located in the serine (pro)enzyme active site that the carbonyl carbon atom of the scissile peptide bond is proximal to the side-chain oxygen of Ser195 of the (pro)enzyme. The carbonyl oxygen atom of Lys15 is stabilized by two hydrogen bonds with the amide nitrogen atoms of the (pro)enzyme's Gly193 (side chain) and Ser195 (main chain). The positively charged BPTI-Lys15(P_1) side-chain forms water-mediated hydrogen-bonding interactions with the negatively charged Asp189 and other residues of trypsin-like serine (pro)enzymes. In case of an interaction with bovine α-chymotrypsin, the S_1 site is not charged, and therefore, Lys15(P_1) instead binds via hydrogen bonding to the carbonyl oxygen atoms of Gly216 and Ser217.[141]

3.2 Refolding of cysteine-rich proteins BPTI and hirudin

In the past decades, the pathways by which small cysteine-rich proteins fold to their native conformations have been extensively studied.[142-149] Two prominent proteins with very different folding features are BPTI and hirudin (Figure 3.4).[150, 151] BPTI folding is defined by predominantly occurring 1- and 2-disulfide intermediates that adopt native cysteine pairings and substructures. [152, 153] These intermediates funnel protein conformations towards the native state and prevent the accumulation of three-disulfide scrambled isomers. [154-156] In BPTI, Cys14-Cys38 is the single solvent-exposed disulfide bridge and is the first disulfide formed in the oxidative folding mechanism of the protein (Figure 3.5). [149, 157] It rearranges to rapidly into the other disulfide bridges Cys30-Cys51 and Cys5-Cys55.[146] According to the folding pathway described by Weissman and Kim iprprimary1,[158] two major intermediates are formed during the folding process, each containing two disulfide bonds, named N' (14-38, 30-51) and N* (5-55, 14-38).[146, 158-160]

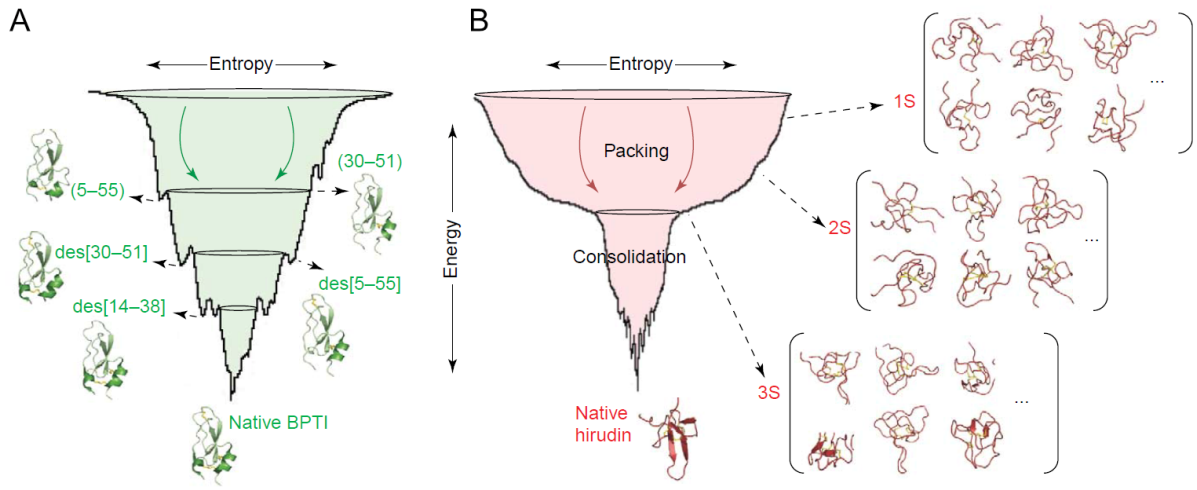


Figure 3.4. Schematic overview of divergent landscapes in oxidative folding. The energy landscape is represented in a two-dimensional manner, with reduced unfolded proteins at the highest energetic level on top and native folded structures at the bottom of the funnel. For BPTI, a limited number of native-like intermediates funnel protein conformations towards the native structure; disulfide pairings of folding intermediates are shown between parentheses (**A**). Hirudin folds through an initial stage of disulfide-bond formation (packing), followed by reshuffling of scrambled isomers (consolidation) to form the native protein structure; 1S, 2S, and 3S indicate the heterogeneous ensembles containing one, two, or three disulfides (**B**); (Adapted with permission from Arolas *et al.*,[150] Copyright 2006, Elsevier).

N' is proposed to rearrange to N^{SH}_{SH} (5-55, 14-38) through unfolding. Intramolecular thiol-disulfide exchange reactions, followed by the formation of native form N. For over 30 years, it was believed N* was a dead-end intermediate.[160, 161] However, Mousa *et al.* showed in 2018 that N* is capable of rearranging to N through an intermediate formed by N* and oxidizing reagent glutathione (GSSG), named N*-SG.[149, 162, 163]

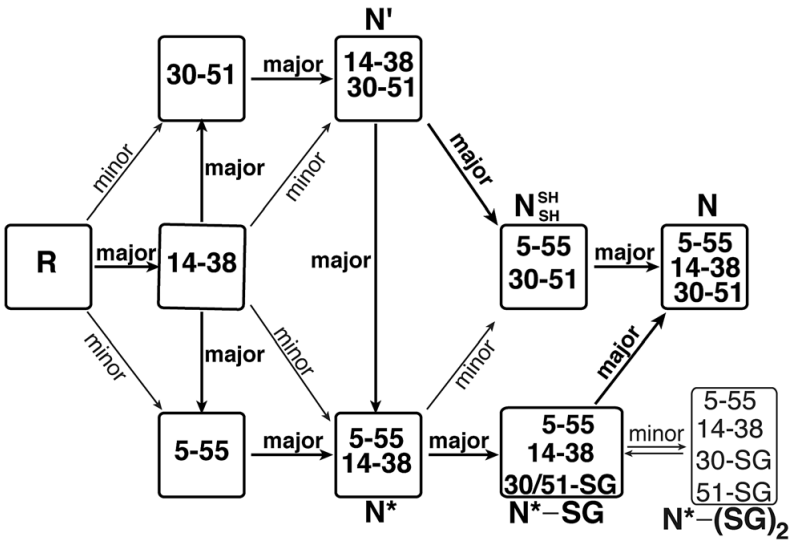


Figure 3.5. Oxidative folding pathway of BPTI, "major" indicates a preferred disulfide formation, whereas "minor" means a less preferred. (Adapted with permission from Mousa *et al.*,[149] Copyright 2018, Royal Society of Chemistry).

The oxidative folding of hirudin differs from that of BPTI drastically. Firstly, the folding intermediates are far more heterogeneous, with over 30 species of one- and two-

disulfide-containing intermediates identified.[164, 165] In addition, no predominant intermediates are formed adopting native disulfides, and high levels of three-disulfide scrambled isomers are formed.[166] The folding mechanism of hirudin can instead be described in a two-step way: At first, in an initial stage, non-specific disulfide bonds are included with many scrambled intermediates (packing), after which reduction and disulfide-reshuffling leads to the native structure (consolidation).[166]

One of the most fundamental questions in biology is the protein folding problem. How does the primary sequence of the protein determine its three-dimensional structure? Exceptional progress has been made in unraveling the causes of why polypeptides fold in their native state conformations.[167, 168] Some of the key findings are that the unfolded state of a protein is defined by native-like intramolecular interactions in which only a limited amount of conformations are searched by the amino acid chain, called the topomer-sampling model.[169] The so-called "hydrophobic collapse" is a hypothesized event that occurs early during the folding process of globular proteins. It is based on the observation that in the folded native conformation of many proteins, a hydrophobic core consisting of non-polar amino acids is present, with the most polar and charged residues facing the solvent-exposed protein surface. According to this theory, the protein chain forms initially localized regions consisting of mainly hydrophobic residues. Due to the interaction of the polypeptide chain with water, it aggregates into a tertiary confirmation with a hydrophobic core. This state is called molten globule and corresponds to a partially folded state in which the energy is lower than in the denatured but higher than in the native state.[170] Because of the time scale, experimental evidence of early folding events are challenging to obtain, which is why molecular dynamics simulations mainly study this stage.[171]

Nowadays, new algorithms can predict the correct native state of an amino acid sequence. This is mainly due to increased calculation capacities and novel technologies like machine learning and artificial intelligence (AI). Today's most prominent example is probably *AlphaFold*,[172] a deep learning program that reaches the highest scores in the protein-structure prediction challenge called *Critical Assessment of Structure Prediction* (CASP).[173] As seen in Figure 3.6, the *AlphaFold*-predicted structure [174, 175] of BPTI is the same as the structure obtained from X-ray analysis.

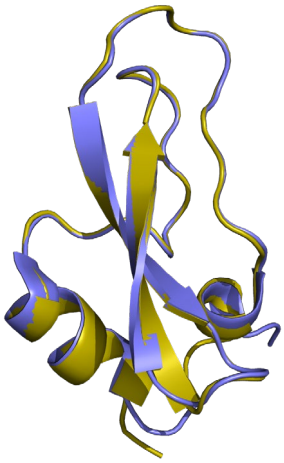


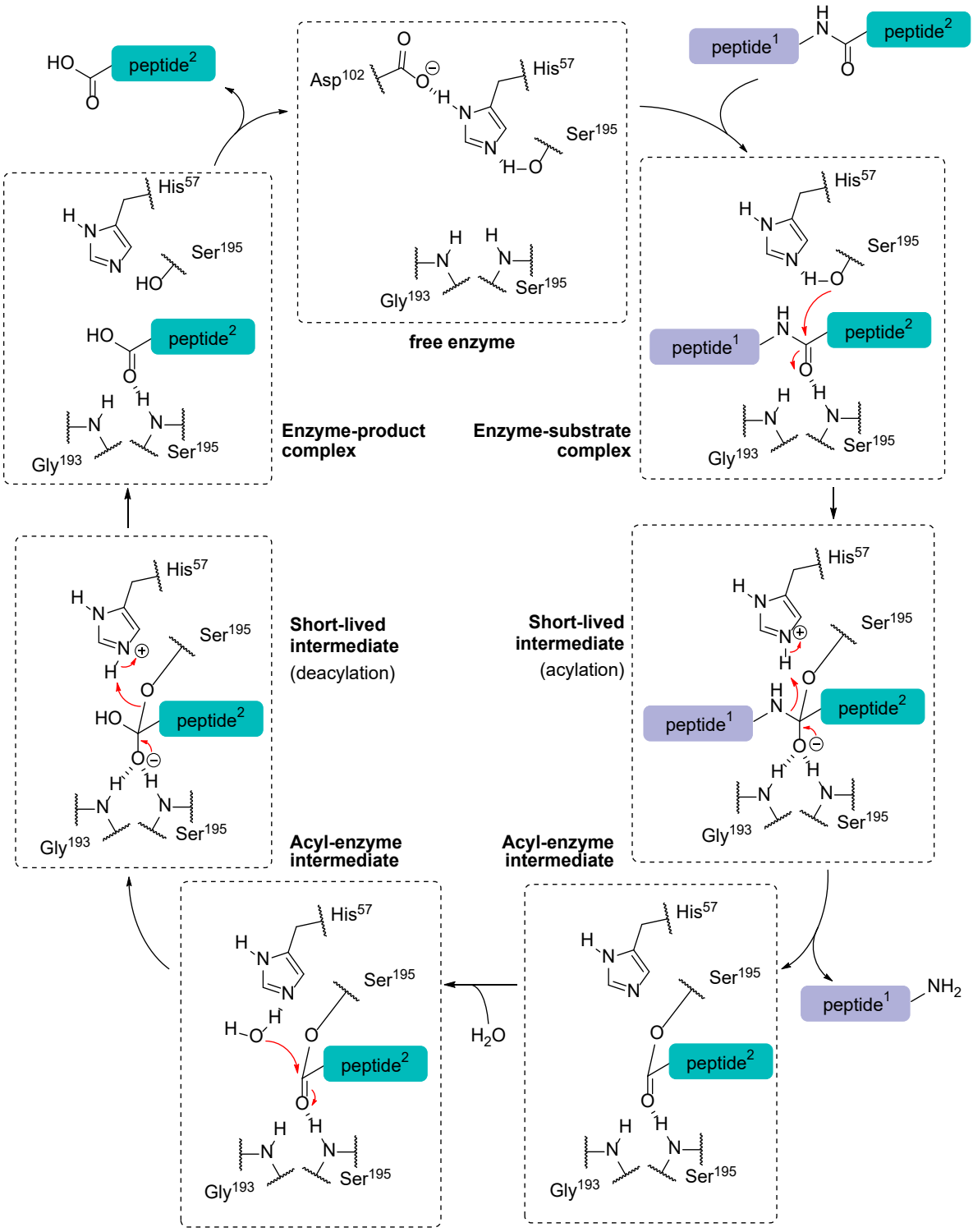
Figure 3.6. Predicted structure of BPTI by *AlphaFold* (olive) [174, 175] aligned to X-ray structure (purple, PDB code: 4Y0Y).[100]

With today's rapid development in computing techniques, in the near future, even the most complex proteins structures might be predictable with low knowledge of molecular dynamic simulations and little calculation cost, thus opening a deeper insight into the complexity of protein folding pathways.

3.3 Proteases – classification, catalytic activity, and inhibition

Proteases are enzymes that catalyze proteolysis. They can be classified into seven groups: serine-, cysteine-, threonine-, aspartic-, glutamic-, metalloproteases, and asparagine peptide lyases.[176] Almost one-third of all proteases are serine proteases.[177] Among the class of serine proteases, there are 16 superfamilies, each containing several families.[178] chymotrypsin-like proteases from the PA clan (superfamily) are the most abundant in nature,[176] involved in a broad range of physiological processes. These diverse functions result in proteases with widely varied specificities, from digestive enzymes, which cleave after small hydrophobic, bulky aromatic, or positively charged residues.[179] The specificity of serine proteases can usually be explained by the substrate-binding site's topology in the enzyme's active center. Adjacent to the binding site, the Asp-His-Ser charge relay system in the catalytic active center of these enzymes, also called the "catalytic triad", can be found. The catalytic triad consists of the residues Asp102, His57, and Ser195 (Scheme 3.1). It is part of an extensive hydrogen-bonding network. Upon substrate binding, the enzyme-substrate complex is formed, and His57 and Ser195 generate an alkoxide ion on Ser195, which attacks the carbonyl carbon of the amide bond, resulting in a short-lived tetrahedral intermediate. Due to the instability of the negative charge at the carbonyl oxygen of the substrate, the intermediate collapses, releasing the first product under the

formation of an acyl-enzyme intermediate. An incoming water molecule is deprotonated by general base catalysis under the formation of a hydroxide ion, which attacks the carbonyl carbon of the acyl-enzyme ester-forming a second short-lived tetrahedral intermediate. Its collapse forms the second product, which upon diffusion from the active site regenerates the free enzyme.



Scheme 3.1. Catalytic mechanism of peptide bond hydrolysis by chymotrypsin. Backbone NH and side chain OH of Ser195 are not connected for guaranteeing clarity in this scheme.

The backbone NHs of Gly193 and Ser195 form the oxyanion hole, a positively charged pocket, through which the carbonyl of the scissile peptide bond is activated, and the negatively charged oxyanion of the tetrahedral intermediate is stabilized.[180]

As digestive enzymes break down peptides and proteins into smaller building blocks, organisms have evolved inhibitors, which downregulate the activity of these enzymes. Generally, enzyme inhibition can be irreversible or reversible. Among reversible inhibitors, there are different modes by which enzymes can inhibit through competitive, uncompetitive, or mixed (non-competitive) inhibition (Figure 3.7). In competitive inhibition, the substrate and the inhibitor compete for the active site of an enzyme. While the inhibitor binds to the enzyme, the active site is blocked, preventing hydrolysis of the substrate. In many cases, competitive inhibitors mimic the structural motives of the substrate to enable binding to the enzyme. Thus, leading to the formation of the enzyme-inhibitor complex disrupting the catalytic properties of the enzyme. In case of an uncompetitive inhibition, the inhibitor is binding to a site distinct from the active site, the allosteric site, which is formed only upon the formation of the enzyme-substrate complex. A mixed inhibitor can bind to an enzyme or an enzyme-substrate complex since an allosteric site is already present in both cases.[181]

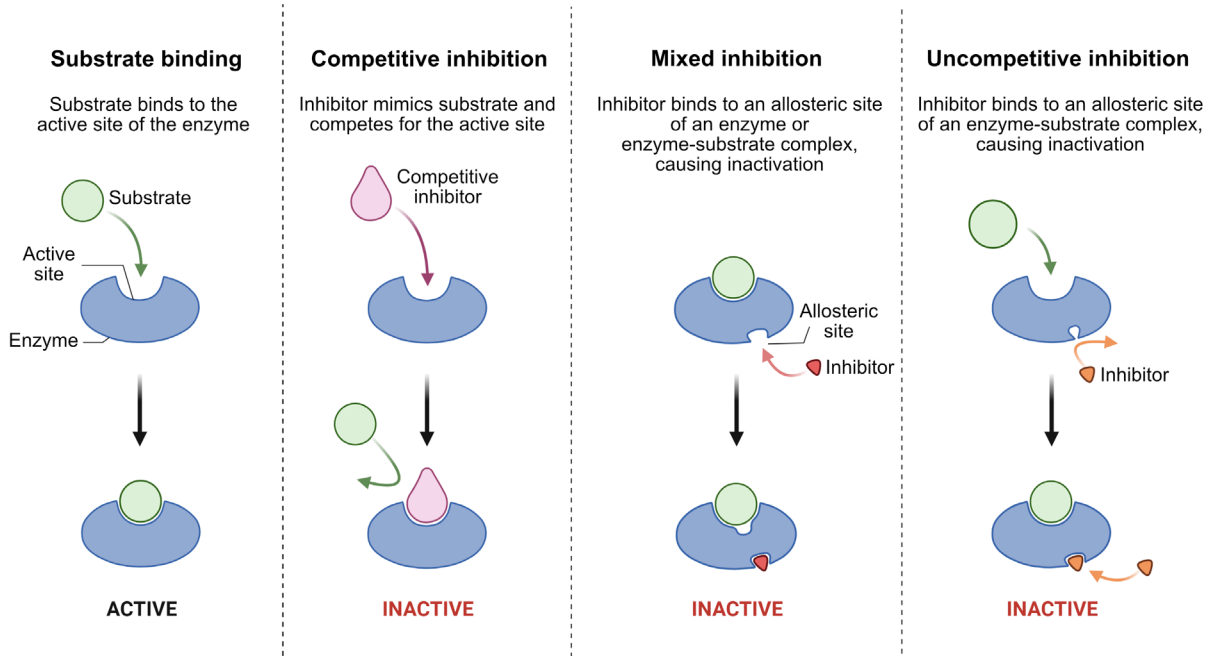


Figure 3.7. Schematic representation of different modes of reversible enzyme inhibition. Created with BioRender.com

BPTI follows a classical competitive inhibition mechanism. The minimal notation for the reaction scheme for the formation of a serine (pro)enzyme-BPTI complex is:

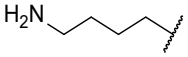
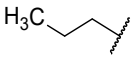
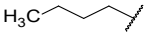
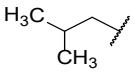
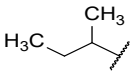
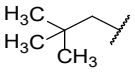
$$E + BPTI \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} L \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} E{:}BPTI \tag{3.1}$$

In which E is the serine (pro)enzyme, BPTI is the native inhibitor, L stands for a loose complex of E with BPTI, respectively, E:BPTI is the serine (pro)enzyme-BPTI complex. k_1 denotes the second-order rate constant for the formation of L, k_{-1} is the first-order rate constant for the dissociation of L, k_2 stands for the first-order rate constant for the conversion of L to E:BPTI and is the rate-limiting step for the E:BPTI complex formation, k_{-2} is the first-order rate constant for the conversion of E:BPTI to L and is the rate-limiting step for the E:BPTI complex dissociation.

$$K = \frac{k_{-2}}{k_2 / \frac{k_{-1}}{k_1}} \tag{3.2}$$

Serine (pro)enzyme-BPTI complexes were investigated intensively in the past to gain more profound knowledge of protein-protein interactions. BPTI has relatively broad specificity. It inhibits chymotrypsin-, elastase- and trypsin-like (pro)enzymes (Table 3.2). The affinity of BPTI for the different serine proteases is mainly determined by the contacts between the enzyme (S_1) and the inhibitor (P_1). Replacing the Lys15 of BPTI with an apolar side chain, such as *nor*Val, *nor*Leu, or Phe, forces the inhibitor P_1 hydrophobic side chain in the protease S_1 specificity pocket of trypsin, which contains Asp189, causing six to nine orders of magnitude drop in inhibitor affinity.[182] On the other hand, the affinity of wild-type BPTI for α-chymotrypsin is quite close to that of the semisynthetic BPTI analogues bearing the apolar P1 residues *nor*Val, *nor*Leu, or Phe (Table 3.1).[183] As mentioned before, such an observation can be interpreted by assuming that the cost of restructuring the P_1 side chain outside of the α-chymotrypsin S_1 subsite is essentially compensated by hydrogen bonding to the carbonyl oxygen atoms of Gly216 and Ser217.[141]

Table 3.1: Dissociation constants of BPTI or P_1-variants in complex with two serine proteases. Weak inhibition is denoted by "+", no inhibition by "-".[184]

P_1 residue	**P_1 side chain**	**vdW volume** [Å^3]	***K*** **(β-trypsin)** [mol · L^{-1}]	***K*** **(α-chymotrypsin)** [mol · L^{-1}]
Lysine (wild type)		87.44	$6 \cdot 10^{-14}$	$9 \cdot 10^{-9}$
Glycine		7.24	-	-
Alanine		24.54	+	-
2-Aminobutanoic acid		41.84	+	+
Norvaline		59.14	$1.3 \cdot 10^{-7}$	$1.4 \cdot 10^{-8}$
Valine		59.14	-	-
Norleucine		76.44	$2.0 \cdot 10^{-8}$	$6.7 \cdot 10^{-9}$
Leucine		76.44	-	-
Isoleucine		76.44	-	-
tert-Leucine		76.44	-	-
tert-Butylalanine		93.74	+	+

A semisynthetic approach to modulating the affinity of BPTI for human leukocyte elastase demonstrates the individual contribution of the P_1 side chain to the inhibition of a particular target protease. The mutant Val15 species shows the most powerful inhibition of human leukocyte elastase.[184]

Tryptases are trypsin-like serine proteases found in granules of mast cells where BPTI is also localized.[185] Human tryptase is a ring-like tetramer formed by four identical subunits with the active site facing the center of the ring.[186] Recombinant BPTI variants inhibit tryptase TL2, a serine esterase from human T4+ lymphocytes, and HIV-1 replication in H9 cells. Interestingly TL2 is a binding component to HIV-1 glycoprotein gp120, playing a role in the HIV-1 infection mechanism. In contrast to wild type BPTI, the semisynthetic BPTI analogue with Arg15 at P_1, with a reactive-site sequence homologous to the V3 loop of HIV-1 gp120, inhibits specifically human tryptase TL2 activity.[187]

Table 3.2: Values of the apparent thermodynamic and kinetic parameters for binding wild-type BPTI to representative serine proteases.

Protease	***K*** [M]	$\mathbf{k_1}$ [$M^{-1}s^{-1}$]	$\mathbf{k_{-2}}$ [s^{-1}]	**Reference**
H. trypsin-1	$< 2.0 \cdot 10^{-10}$	$1.0 \cdot 10^{5}$	$< 2.0 \cdot 10^{-5}$	188
H. trypsin-2	$< 1.0 \cdot 10^{-9}$	$2.4 \cdot 10^{5}$	$< 2.0 \cdot 10^{-4}$	188
B. β-trypsin	$6.0 \cdot 10^{-14}$	$1.1 \cdot 10^{6}$	$6.6 \cdot 10^{-8}$	131
B. α-chymotrypsin	$1.5 \cdot 10^{-9}$	$3.7 \cdot 10^{5}$	$5.5 \cdot 10^{-4}$	132, 183, 184, 189
H. chymotrypsin-A	$1.7 \cdot 10^{-7}$	$\sim 10^{4}$	$2.0 \cdot 10^{-3}$	188
B. pseudotrypsin	$9.0 \cdot 10^{-9}$	$7.0 \cdot 10^{4}$	$6.3 \cdot 10^{-4}$	131
B. anhydrotrypsin	$1.1 \cdot 10^{-13}$	$7.7 \cdot 10^{5}$	$8.5 \cdot 10^{-8}$	133
B. trypsinogen	$2.0 \cdot 10^{-6}$	$2.0 \cdot 10^{4}$	$4.0 \cdot 10^{-2}$	134, 135, 190-192
B. trypsinogen: Ile-Val	$< 1.0 \cdot 10^{-9}$	$2.0 \cdot 10^{5}$	$< 2.0 \cdot 10^{-4}$	135, 190, 191
H. u. kallikrein	$9.1 \cdot 10^{-11}$	-	-	135, 193
P. p. β-kallikrein-A	$8.4 \cdot 10^{-10}$	$3.0 \cdot 10^{5}$	$2.7 \cdot 10^{-4}$	135, 193
P. p. β-kallikrein-B	$8.4 \cdot 10^{-10}$	$3.0 \cdot 10^{5}$	$2.7 \cdot 10^{-4}$	135, 193
H. α-thrombin	$8.0 \cdot 10^{-4}$	$4.0 \cdot 10^{4}$	$7.0 \cdot 10^{1}$	194, 195
H. β-thrombin	$4.0 \cdot 10^{-4}$	$5.0 \cdot 10^{4}$	$3.0 \cdot 10^{1}$	194, 195
H. γ-thrombin	$1.1 \cdot 10^{-4}$	$8.5 \cdot 10^{4}$	$1.0 \cdot 10^{1}$	194, 195
H. Glu192Gln-thrombin	$2.4 \cdot 10^{-6}$	-	-	196
B. factor Xa	$4.8 \cdot 10^{-6}$	-	-	197
H. factor Xa	$4.8 \cdot 10^{-6}$	-	-	197
H. factor XII	$9.1 \cdot 10^{-5}$	-	-	198
H. protein C	$1.6 \cdot 10^{-7}$	-	-	198
H. α-plasmin	$1.9 \cdot 10^{-10}$	-	-	198
H. plasmin-Glu1	$1.0 \cdot 10^{-9}$	$1.6 \cdot 10^{5}$	$1.6 \cdot 10^{-4}$	199
H. plasmin-Lys77	$8.3 \cdot 10^{-10}$	$1.8 \cdot 10^{5}$	$1.5 \cdot 10^{-4}$	199, 200
H. plasmin-Val442	$7.7 \cdot 10^{-10}$	$2.0 \cdot 10^{5}$	$1.5 \cdot 10^{-4}$	199
H. cathepsin G	$4.0 \cdot 10^{-6}$	$2.3 \cdot 10^{5}$	$9.0 \cdot 10^{-1}$	201
H. l. elastase	$3.5 \cdot 10^{-6}$	-	-	184
H. p. elastase	$> 2.0 \cdot 10^{-3}$	-	-	188
P. p. elastase	$1.0 \cdot 10^{-3}$	-	-	202

The inability to produce adequate pancreatic trypsin inhibitor can lead to severe tissue damage in the pancreas. This can lead to inflammation of the pancreas, a medical condition called acute pancreatitis.

3.4 Biomedical Applications

BPTI inhibits several serine proteases, and thus over the past few decades, its pharmacological applications were focused on clinical syndromes associated with dysfunctions of digestive and coagulation enzymes.

BPTI was used as an intensive-care drug for acute pancreatitis.[203] As mentioned previously, it forms strong complexes with human pancreatic trypsin-1 and -2.[204] Yet, BPTI shows no significant inhibition of human pancreatic chymotrypsin-A and human pancreatic elastase.[205] For this reason, its therapeutic efficacy in digestive dysfunctions has been questioned in the past.[206] Randomized trials treating human acute pancreatitis with antiproteolytic drugs, such as BPTI and gabexate mesylate, have failed to show a benefit in this clinical setting. Hence, the pathway of severe acute pancreatitis is not significantly influenced by BPTI.[207]

Another field of clinical application of BPTI is the treatment of dysfunctions of coagulation/fibrinolysis enzymes. It shows promising efficacy in clinical trials. In the intrinsic pathway of coagulation, the activation of the coagulation cascade can occur through activation of some serine proteases, such as coagulation factor XII, also called the Hageman factor. [208] XII triggers the contact system, which binding factor XII initiates to a negatively charged surface. This interaction leads to autoactivation of XII, converting the zymogen form into the active enzyme factor XIIa.[209] XIIa converts pre-kallikrein to kallikrein, which activates XII, cleaves high molecular weight kininogen, generating vasoactive bradykinin, and converts pro-urokinase to urokinase. Fibrinolysis consists of plasminogen activation to the enzyme plasmin, which digests fibrin fibrils. Plasminogen is activated by urokinase and by tissue plasminogen activator.[208] It is known that BPTI inhibits kallikrein and, therefore, indirectly inhibits the activation of XII to XIIa.[209] Thus, BPTI inhibits the initiation of both coagulation and fibrinolysis.

Furthermore, BPTI inhibits nitric oxide synthesis catalyzed by rat nitric oxide synthase type I and II (NOS-I and NOS-II); however, the binding mechanism between NOS-I/II and BPTI remains unknown.[210]

Large conductance Ca^{2+}-activated K^+ channels contain intracellular binding sites for BPTI. The binding of BPTI to one inhibitor-binding domain (out of four available) induces discrete sub-state events, which results in the formation of a Ca^{2+}-activated K^+ channel that does not allow normal movement of the voltage sensor in that subunit. Thus, channel opening is impaired when one of the four subunits is in a conformation different from the other.

The sub-state current level is unaffected by most of the mutations of BPTI except for substitutions of Lys15 at the P_1 site. Different residues at this position modulate the conductance of the BPTI-induced sub-state from 0 % with Gly15 to 55 % with Val15 of the open state at +20 mV. This indicates that the BPTI inhibitory loop that contacts serine (pro)enzymes forms the Ca^{2+}-activated K^+ channel-BPTI complex.[211] A recent study has shown that SARS-CoV-2 might be an exciting target for BPTI since it could prevent viral replication.[212]

4 Aim of this study

In this work, the site-specific incorporation of fluorine into a model protein and the resulting changes in the properties of the protein will be investigated. The main interest will focus on the varying properties in protein-protein interactions, as previous studies have shown, that site-specific fluorination can have a beneficial impact on inhibition properties towards serine proteases. Ye *et al.* have shown that the introduction of fluoroalkyl amino acid side chains, namely DfeGly and TfeGly, in the P_1 position of BPTI is capable of restoring inhibitory activity towards β-trypsin compared to a mutant that contains the corresponding hydrocarbon side chain Abu.[100] Crystal structure analysis of the serine protease-inhibitor complexes revealed changes in the stoichiometry and dynamics of water in the specificity pocket of trypsin. Starting from this point, the library of fluorinated BPTI variants will be extended to the still undescribed MfeGly-containing BPTI.

Since access to fluorinated amino acids proves to be challenging and not all synthetic pathways are easily feasible, a novel approach to accessing Fmoc-protected MfeGly will be designed. This approach will have to be superior to existing methods to overcome the availability bottleneck about other upcoming research projects. MfeGly is commercially not available and has not been described in the PDB as ligand so far.

Once the fluorinated derivatives of Abu are synthesized, they will be site-specifically incorporated into position 15 of BPTI. In a oxidative refolding process, the cysteine-rich serine protease inhibitor will be refolded into its native state to be used in further investigations. Correct folding of the protein will be monitored and verified by UV-monitored RP-HPLC and CD spectroscopy. The activity of proteins with the native conformation will be tested by UV-based inhibition assays towards several serine proteases. BPTI variants will be crystallized with chymotrypsin for in-depth structural analyses to gain deeper insight into the structural relationship between protease and inhibitor.

5 Published work

Overview of peer-reviewed publications involved in this thesis:

- J. Leppkes, T. Hohmann, B. Koksch, Improved enantioselective gram scale synthesis route to *N*-Fmoc-protected monofluoroethylglycine. *J. Fluorine Chem.* **2020**, *232*, 109453.
- J. Leppkes, N. Dimos, B. Loll, T. Hohmann, M. Dyrks, A. Wieseke, B. G. Keller, B. Koksch, Fluorine-induced polarity increases inhibitory activity of BPTI towards chymotrypsin, *RSC Chem. Biol.* **2022**, *Accepted Manuscript* (DOI: doi.org/10.1039/D2CB00018K)

5.1 Improved enantioselective gram scale synthesis route to *N*-Fmoc-protected monofluoroethylglycine

J. Leppkes, T. Hohmann, B. Koksch, *J. Fluorine Chem.* **2020**, *232*, 109453.

Submitted: 6 November 2019; First published: 23 January 2020
Published by Elsevier B.V., Amsterdam, Netherlands.[213]
The published work is available online:
https://doi.org/10.1016/j.jfluchem.2020.109453

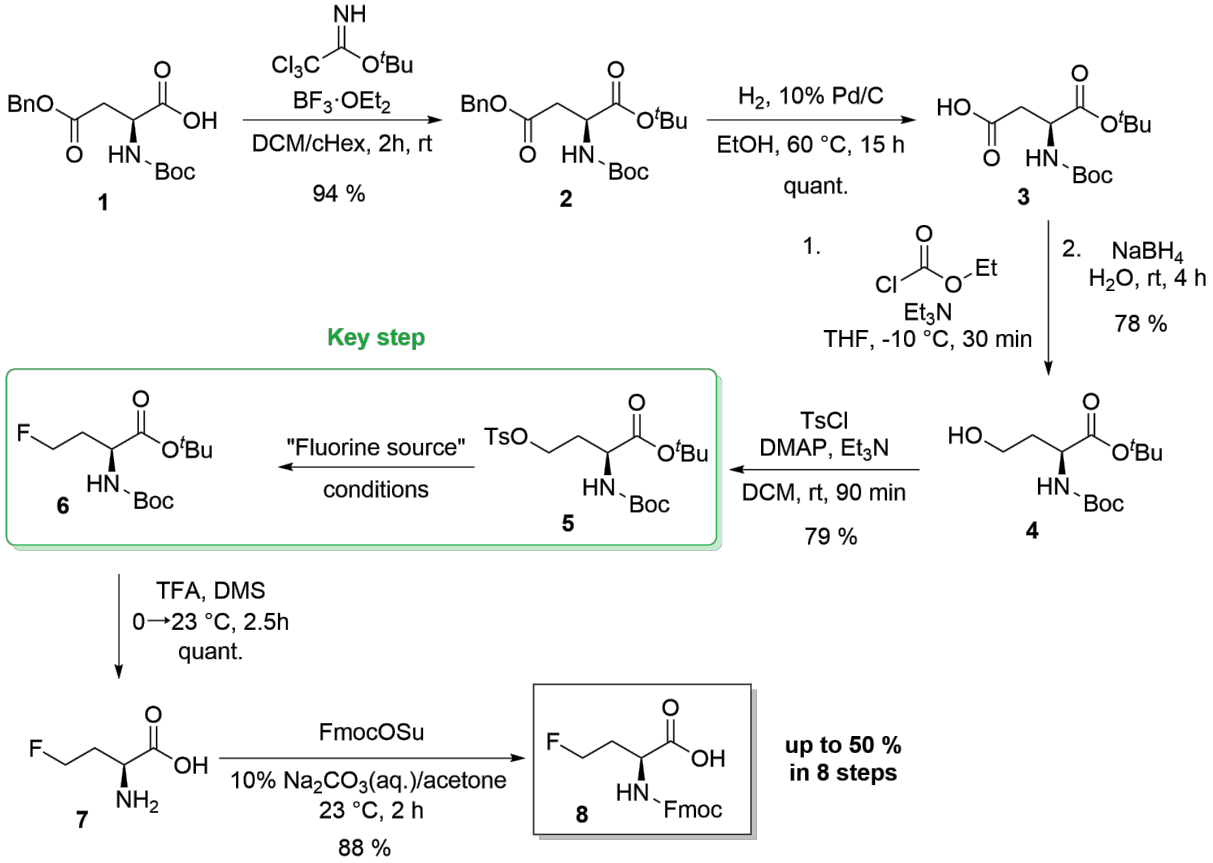
5.1.1 Contributions of Authors

Beate Koksch and I developed the general concept of this project. Beate Koksch supervised the project. I wrote the manuscript, and Beate Koksch revised it. Thomas Hohmann and I performed the synthesis of the studied compounds. I carried out all measurements and analyses.

5.1.2 Rationale and summary of the project

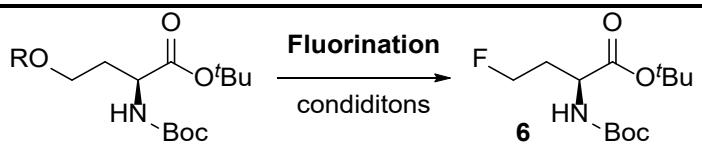
In this project, the novel synthesis pathway of Fmoc-protected monofluoroethylglycine for the incorporation into peptides and proteins by solid-phase peptide synthesis was established. This was of great interest, since MfeGly is commercially not available and its access was further limited due to a demanding synthesis.[214] However, MfeGly and its other fluorinated derivatives, DfeGly and TfeGly, are of great interest and have been extensively studied in the field of peptide and protein engineering.[129, 215-219] Haufe *et al.* first described the synthesis of enantiopure MfeGly in 2000.[214] The new approach aimed

to develop a convenient synthesis of obtaining the Fmoc-protected amino acid in a gram-scale (Scheme 5.1). *N*-α-Boc-protected L-aspartic acid was chosen as staring material, a relatively inexpensive precursor for non-canonical amino acids. In a five-step synthesis, tosylated homoserine (Hse) **5** was obtained. The critical step in this study was to obtain Boc-L-MfeGly-O^{t}Bu **6**; for this reason, multiple nucleophilic fluorination sources were screened (Table 5.1). Using TASF and triethylamine trihydrofluoride led to a successful conversion of Hse **5** to Mfe **6** in an almost quantitive yield. However, due to limited accessibility, high price, and large excess needed of TASF, further fluoride sources were explored. Based on studies by Kim *et al.*,[220, 221] tetra-*N*-butylammonium fluoride (TBAF) trihydrate was tested as a less-expensive nucleophilic fluorination reagent being able to synthesize MfeGly. By employing TBAF(^{t}BuOH)$_4$ as nucleophilic fluorination reagent in acetonitrile, the yield could even be increased to 65 % compared to 56 % using TBAF · 3 $H_2$0. Afterward, enantiopure Fmoc-protected MfeGly **8** was obtained in a two-step synthesis.



Scheme 5.1. Synthesis pathway of Fmoc-protected MfeGly (Reprinted with permission from Leppkes *et al.*,[213] Copyright 2020, Elsevier.)

Table 5.1. Synthetic conditions for nucleophilic fluorination of homoserine.



R = H **4**
R = Ts **5**

Entry	Compound	Fluorine source	Solvent	Temperature [°C]	Time [h]	Yield [%] [a]
1	**4**	DAST	CH_2Cl_2	-10	20	-[b]
2	**4**	XtalFluor-E (3 equiv.), $Et_3N{\cdot}3$ HF (1.5 equiv.)	CH_2Cl_2	23	16	-[b]
3	**5**	TASF (5 equiv.), $Et_3N{\cdot}3$ HF (**0 equiv.**)	CH_2Cl_2/THF	50	16	40
4	**5**	TASF (5 equiv.), $Et_3N{\cdot}3$ HF (**0.5 equiv.**)	CH_2Cl_2/THF	50	16	74
5	**5**	TASF (5 equiv.), $Et_3N{\cdot}3$ HF (**1 equiv.**)	CH_2Cl_2/THF	50	16	48
6	**5**	TASF (5 equiv.), $Et_3N{\cdot}3$ HF (**1.5 equiv.**)	CH_2Cl_2/THF	50	16	98
7	**5**	TASF (5 equiv.), $Et_3N{\cdot}3$ HF (**2 equiv.**)	CH_2Cl_2/THF	50	16	62
8	**5**	TASF (5 equiv.), $Et_3N{\cdot}3$ HF (**2.5 equiv.**)	CH_2Cl_2/THF	50	16	84
9	**5**	TASF (5 equiv.), $Et_3N{\cdot}3$ HF (**3 equiv.**)	CH_2Cl_2/THF	50	16	81
10	**5**	CsF (3 equiv.)	^{t}BuOH	80	6	21
11	**5**	TBAF$\cdot 3\ H_2O$ (3 equiv.)	^{t}BuOH	60	1.5	54
12	**5**	TBAF(^{t}BuOH)$_4$ (2 equiv.)	^{t}BuOH	70	16	45[c]
13	**5**	TBAF(^{t}BuOH)$_4$ (3 equiv.)	MeCN	70	16	38
14	**5**	TBAF(^{t}BuOH)$_4$ (2 equiv.)	MeCN	70	3	65

[a] Isolated yields [b] Starting material recovered [c] Yield determined by ^{19}F NMR using trifluorotoluene as internal standard

This amino acid was particularly important since it was needed for studies described in chapter 5.2 and other projects in the Koksch working group.

For copyright reasons, the article is not included in the online version of this thesis.
An electronic version of the article is available
(DOI: https://doi.org/10.1016/j.jfluchem.2020.109453)

5.2 Fluorine-induced polarity increases inhibitory activity of BPTI towards chymotrypsin

J. Leppkes, N. Dimos, B. Loll, T. Hohmann, M. Dyrks, A. Wieseke, B. G. Keller, B. Koksch, *RSC Chem. Biol.* **2022**, A*ccepted Manuscript* (DOI: doi.org/10.1039/D2CB00018K)
Submitted: 21 January 2022; First published: 16 May 2022
Published by Royal Society of Chemistry, London, UK.[222]
The published work is available online: https://doi.org/10.1039/D2CB00018K

5.2.1 Contributions of Authors

I designed and performed syntheses of BPTI variants, synthesized fluorinated amino acids, performed inhibition and CD measurements, analyzed the data, prepared figures, and wrote the manuscript and the ESI. Nicole Dimos performed crystallization experiments and performed structural refinement. Bernhard Loll performed X-ray diffraction measurements and performed structural refinement. Thomas Hohmann provided fluorinated amino acids, DfpGly and PfpGly, and measured hydrophobicity of Fmoc-protected amino acids. Michael Dyrks and Ariane Wieseke assisted in protein synthesis and refolding. Bettina G. Keller helped writing the manuscript and evaluating inhibition data. Beate Koksch designed and supervised the entire project and the writing of the paper.

5.2.2 Rationale and summary of the project

In this project, the impact of site-specific fluorination on the inhibition properties of BPTI towards the serine protease α-chymotrypsin was investigated. Based on previous work,[129] about fluorine restoring inhibitory properties of BPTI towards β-trypsin, a library of BPTI variants containing fluorinated amino acids at position 15 was synthesized. Kent *et al.* described a SPPS-based synthesis of BPTI, based on the synthesis of two fragments, Arg1-Gly37 and Cys38-Ala58, which were connected by NCL.[112] In this study, the first full-length synthesis of BPTI by microwave-assisted SPPS was presented, in which yields in peptide synthesis were significantly increased and DMF usage was cut to 19 % compared to the initial synthesis approach (Figure 5.1). In addition, it was shown that in the 43rd coupling cycle of SPPS, non-canonical amino acids were incorporated with only 1.5 equiv. as compared to 5 equiv. in a standard coupling. After purification of the BPTI variants, they were oxidatively refolded to generate the correctly folded proteins. An extended overview of the refolding process is discussed in section 6 “Unpublished

work", subsection 6.1 "Oxidative refolding of BPTI". Secondary structure analysis was performed by CD spectroscopy to confirm the native conformation.

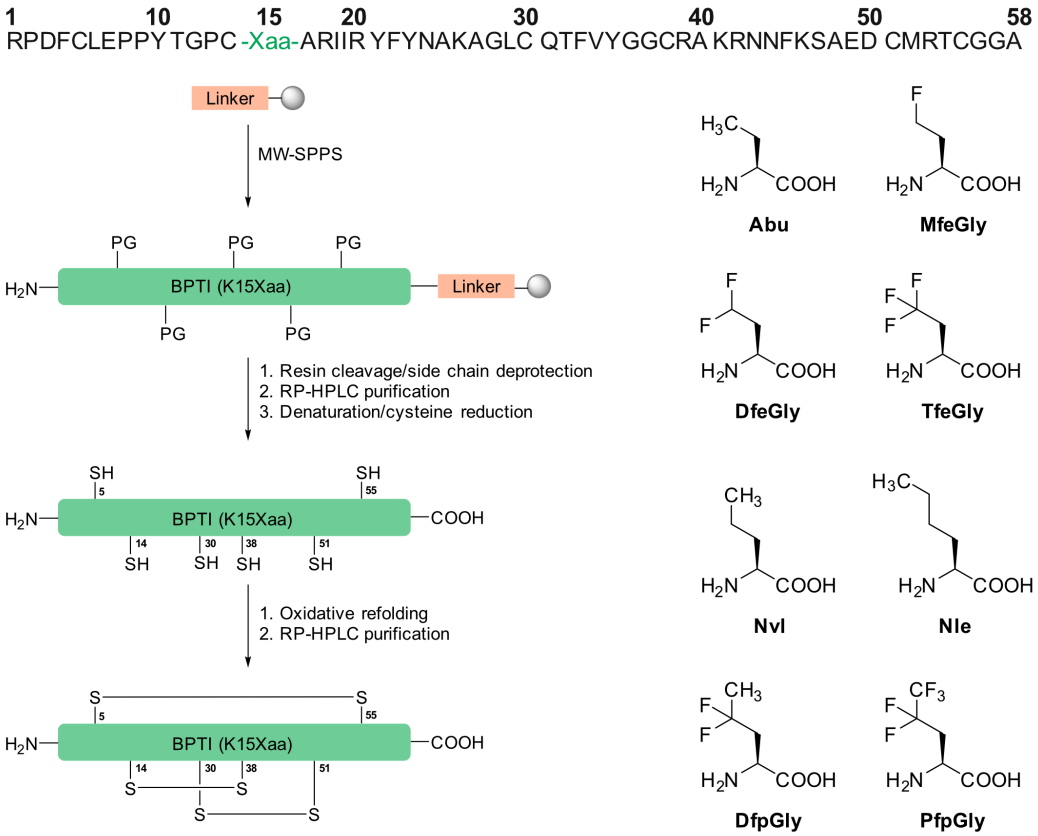


Figure 5.1. Primary sequence of BPTI (**top**); full-length synthesis and refolding strategy of BPTI(K15Xaa) (**left**); non-canonical amino acids incorporated at position 15 in BPTI (**right**)

The inhibitory activity of the BPTI K15 variants, containing the non-canonical amino acids shown in Figure 5.1, was tested in a UV-based assay towards α-chymotrypsin. The findings showed the lowest activity at BPTI containing Abu with 1.37 ± 0.17 µM. Interestingly, -CF_2H- (K15DfeGly) and -CF_2CH_3-containing (K15DfpGly) BPTI showed the best inhibitory activity towards chymotrypsin, even than wild-type BPTI. Highly fluorinated TfeGly and PfpGly containing showed activities of 0.44 ± 0.12 µM and 277 ± 42 nM, respectively. This is why the varying activities might be explained due to fluorine-induced differences in the polarity of the side chain.

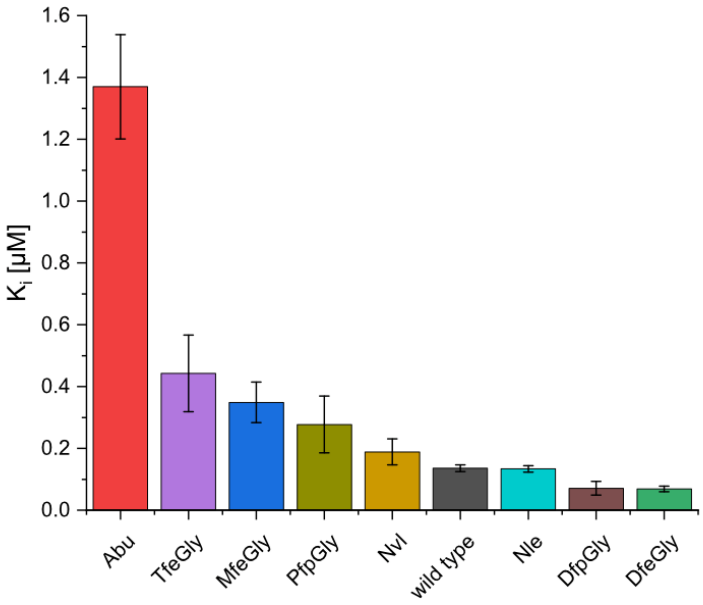


Figure 5.2. Inhibition of α-chymotrypsin with BPTI variants; comparison of K_i [µM] of BPTI variants, Abu, TfeGly, MfeGly, PfpGly, Nvl, wild type, Nle, DfpGly, and DfeGly

Furthermore, BPTI K15 variants were crystallized in complex with serine protease α-chymotrypsin at high resolutions, ranging from 1.83 to 1.99 Å. The complex of BPTI K15MfeGly and chymotrypsin (Figure 5.3, **B**) is the first crystal structure reported in the PDB, with MfeGly incorporated in its protein sequence.

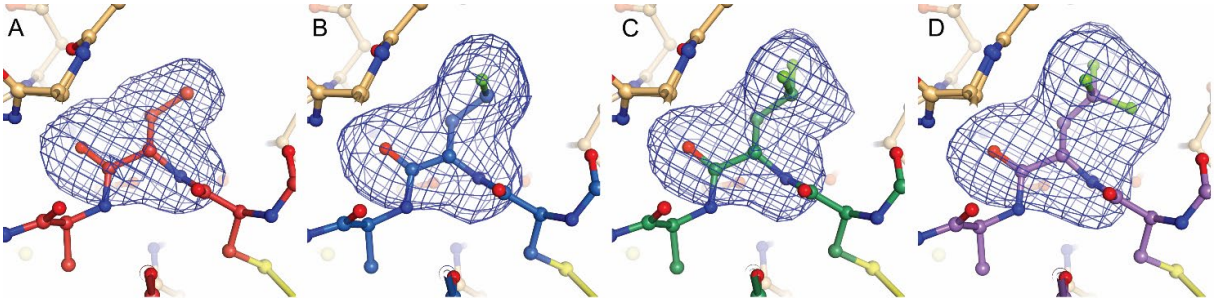


Figure 5.3. Sideview of the P1 residue of BPTI that has been omitted for calculation of a simulated annealing omit map contoured at 3.0 σ shown as blue mesh; K15Abu (**A**), K15MfeGly (**B**), K15DfeGly (**C**), K15TfeGly (**D**)

It is argued, that fluorine induces polarity in otherwise hydrophobic aliphatic side chains of amino acids. Thus, employing fluorinated amino acids as bioorthogonal tools in protein engineering will find applications in fields like drug development, influencing protein-protein interactions.

For copyright reasons, the article is not included in the online version of this thesis.
An electronic version of the article is available
(DOI: https://doi.org/10.1039/D2CB00018K)

6 Unpublished work

6.1 Oxidative refolding of BPTI

Purified full-length BPTI, synthesized according to chapter 5.2, was denatured by treatment with TCEP and 6 M GdmCl, to assure no undesired disulfide bridges were formed within the protein chain. Isolation of the denatured and reduced BPTI was afforded by gel filtration over a Sephadex G-25 matrix in an almost quantitative yield with no formation of side products.

BPTI is containing six cysteines which are forming distinct disulfide bridges Cys5-Cys55, Cys14-Cys38, and Cys30-Cys51. The folding pathway of BPTI was previously discussed in chapter 3.2. Even though refolding of BPTI has been discussed and analyzed in literature multiple times,[112, 142, 143, 146, 149, 155, 158, 168, 223] it remains challenging from a chemical synthesis perspective. Initial *in vitro* experiments for testing the refolding conditions of BPTI, to form the native structure lead to the formation of physical aggregates, which precipitated in the refolding buffer. Therefore, multiple parameters had to be adjusted before refolded BPTI could be successfully isolated. Reproducing the procedure described by Ye *et al.* with an open-air oxidation process was not successful.

The conditions, which are of great importance for successful results in refolding BPTI, are protein concentration, pH, ionic strength, redox agents, temperature and further additives.

It is described in literature,[149, 165] that refolding is significantly faster at pH 8.7, this is why Tris-HCl buffer was chosen, since it is buffering in a range from pH 7.0 to 9.0. A suitable redox system assisting in the disulfide bond formation is consisting of reduced (GSH) and oxidized glutathione (GSSG). In literature, varying concentrations of GSH and GSSG are reported, best results were performed with 5-fold excess compared to the protein concentration.[149] The GSH/GSSG redox system is replacing the reshuffling properties of the protein disulfide-isomerase (PDI) which is employed in *in vivo* refolding processes.[150]

To guarantee a sufficient level of ionic strength in the refolding buffer, KCl was added. However, all refolding experiments were unsuccessful, until non-denaturing conditions (1 M) GdmCl were added to the buffer, preventing the formation of physical aggregates. After GdmCl was added to the refolding buffer, correctly folded BPTI could be isolated and purified by RP-HPLC (chapter 6.4.1).

6.2 Inhibition of β-trypsin

The activity of refolded BPTI variants containing Abu, DfeGly, and TfeGly were tested towards serine protease β-trypsin. The results were in agreement with the previously published data by Ye *et al.*.[100] The library was extended by BPTI(K15MfeGly) which showed increased inhibitory activity compared to the Abu variant, yet lower activity than the difluorinated BPTI.

In addition to that, the activity of BPTI variants containing Nvl, Nle, DfpGly and PfpGly were tested in which PfpGly showed the lowest activity of all variants, whereas activity increased from Nvl to Nle. Interestingly, DfpGly-containing BPTI showed enhanced activity compared to the hydrocarbon analogues.

A detailed overview of all K_i values in comparison to literature values can be found in Table 6.1.

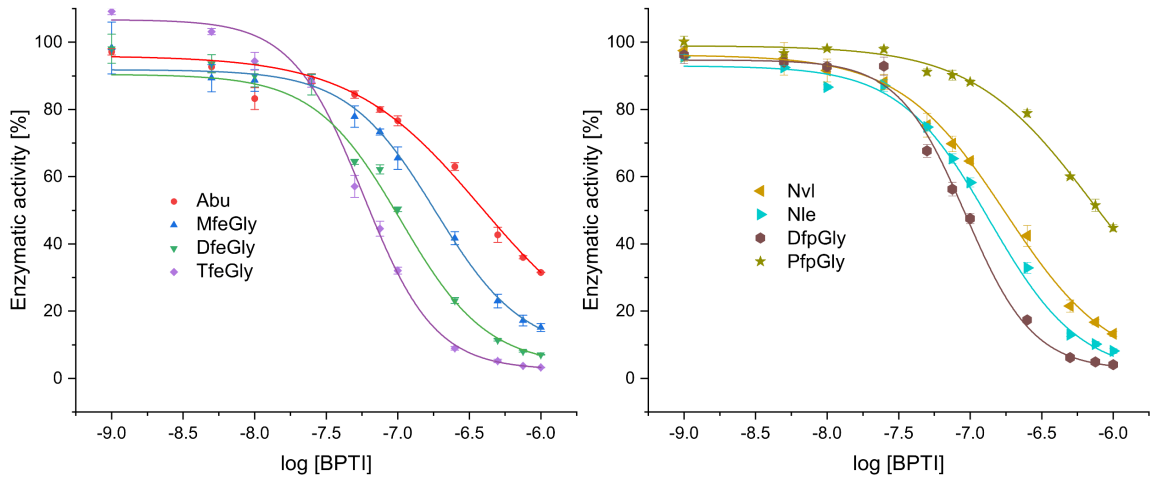


Figure 6.1. Inhibition of β-trypsin with BPTI variants. Enzymatic activity plotted against log[BPTI] showing sigmoidal curve, wild type, Abu, MfeGly, DfeGly, TfeGly (**left**), wild type, Nvl, Nle, DfpGly, PfpGly (**right**).

K_i values for wild type BPTI could not be determined as the performed activity assay is limited by the enzyme concentration [E]. In case of complexes with a K_i lower than the sensitivity range, the measured activity curve is determined by [E] and not by the inhibitor concentration [I].[224] For wild type BPTI the described literature values were determined with a different method[132, 184] resulting in values of 10^{-13} to 10^{-14} M (10^{-4} to 10^{-5} nM). As in the performed assay [E] = 100 nM, the resulting detection limit for K_i was 50 nM, thus making it not possible to determine an accurate K_i for wild type BPTI in complex with trypsin. Since wild type BPTI and trypsin for such a strong complex, it is experimentally very difficult to determine the corresponding binding or inhibitor constant and has only been described twice in literature so far. [132, 184] The described K_i

values for wild type BPTI by Ye *et al.* should be seen as not accurate as they were measured in a similar activity assay the performed one.[100]

Table 6.1. Comparison of K_i values determined in this study to K_i values from literature.

BPTI variant	K_i (exp.) [M]	K_i (lit.) [M]
wild type	-	$1.93 \cdot 10^{-8}$ [a]/ $6 \cdot 10^{-14}$ [b]
K15Abu	$3.44 \cdot 10^{-7}$	-[c]
K15Nvl	$1.59 \cdot 10^{-7}$	$1.3 \cdot 10^{-7}$
K15Nle	$1.23 \cdot 10^{-7}$	$2.0 \cdot 10^{-8}$
K15MfeGly	$1.64 \cdot 10^{-7}$	- [c]
K15DfeGly	$9.41 \cdot 10^{-8}$	$2.58 \cdot 10^{-8}$
K15TfeGly	$5.15 \cdot 10^{-8}$	$1.92 \cdot 10^{-8}$
K15DfpGly	$8.36 \cdot 10^{-8}$	- [c]
K15PfpGly	$6.22 \cdot 10^{-7}$	- [c]

[a] Ref.[100] [b] Ref. [184] [c] Not described in literature so far.

6.3 Protein crystallography

Crystallization experiments afforded single crystals of trypsin-BPTI(K15MfeGly) (Figure 6.2).

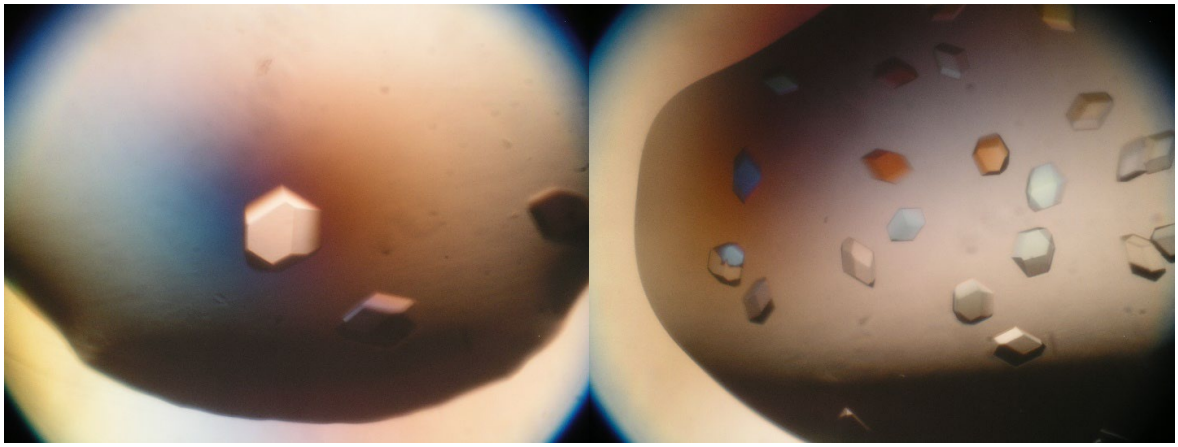


Figure 6.2. Crystals of β-trypsin-BPTI(K15MfeGly) complex under microscope (~ 60 x)

Data set for the protease-inhibitor complex were obtained in a resolution of 1.18 Å and structure was deposited in the PDB under 7PH1.[225] The crystallographic analysis of the MfeGly-containing complex showed, that as published by Ye *et al.*[100] for the complexes containing Abu and its fluorinated derivatives DfeGly and TfeGly, changes compared to

wild type BPTI in the stoichiometry and dynamics of the structural water molecules in the S1 pocket of trypsin were observed (Figure 6.3).

Further analysis revealed, the B-factors of structural water molecules B and C (Table 6.2) are 23.5 and 24.1 Å^2, which is in agreement to values previously published by Koksch and coworkers.[100]

Table 6.2. Comparison of B-factors of water molecules in the S1-binding pocket in β-trypsin-BPTI complexes [Å^2]

	Abu	MfeGly	DfeGly	TfeGly	wild type
A/ A'	21.8/–	21.1/–	17.5/–	18.4/–	–/13.5
B	27.7	23.5	20.0	20.5	—
C	31.2	24.1	22.2	20.7	—
D/D'	23.0/–	18.6/–	17.8/–	17.9/–	–/15.4
E/E'	20.7/–	24.2/–	19.1/–	19.4/–	–/14.1

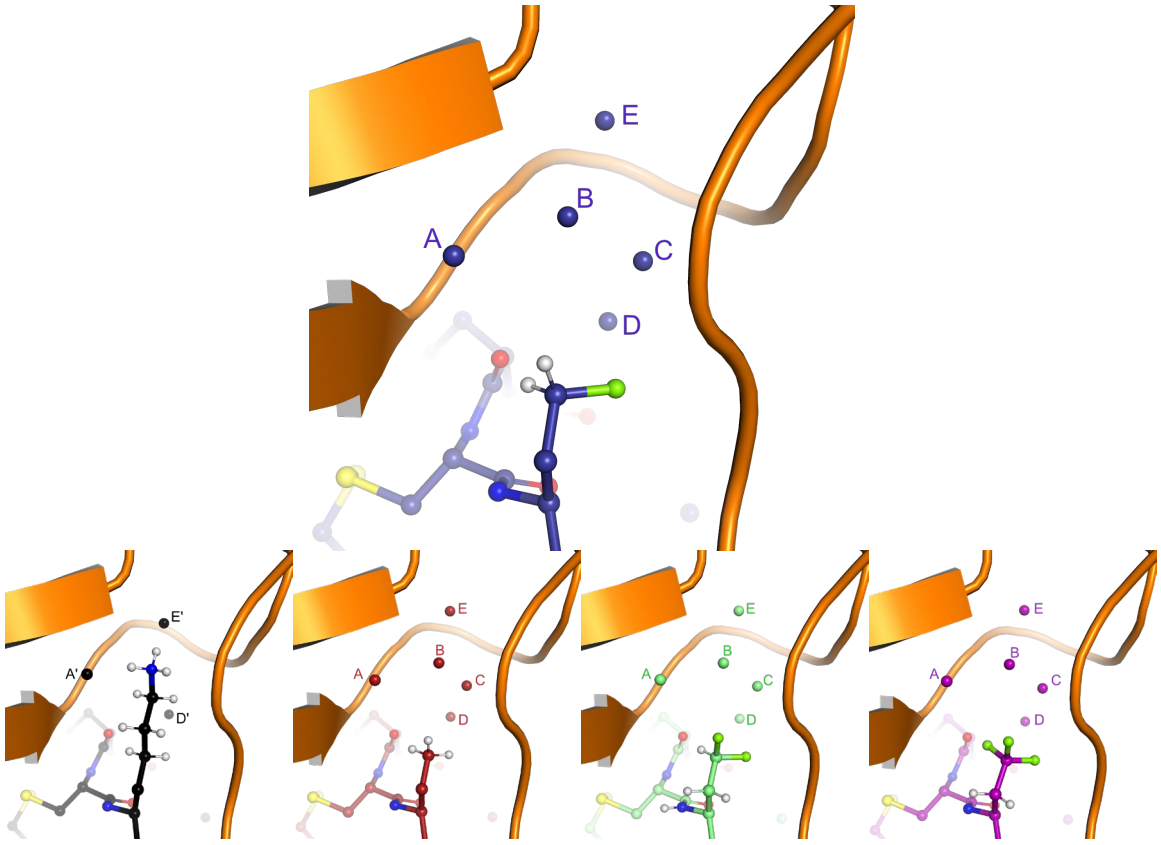


Figure 6.3. Structural context of the P1 side chain in the S1 binding pocket. β-trypsin shown in cartoon representation, BPTI K15MfeGly and structural water molecules in blue (PDB code: 7PH1)(**top**), K15wt and structural water molecules in black (PDB code: 4Y0Y) (**bottom, left**), BPTI K15Abu and structural water molecules in red (PDB code: 4Y0Z) (**bottom, center left**), BPTI K15DfeGly and structural water molecules in green (PDB code: 4Y10) (**bottom, center right**), BPTI K15TfeGly and structural water molecules in purple(PDB code: 4Y11) (**bottom, right**). Structural water molecules named A-E for BPTI variants and A', D', and E' for wild type BPTI, according to Ye *et al.*[100]

6.4 Experimental section

6.4.1 Protein refolding

Purified BPTI variants were dissolved in buffer containing 6 M GdmCl, 0.1 M Na_2HPO_4, 20 mM TCEP at pH 7.0 with a final concentration of 1.5 mM. After shaking for 24 h at 25 °C, buffer exchange to 10 mM HCl was performed with a PD-10 desalting column (Sephadex G-25 M, GE Healthcare). Flow through was collected and lyophilized overnight. Lyophilized product was dissolved in 10 mM HCl and dropwise added to degassed oxidation buffer containing 0.1 M Tris, 0.2 M KCl, 1 mM EDTA, 1 M GdmCl, 150 µM GSSG and 300 µM GSH at pH 8.7 with a final peptide concentration of 30 µM. Mixture was shaken for 24 h at 37 °C. Solution was concentrated in Amicon® Ultra-15 3 kDa Centrifugal Filter Unit (Merck, Darmstadt, Germany), washed with 10 mM HCl and lyophilized. Crude refolded BPTI variants were purified with RP-HPLC according to gradient shown in Table 6.3.

6.4.2 Preparative HPLC of refolded BPTI derivatives

Purification of crude proteins was performed using a reversed-phase HPLC LaPrep∑ low-pressure HPLC system (VWR, Darmstadt, Germany). The system consists of a LaPrep∑ LP 1200 preparative solvent pump with a 100 mL titanium pump head, a ternary low-pressure gradient, a dynamic mixing chamber, a 6-port-3-channel injection valve with an automated preparative 20 mL sample loop, a LaPrep∑ LP 3101 1-channel UV-detector, a LaPrep∑ semi-preparative flow cell with a 0.5 mm path length and a LaPrep∑ LP 2016 17-port/1-channel fraction valve. As stationary phase a Kinetex® RP-C18 TMS end-capped LC column (5 µm, 100 Å, 250 × 21.2 mm, Phenomenex®, Torrance, CA, USA) was used. A SecurityGuard PREP Cartridge Holder (21.2 mm, Phenomenex, Torrance, CA, USA) equipped with a C18 cartridge (15 × 21.2 mm, Phenomenex®, Torrance, CA, USA) was used as precolumn. Both eluents, deionized water (solvent A) and acetonitrile (solvent B), contained 0.1% (v/v) TFA. HPLC runs were performed as described in Table 6.3:

Table 6.3. Gradient method used for preparative HPLC for purification of refolded BPTI.

Time [min]	A [%]	B [%]
0	90	10
18	20	80
20	0	100
23	0	100
24	90	10
27	90	10

Flow rate = 15 mL min^{-1}

Data analysis was performed with EZChrom Elite software (Version 3.3.2 SP2, Agilent Technologies, Santa Clara, CA, USA).

6.4.3 Inhibitory activity assay

For inhibitory activity assays increasing concentrations of BPTI variants (0, 1, 5, 10, 25, 50, 75, 100, 250, 500, 750, 1000 nM) were incubated with 20 µL β-trypsin (100 nM, Sigma-Aldrich) in 200 mM Triethanolamine, 20 mM $CaCl_2$, pH 7.8 in a 96-well plate for 1 h. 20 µL *N*α-Benzoyl-L-arginine 4-nitroanilide hydrochloride (BApNA, Sigma-Aldrich) (1 mM) were added to 180 µLs preincubated enzyme/inhibitor mixture. Hydrolysis of BApNA and formation of the product p-Nitroaniline (pNA) was monitored by measuring absorbance at 405 nm in an Infinite M200 microplate reader (Tecan Group AG, Männedorf, Switzerland) for 30 min at 25 °C. Initial velocities were determined by plotting absorbance against reaction time. The IC_{50} values were determined by plotting residual enzyme activity against logarithmic inhibitor concentration using OriginPro 2020b (OriginLab Corporation, Northampton, MA, USA).

6.4.4 Protein crystallography

To form complex between trypsin and BPTI variants, 60 µL of 2.0 mM BPTI in 25 mM Tris, 10 mM $CaCl_2$, pH 7.4 were added to 50 µL of bovine trypsin in 25 mM Tris, 10 mM $CaCl_2$, pH 7.4. Solution was incubated for 12 h at 4 °C. β-trypsin-BPTI complex was purified by dialysis. Initial crystals were obtained by the sitting-drop vapor-diffusion method at 18 °C with a reservoir solution containing 2.2 M ammonium sulfate. Inter-grown crystals were used to prepare a seed stock. With a cat whisker, seeds were transferred to a freshly prepared crystallization drop. Well-formed crystals were soaked in 30% glycerol plus reservoir solution and frozen in liquid nitrogen. Data was collected at Berlin BESSYII, beamline 14.2. X-ray data collection was performed at 100 K. Diffraction data were processed with the XDS in space group *I*222 (Table 6.4).[226] The structure of the Trypsin-BPTI-MfeGly was solved by molecular replacement with the coordinates of Chymotrypsin-BPTI (PDB: 4Y11),[227] downloaded from the PDB-REDO[228] server as search models using PHASER.[229] All remaining structures were determined by isomorphous replacement. The structures were refined by maximum-likelihood restrained refinement and TLS refinement[230] using PHENIX followed by iterative, manual model building cycles with COOT.[231] Model quality was evaluated with MolProbity.[232] Figures were prepared using PyMOL (Schroedinger Inc). The atomic coordinates and structure factor amplitudes

have been deposited in the Protein Data Bank under the accession code 7PH1 (Trypsin-BPTI-MfeGly).

Table 6.4. Crystallographic data collection, refinement, and validation statistics.

Dataset	Trypsin/BPTI MfeGly
PDB entry	7PH1
Data Collection	
Wavelength [Å]	0.9184
Temperature [K]	100
Space group	*I*222
Unit Cell Parameters	
a, b, c [Å]	74.97, 81.29, 124.25
α, β, γ [°]	90.0, 90.0, 90.0
Resolution [Å][a]	30.00 - 1.18 (1.25 - 1.18)
Reflections [a]	
Unique [a]	124,357 (19,736)
Completeness [%][a]	99.8 (98.9)
Multiplicity [a]	7.4 (7.5)
Data quality [a]	
Intensity [I/σ(I)] [a]	10.02 (0.91)
R_{meas} [%][a,b]	9.3 (204.9)
$CC_{1/2}$ [a,c]	99.9 (45.6)
Wilson B value [Å^2]	19.0
Refinement	
Resolution [Å][a]	30.00 - 1.18 (1.21 - 1.18)
Reflections [a]	
Number	124,220
Test Set [%]	1.5
R_{work} [%][a]	14.9 (34.3)
R_{free} [%][a]	16.6 (35.8)
Asymmetric Unit	
Protein: Residues, Atoms	223 (E), 3,693 (E) 57 (I), 1,029 (I)
Ligands: Molecules	1 (Ca^{2+}), 8 (glycerol), 14 (SO_4^{2-})
Water molecules	322
Mean Temperature factors [Å^2][b]	
All Atoms	22.9
Macromolecules	23.8 (E), 20.5 (I)
Ligands	23.2 (Ca^{2+}), 36.9 (glycerol) 36.5 (SO_4^{2-})
Water molecules	32.4

RMSD from Target Geometry [d]	
Bond Lengths [Å]	0.015
Bond Angles [°]	1.512
Validation Statistics	
Ramachandran Plot [f]	
Residues in Allowed Regions [%]	1.5
Residues in Favored Regions [%]	98.5
Ramachandran plot *Z*-score [f] (RMSD)	
whole	-0.25 (0.41)
helix	1.92 (0.74)
sheet	-0.97 (0.49)
loop	-0.20 (0.39)
MOLPROBITY Clashscore [g]	1.62
MOLPROBITY score [f]	0.88

[a] data for the highest resolution shell in parenthesis

[b] $R_{meas}(I) = \sum_h [N/(N-1)]^{1/2} \sum_i | I_{ih} - \langle I_h \rangle | / \sum_h \sum_i I_{ih}$, in which $\langle I_h \rangle$ is the mean intensity of symmetry-equivalent reflections h, I_{ih} is the intensity of a particular observation of h and N is the number of redundant observations of reflection h. [233]

[c] $CC_{1/2} = (\langle I^2 \rangle - \langle I \rangle^2) / (\langle I^2 \rangle - \langle I \rangle^2) + \sigma^2_\varepsilon$, in which σ^2_ε is the mean error within a half-dataset.[234]

[d] RMSD – root mean square deviation

[e] calculated with PHENIX [235]

[f] calculated with MOLPROBITY [232]

[g] Clashscore is the number of serious steric overlaps (> 0.4) per 1,000 atoms.[232]

7 Summary and Outlook

This work aimed at the site-specific incorporation of fluorinated amino acids into the serine protease inhibitor BPTI to investigate fluorine's impact on protein-protein interactions. The overarching goal was the investigation of the influence fluorination has on the inhibitory properties of BPTI towards the serine protease α-chymotrypsin.

In the first part of this project, a new synthetic strategy to obtain Fmoc-protected MfeGly in a gram-scale was established since access to this fluorinated amino acid has been limited in the past and the amino acid is commercially not available. In this novel synthesis, the amino acid could be produced in eight steps with an overall yield of up to 50 %.[213] Alongside MfeGly, further Fmoc-protected fluorinated derivatives of Abu were synthesized according to procedures described in the literature.[70, 72, 236]

With the fluorinated amino acids at hand, a full-length SPPS protocol of BPTI, containing non-canonical amino acids at position 15 of the sequence, was developed, increasing the overall yield of the peptide and reducing solvent consumption and reaction time to a fraction of previously employed synthesis strategies.[112, 129] Furthermore, for the first time a coupling cycle in SPPS using only an excess of 1.5 equiv. of fluorinated amino acid compared to the usual 5 equiv. was described. The full-length BPTI variants were oxidatively folded to their native conformations, for which a folding procedure was developed in this work. The native protein state was verified by secondary structure analysis by CD spectroscopy. A library of BPTI variants containing various fluorinated amino acids and their hydrocarbon analogues at position 15 was obtained: Abu, MfeGly, DfeGly, TfeGly, Nvl, DfpGly, PfpGly, and Nle. The inhibitory activity of the BPTI variants was tested towards the serine protease α-chymotrypsin. A key finding was that the variants containing DfeGly and DfpGly showed the best inhibition properties, even better than Lys-containing wild-type BPTI. In addition, crystallization experiments and X-ray diffraction analysis of four BPTI variants, K15Abu, K15MfeGly, K15DfeGly, and K15TfeGly, in complex with α-chymotrypsin were performed. The resolution of the crystals obtained ranged from 1.83 to 1.99 Å. The MfeGly-containing crystal structure of the BPTI-chymotrypsin complex was the first one described in the PDB with this residue in its sequence.

In the future, the inhibitory activity of the BPTI variants towards other serine proteases should be tested. This work showed that DfeGly- and DfpGly-containing BPTI showed increased inhibition towards chymotrypsin. In contrast, in the study of Ye *et al.* from 2015,

the K15TfeGly variant showed the best inhibitory activity towards β-trypsin. Unpublished initial screenings showed that the fluorinated BPTI K15TfeGly does not inhibit the coagulation Factor XIa, while wild-type BPTI inhibits this enzyme in a μM range. The results demonstrate that fluorination can have a beneficial impact on protein-protein interactions however it is still challenging to predict the precise changes in properties. Yet, the results demonstrate, that the modification of the P1 position in an enzyme inhibitor with a broad specificity as BPTI can drastically influence the specificity towards serine proteases. This can be utilized in fine-tuning the activity of inhibitors towards proteases. Understanding the underlying mechanisms fluorine is influencing will be a great challenge in the years to come.

In further studies, it will be beneficial to investigate other less synthetically challenging systems than BPTI to study fluorinated amino acids in the context of protein-protein interactions. The factor of size might not pose as the biggest obstacle in obtaining BPTI; still, oxidative refolding of cysteine-rich proteins *in vitro* remains not trivial and is the bottleneck in the overall synthetic pathway. Once other systems of interest are identified, the focus should lay on widening the range of amino acids to investigate. Since polar and charged residues are involved in many examples of molecular recognition, fluorinated derivatives of these amino acids need to be synthesized and systematically studied in the context of a peptide/protein environment.

Another important aspect in future studies will be the possibility of incorporating any given non-canonical amino acid into proteins by *in vivo* methods. Even though efforts are being undertaken to enable the incorporation of ncAA into proteins, it remains challenging.[215]

Fluorine in the context of protein engineering still remains an exciting research field that will benefit from technological developments in the years to come.

8 Curriculum Vitae

For data security reasons the curriculum vitae has been omitted from the published version.

9 Bibliography

1. D. L. Nelson, M. M. Cox and A. L. Lehninger, *Lehninger principles of biochemistry*, W.H. Freeman and Company, New York, NY, Seventh edition edn., 2017.
2. A. L. Weber and S. L. Miller, *J. Mol. Evol.*, 1981, **17**, 273-284.
3. H. Moissan, *C. R. Acad. Sci.*, 1886, **102**, 1543–1544.
4. R. E. Banks, *Journal of Fluorine Chemistry*, 1986, **33**, 3-26.
5. M. Salwiczek, E. K. Nyakatura, U. I. M. Gerling, S. Ye and B. Koksch, *Chem. Soc. Rev.*, 2012, **41**, 2135-2171.
6. P. Vlieghe, V. Lisowski, J. Martinez and M. Khrestchatisky, *Drug Discov. Today*, 2010, **15**, 40-56.
7. F. Vella, *Biochemical Education*, 1992, **20**, 121-122.
8. W. E. Stites, *Chem. Rev.*, 1997, **97**, 1233-1250.
9. C. Jäckel and B. Koksch, *European Journal of Organic Chemistry*, 2005, **2005**, 4483-4503.
10. D. O'Hagan, *Chem. Soc. Rev.*, 2008, **37**, 308-319.
11. A. Bondi, *The Journal of Physical Chemistry*, 1964, **68**, 441-451.
12. J. W. Banks, A. S. Batsanov, J. A. K. Howard, D. O'Hagan, H. S. Rzepa and S. Martin-Santamaria, *Journal of the Chemical Society, Perkin Transactions 2*, 1999, DOI: 10.1039/A907452J, 2409-2411.
13. T. Allmendinger, P. Furet and E. Hungerbühler, *Tetrahedron Lett.*, 1990, **31**, 7297-7300.
14. T. Allmendinger, E. Felder and E. Hungarbühler, *Tetrahedron Lett.*, 1990, **31**, 7301-7304.
15. J. J. Urban, B. G. Tillman and W. A. Cronin, *The Journal of Physical Chemistry A*, 2006, **110**, 11120-11129.
16. S. Couve-Bonnaire, D. Cahard and X. Pannecoucke, *Org. Biomol. Chem.*, 2007, **5**, 1151-1157.
17. S. Oishi, H. Kamitani, Y. Kodera, K. Watanabe, K. Kobayashi, T. Narumi, K. Tomita, H. Ohno, T. Naito, E. Kodama, M. Matsuoka and N. Fujii, *Org. Biomol. Chem.*, 2009, **7**, 2872-2877.
18. A. Sun, D. C. Lankin, K. Hardcastle and J. P. Snyder, *Chemistry – A European Journal*, 2005, **11**, 1579-1591.
19. C. Thiehoff, Y. P. Rey and R. Gilmour, *Isr. J. Chem.*, 2017, **57**, 92-100.
20. F. Akkerman, J. Buschmann, D. Lentz, P. Luger and E. Rödel, *J. Chem. Crystallogr.*, 2003, **33**, 969-975.
21. K. B. Wiberg, M. A. Murcko, K. E. Laidig and P. J. MacDougall, *The Journal of Physical Chemistry*, 1990, **94**, 6956-6959.
22. P. Huber-Wälchli and H. H. Günthard, *Chem. Phys. Lett.*, 1975, **30**, 347-351.
23. J. D. Dunitz and R. Taylor, *Chemistry – A European Journal*, 1997, **3**, 89-98.
24. F. R. Fischer, W. B. Schweizer and F. Diederich, *Angew. Chem. Int. Ed.*, 2007, **46**, 8270-8273.
25. G. T. Giuffredi, V. Gouverneur and B. Bernet, *Angew. Chem. Int. Ed.*, 2013, **52**, 10524-10528.
26. H.-J. Schneider, *Chemical Science*, 2012, **3**, 1381-1394.
27. B. Breiten, M. R. Lockett, W. Sherman, S. Fujita, M. Al-Sayah, H. Lange, C. M. Bowers, A. Heroux, G. Krilov and G. M. Whitesides, *Journal of the American Chemical Society*, 2013, **135**, 15579-15584.
28. C. Dalvit, C. Invernizzi and A. Vulpetti, *Chemistry – A European Journal*, 2014, **20**, 11058-11068.

29. C. Dalvit and A. Vulpetti, *Chemistry – A European Journal*, 2016, **22**, 7592-7601.
30. B. Biswas and P. C. Singh, *Journal of Fluorine Chemistry*, 2019, DOI: https://doi.org/10.1016/j.jfluchem.2019.109414, 109414.
31. J. F. Darby, A. P. Hopkins, S. Shimizu, S. M. Roberts, J. A. Brannigan, J. P. Turkenburg, G. H. Thomas, R. E. Hubbard and M. Fischer, *Journal of the American Chemical Society*, 2019, DOI: 10.1021/jacs.9b06275.
32. I. Hyla-Kryspin, G. Haufe and S. Grimme, *Chemistry – A European Journal*, 2004, **10**, 3411-3422.
33. K. Müller, C. Faeh and F. Diederich, *Science*, 2007, **317**, 1881-1886.
34. G. W. Gribble, *Acc. Chem. Res.*, 1998, **31**, 141-152.
35. D. O'Hagan and D. B. Harper, *Journal of Fluorine Chemistry*, 1999, **100**, 127-133.
36. J. Tannock, *Rhod. J. Agric. Res.*, 1975, **13**, 67.
37. G. O. Morton, J. E. Lancaster, G. E. Van Lear, W. Fulmor and W. E. Meyer, *Journal of the American Chemical Society*, 1969, **91**, 1535-1537.
38. X. M. Zhu, S. Hackl, M. N. Thaker, L. Kalan, C. Weber, D. S. Urgast, E. M. Krupp, A. Brewer, S. Vanner, A. Szawiola, G. Yim, J. Feldmann, A. Bechthold, G. D. Wright and D. L. Zechel, *ChemBioChem*, 2015, **16**, 2498-2506.
39. J. Han, L. Kiss, H. Mei, A. M. Remete, M. Ponikvar-Svet, D. M. Sedgwick, R. Roman, S. Fustero, H. Moriwaki and V. A. Soloshonok, *Chem. Rev.*, 2021, **121**, 4678-4742.
40. D. O'Hagan, C. Schaffrath, S. L. Cobb, J. T. G. Hamilton and C. D. Murphy, *Nature*, 2002, **416**, 279-279.
41. C. D. Cadicamo, J. Courtieu, H. Deng, A. Meddour and D. O'Hagan, *ChemBioChem*, 2004, **5**, 685-690.
42. C. Schaffrath, H. Deng and D. O'Hagan, *FEBS Lett.*, 2003, **547**, 111-114.
43. D. O'Hagan and H. Deng, *Chem. Rev.*, 2015, **115**, 634-649.
44. T. Kittilä, P. Calero, F. Fredslund, P. T. Lowe, D. Tezé, M. Nieto-Domínguez, D. O'Hagan, P. I. Nikel and D. H. Welner, *Microbial Biotechnology*, 2022, **n/a**.
45. K.-H. van Pee, *Curr. Org. Chem.*, 2012, **16**, 2583-2597.
46. S. Huang, L. Ma, M. H. Tong, Y. Yu, D. O'Hagan and H. Deng, *Org. Biomol. Chem.*, 2014, **12**, 4828-4831.
47. A. S. Eustáquio, D. O'Hagan and B. S. Moore, *J. Nat. Prod.*, 2010, **73**, 378-382.
48. L. Martarello, C. Schaffrath, H. Deng, A. D. Gee, A. Lockhart and D. O'Hagan, *Journal of Labelled Compounds and Radiopharmaceuticals*, 2003, **46**, 1181-1189.
49. H. Deng, S. L. Cobb, A. D. Gee, A. Lockhart, L. Martarello, R. P. McGlinchey, D. O'Hagan and M. Onega, *Chem. Commun.*, 2006, DOI: 10.1039/B516861A, 652-654.
50. D. E. Ponde, C. S. Dence, N. Oyama, J. Kim, Y.-C. Tai, R. Laforest, B. A. Siegel and M. J. Welch, *J. Nucl. Med.*, 2007, **48**, 420.
51. J. Marik, A. Ogasawara, B. Martin-McNulty, J. Ross, J. E. Flores, H. S. Gill, J. N. Tinianow, A. N. Vanderbilt, M. Nishimura, F. Peale, C. Pastuskovas, J. M. Greve, N. van Bruggen and S. P. Williams, *J. Nucl. Med.*, 2009, **50**, 982.
52. X.-G. Li, J. Domarkas and D. O'Hagan, *Chem. Commun.*, 2010, **46**, 7819-7821.
53. X.-G. Li, S. Dall'Angelo, L. F. Schweiger, M. Zanda and D. O'Hagan, *Chem. Commun.*, 2012, **48**, 5247-5249.
54. X.-G. Li, A. Autio, H. Ahtinen, K. Helariutta, H. Liljenbäck, S. Jalkanen, A. Roivainen and A. J. Airaksinen, *Chem. Commun.*, 2013, **49**, 3682-3684.
55. X.-G. Li, K. Helariutta, A. Roivainen, S. Jalkanen, J. Knuuti and A. J. Airaksinen, *Nat. Protoc.*, 2014, **9**, 138-145.
56. H. Cai and P. S. Conti, *Journal of Labelled Compounds and Radiopharmaceuticals*, 2013, **56**, 264-279.

57. S. Dall'Angelo, Q. Zhang, I. N. Fleming, M. Piras, L. F. Schweiger, D. O'Hagan and M. Zanda, *Org. Biomol. Chem.*, 2013, **11**, 4551-4558.
58. M. E. Sergeev, F. Morgia, M. R. Javed, M. Doi and P. Y. Keng, *J. Mol. Catal. B: Enzym.*, 2013, **97**, 74-79.
59. S. Thompson, Q. Zhang, M. Onega, S. McMahon, I. Fleming, S. Ashworth, J. H. Naismith, J. Passchier and D. O'Hagan, *Angew. Chem. Int. Ed.*, 2014, **53**, 8913-8918.
60. Y. Ogawa, E. Tokunaga, O. Kobayashi, K. Hirai and N. Shibata, *iScience*, 2020, **23**.
61. M. Inoue, Y. Sumii and N. Shibata, *ACS Omega*, 2020, **5**, 10633-10640.
62. J. Han, A. M. Remete, L. S. Dobson, L. Kiss, K. Izawa, H. Moriwaki, V. A. Soloshonok and D. O'Hagan, *Journal of Fluorine Chemistry*, 2020, **239**, 109639.
63. D. O'Hagan, *Journal of Fluorine Chemistry*, 2010, **131**, 1071-1081.
64. S. Purser, P. R. Moore, S. Swallow and V. Gouverneur, *Chem. Soc. Rev.*, 2008, **37**, 320-330.
65. H. J. Böhm, D. Banner, S. Bendels, M. Kansy, B. Kuhn, K. Müller, U. Obst-Sander and M. Stahl, *ChemBioChem*, 2004, **5**, 637-643.
66. B. E. Smart, *Journal of Fluorine Chemistry*, 2001, **109**, 3-11.
67. K. Ganesan, S. K. Raza and R. Vijayaraghavan, *Journal of Pharmacy and Bioallied Sciences*, 2010, **2**, 166 - 178.
68. R. Stone, *Science*, 2018, **359**, 1314-1315.
69. J. Moschner, V. Stulberg, R. Fernandes, S. Huhmann, J. Leppkes and B. Koksch, *Chem. Rev.*, 2019, **119**, 10718-10801.
70. T. Tsushima, K. Kawada, S. Ishihara, N. Uchida, O. Shiratori, J. Higaki and M. Hirata, *Tetrahedron*, 1988, **44**, 5375-5387.
71. H. A. Lindner, V. V. Lunin, A. Alary, R. Hecker, M. Cygler and R. Ménard, *J. Biol. Chem.*, 2003, **278**, 44496-44504.
72. D. Winkler and K. Burger, *Synthesis*, 1996, **1996**, 1419-1421.
73. S. Kröger and G. Haufe, *Amino Acids*, 1997, **12**, 363-372.
74. W. J. Middleton, *The Journal of Organic Chemistry*, 1975, **40**, 574-578.
75. A. A. Berger, J.-S. Völler, N. Budisa and B. Koksch, *Acc. Chem. Res.*, 2017, **50**, 2093-2103.
76. W. Hoffmann, J. Langenhan, S. Huhmann, J. Moschner, R. Chang, M. Accorsi, J. Seo, J. Rademann, G. Meijer, B. Koksch, M. T. Bowers, G. von Helden and K. Pagel, *Angew. Chem. Int. Ed.*, 2019, **58**, 8216-8220.
77. J. R. Robalo, L. M. Streacker, D. Mendes de Oliveira, P. Imhof, D. Ben-Amotz and A. Vila Verde, *Journal of the American Chemical Society*, 2019, DOI: 10.1021/jacs.9b06862.
78. J. R. Robalo, S. Huhmann, B. Koksch and A. Vila Verde, *Chem*, 2017, **3**, 881-897.
79. A. K. Sato, M. Viswanathan, R. B. Kent and C. R. Wood, *Curr. Opin. Biotechnol.*, 2006, **17**, 638-642.
80. K. Sanford and M. Kumar, *Curr. Opin. Biotechnol.*, 2005, **16**, 416-421.
81. R. B. Merrifield, *Journal of the American Chemical Society*, 1963, **85**, 2149-2154.
82. R. Behrendt, P. White and J. Offer, *Journal of Peptide Science*, 2016, **22**, 4-27.
83. J. P. Tam, J. Xu and K. D. Eom, *Peptide Science*, 2001, **60**, 194-205.
84. S. L. Pedersen, A. P. Tofteng, L. Malik and K. J. Jensen, *Chem. Soc. Rev.*, 2012, **41**, 1826-1844.
85. N. Hartrampf, A. Saebi, M. Poskus, Z. P. Gates, A. J. Callahan, A. E. Cowfer, S. Hanna, S. Antilla, C. K. Schissel, A. J. Quartararo, X. Ye, A. J. Mijalis, M. D. Simon, A. Loas, S. Liu, C. Jessen, T. E. Nielsen and B. L. Pentelute, *Science*, 2020, **368**, 980.
86. H. Sun and A. Brik, *Accounts of Chemical Research*, 2019, **52**, 3361-3371.

87. B. C. Buer, J. L. Meagher, J. A. Stuckey and E. N. G. Marsh, *Proceedings of the National Academy of Sciences*, 2012, **109**, 4810.
88. C. Jäckel, M. Salwiczek and B. Koksch, *Angew. Chem. Int. Ed.*, 2006, **45**, 4198-4203.
89. M. Salwiczek and B. Koksch, *ChemBioChem*, 2009, **10**, 2867-2870.
90. M. Salwiczek, S. Samsonov, T. Vagt, E. Nyakatura, E. Fleige, J. Numata, H. Cölfen, M. T. Pisabarro and B. Koksch, *Chemistry – A European Journal*, 2009, **15**, 7628-7636.
91. O. D. Monera, C. M. Kay and R. S. Hodges, *Protein Sci.*, 1994, **3**, 1984-1991.
92. L. M. Gottler, H.-Y. Lee, C. E. Shelburne, A. Ramamoorthy and E. N. G. Marsh, *ChemBioChem*, 2008, **9**, 370-373.
93. L. M. Gottler, R. de la Salud Bea, C. E. Shelburne, A. Ramamoorthy and E. N. G. Marsh, *Biochemistry*, 2008, **47**, 9243-9250.
94. T. Panchenko, W. W. Zhu and J. K. Montclare, *Biotechnology and Bioengineering*, 2006, **94**, 921-930.
95. T. H. Yoo, A. J. Link and D. A. Tirrell, *Proceedings of the National Academy of Sciences*, 2007, **104**, 13887-13890.
96. P. P. Pal, J. H. Bae, M. K. Azim, P. Hess, R. Friedrich, R. Huber, L. Moroder and N. Budisa, *Biochemistry*, 2005, **44**, 3663-3672.
97. N. Budisa, P. P. Pal, S. Alefelder, P. Birle, T. Krywcun, M. Rubini, W. Wenger, J. H. Bae and T. Steiner, *Journal*, 2004, **385**, 191.
98. B. Holzberger and A. Marx, *Journal of the American Chemical Society*, 2010, **132**, 15708-15713.
99. B. Holzberger, M. Rubini, H. M. Möller and A. Marx, *Angew. Chem. Int. Ed.*, 2010, **49**, 1324-1327.
100. S. Ye, B. Loll, A. A. Berger, U. Mulow, C. Alings, M. C. Wahl and B. Koksch, *Chemical Science*, 2015, **6**, 5246-5254.
101. B. C. Buer and E. N. G. Marsh, *Protein Sci.*, 2012, **21**, 453-462.
102. R. Huber, D. Kukla, A. Rühlmann and W. Steigemann, *Cold Spring Harbor Symposia on Quantitative Biology*, 1972, **36**, 141-150.
103. R. Huber, D. Kukla, A. Ruhlmann, O. Epp and H. Formanek, *Naturwissenschaften*, 1970, **57**, 389-392.
104. P. Ascenzi, A. Bocedi, M. Bolognesi, A. Spallarossa, M. Coletta, R. De Cristofaro and E. Menegatti, *Curr. Protein Peptide Sci.*, 2003, **4**, 231-251.
105. H. Kraut, E. K. Frey and E. Werle, *Journal*, 1930, **192**, 1.
106. M. Kunitz and J. H. Northrop, *The Journal of General Physiology*, 1936, **19**, 991-1007.
107. J. Deisenhofer and W. Steigemann, *Acta Crystallographica Section B*, 1975, **31**, 238-250.
108. J. A. McCammon, B. R. Gelin and M. Karplus, *Nature*, 1977, **267**, 585-590.
109. G. Wagner and K. Wüthrich, *J. Mol. Biol.*, 1982, **155**, 347-366.
110. A. Paolo, B. Alessio, B. Martino, S. Andrea, C. Massimo, C. Raimondo De and M. Enea, *Current Protein & Peptide Science*, 2003, **4**, 231-251.
111. M. Ferrer, C. Woodward and G. Barany, *International Journal of Peptide and Protein Research*, 1992, **40**, 194-207.
112. W. Lu, M. A. Starovasnik and S. B. H. Kent, *FEBS Lett.*, 1998, **429**, 31-35.
113. A. Meta, H. Nakatake, T. Imamura, C. Nozaki and K. Sugimura, *Protein Expression and Purification*, 2009, **66**, 22-27.
114. B. Kassell, M. Radicevic, M. J. Ansfield and M. Laskowski, Sr., *Biochem Biophys Res Commun*, 1965, **18**, 255-258.
115. J. Siekmann, H. R. Wenzel, W. Schroder and H. Tschesche, *Biol. Chem. Hoppe Seyler*, 1988, **369**, 157-163.

116. E. Wachter, K. Deppner, K. Hochstrasser, K. Lempart and R. Geiger, *FEBS Lett.*, 1980, **119**, 58-62.
117. M. J. Burgering, L. P. Orbons, A. van der Doelen, J. Mulders, H. J. Theunissen, P. D. Grootenhuis, W. Bode, R. Huber and M. T. Stubbs, *J. Mol. Biol.*, 1997, **269**, 395-407.
118. C. A. Sprecher, F. J. Grant, G. Grimm, P. J. O'Hara, F. Norris, K. Norris and D. C. Foster, *Biochemistry*, 1993, **32**, 4481-4486.
119. R. E. Tanzi, J. F. Gusella, P. C. Watkins, G. A. Bruns, P. St George-Hyslop, M. L. Van Keuren, D. Patterson, S. Pagan, D. M. Kurnit and R. L. Neve, *Science*, 1987, **235**, 880-884.
120. E. Levy, A. Amorim, B. Frangione and L. C. Walker, *Neurobiol. Aging*, 1995, **16**, 805-808.
121. B. De Strooper, F. Van Leuven and H. Van den Berghe, *Biochim. Biophys. Acta*, 1991, **1129**, 141-143.
122. R. Sandbrink, C. L. Masters and K. Beyreuther, *Biochim Biophys Acta*, 1994, **1219**, 167-170.
123. B. D. Shivers, C. Hilbich, G. Multhaup, M. Salbaum, K. Beyreuther and P. H. Seeburg, *EMBO J.*, 1988, **7**, 1365-1370.
124. Y. Hokama, S. Iwanaga, T. Tatsuki and T. Suzuki, *J. Biochem.*, 1976, **79**, 559-578.
125. I. Kato and N. Tominaga, *Fed. Proc.*, 1979, **38**, 832.
126. M. D. Adams, S. E. Celniker, R. A. Holt, C. A. Evans and J. D. Gocayne, *Science*, 2000, **287**, 2185-2195.
127. H. Tschesche and T. Dietl, *Eur. J. Biochem.*, 1975, **58**, 439-451.
128. H. C. Krebs and G. G. Habermehl, *Naturwissenschaften*, 1987, **74**, 395-396.
129. S. Ye, B. Loll, A. A. Berger, U. Mulow, C. Alings, M. C. Wahl and B. Koksch, *Chem. Sci.*, 2015, **6**, 5246-5254.
130. M. Marquart, J. Walter, J. Deisenhofer, W. Bode and R. Huber, *Acta Crystallographica Section B*, 1983, **39**, 480-490.
131. J. P. Vincent and M. Lazdunski, *Biochemistry*, 1972, **11**, 2967-2977.
132. J.-P. Vincent and M. Lazdunski, *Eur. J. Biochem.*, 1973, **38**, 365-372.
133. J. P. Vincent, M. Peron-Renner, J. Pudles and M. Lazdunski, *Biochemistry*, 1974, **13**, 4205-4211.
134. W. Bode, *J. Mol. Biol.*, 1979, **127**, 357-374.
135. E. Antonini, P. Ascenzi, M. Bolognesi, G. Gatti, M. Guarneri and E. Menegatti, *J. Mol. Biol.*, 1983, **165**, 543-558.
136. J. Walter, R. Huber and S. Brenner, *J. Mol. Biol.*, 1983, **167**, 911-917.
137. A. Wlodawer, J. Walter, R. Huber and L. Sjölin, *J. Mol. Biol.*, 1984, **180**, 301-329.
138. A. Wlodawer, J. Deisenhofer and R. Huber, *J. Mol. Biol.*, 1987, **193**, 145-156.
139. S. Parkin, B. Rupp and H. Hope, *Acta Crystallogr. D Biol. Crystallogr.*, 1996, **52**, 18-29.
140. K. Kawamura, T. Yamada, K. Kurihara, T. Tamada, R. Kuroki, I. Tanaka, H. Takahashi and N. Niimura, *Acta Crystallographica Section D*, 2011, **67**, 140-148.
141. C. Capasso, M. Rizzi, E. Menegatti, P. Ascenzi and M. Bolognesi, *J. Mol. Recognit.*, 1997, **10**, 26-35.
142. N. J. Darby, P. E. Morin, G. Talbo and T. E. Creighton, *J. Mol. Biol.*, 1995, **249**, 463-477.
143. T. E. Creighton and D. P. Goldenberg, *J. Mol. Biol.*, 1984, **179**, 497-526.
144. P. A. Kosen, T. E. Creighton and E. R. Blout, *Biochemistry*, 1981, **20**, 5744-5754.
145. P. A. Kosen, T. E. Creighton and E. R. Blout, *Biochemistry*, 1980, **19**, 4936-4944.
146. T. E. Creighton, *Progress in Biophysics and Molecular Biology*, 1979, **33**, 231-297.

147. G. I. Makhatadze, K.-S. Kim, C. Woodward and P. L. Privalov, *Protein Sci.*, 1993, **2**, 2028-2036.
148. Y. Kuroda and P. S. Kim, *J. Mol. Biol.*, 2000, **298**, 493-501.
149. R. Mousa, S. Lansky, G. Shoham and N. Metanis, *Chemical Science*, 2018, **9**, 4814-4820.
150. J. L. Arolas, F. X. Aviles, J.-Y. Chang and S. Ventura, *Trends in Biochemical Sciences*, 2006, **31**, 292-301.
151. S. E. Radford, *Trends in Biochemical Sciences*, 2000, **25**, 611-618.
152. G. Bulaj and D. P. Goldenberg, 1999, **8**, 1825-1842.
153. Y. Kuroda and P. S. Kim, *Journal of Molecular Biology*, 2000, **298**, 493-501.
154. D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan and W. Wriggers, 2010, **330**, 341-346.
155. M. Qin, W. Wang and D. Thirumalai, *Proceedings of the National Academy of Sciences*, 2015, **112**, 11241-11246.
156. A. P. J. Middelberg, *Trends Biotechnol.*, 2002, **20**, 437-443.
157. M. Dadlez and P. S. Kim, *Biochemistry*, 1996, **35**, 16153-16164.
158. J. Weissman and P. Kim, *Science*, 1991, **253**, 1386-1393.
159. J. S. Weissman and P. S. Kim, *Nat. Struct. Biol.*, 1995, **2**, 1123.
160. J. S. Weissman and P. S. Kim, *Cell*, 1992, **71**, 841-851.
161. D. P. Goldenberg, *Trends in Biochemical Sciences*, 1992, **17**, 257-261.
162. F. M. Kibria and W. J. Lees, *Journal of the American Chemical Society*, 2008, **130**, 796-797.
163. M. Okumura, M. Saiki, H. Yamaguchi and Y. Hidaka, *The FEBS Journal*, 2011, **278**, 1137-1144.
164. B. Chatrenet and J. Y. Chang, *J. Biol. Chem.*, 1993, **268**, 20988-20996.
165. B. S. Mamathambika and J. C. Bardwell, 2008, **24**, 211-235.
166. B.-Y. Lu and J.-Y. Chang, *Analytical Biochemistry*, 2005, **339**, 94-103.
167. K. A. Dill, S. B. Ozkan, M. S. Shell and T. R. Weikl, *Annual Review of Biophysics*, 2008, **37**, 289-316.
168. T. G. Oas and P. S. Kim, *Nature*, 1988, **336**, 42.
169. D. A. Debe, M. J. Carlson and W. A. Goddard, *Proceedings of the National Academy of Sciences*, 1999, **96**, 2596.
170. V. R. Agashe, M. C. R. Shastry and J. B. Udgaonkar, *Nature*, 1995, **377**, 754-757.
171. N. J. Marianayagam and S. E. Jackson, *Biophysical Chemistry*, 2004, **111**, 159-171.
172. A. W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Žídek, A. W. R. Nelson, A. Bridgland, H. Penedones, S. Petersen, K. Simonyan, S. Crossan, P. Kohli, D. T. Jones, D. Silver, K. Kavukcuoglu and D. Hassabis, *Nature*, 2020, **577**, 706-710.
173. E. Callaway, *Nature*, 2020, **588**, 203-204.
174. M. Varadi, S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Žídek, T. Green, K. Tunyasuvunakool, S. Petersen, J. Jumper, E. Clancy, R. Green, A. Vora, M. Lutfi, M. Figurnov, A. Cowie, N. Hobbs, P. Kohli, G. Kleywegt, E. Birney, D. Hassabis and S. Velankar, *Nucleic Acids Res.*, 2021, DOI: 10.1093/nar/gkab1061, gkab1061.
175. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli and D. Hassabis, *Nature*, 2021, **596**, 583-589.

176. K. Oda, *The Journal of Biochemistry*, 2012, **151**, 13-25.
177. D. M. Blow, *Trends in Biochemical Sciences*, 1997, **22**, 405-408.
178. N. D. Rawlings and A. J. Barrett, *Nucleic acids research*, 2000, **28**, 323-325.
179. L. Hedstrom, *Chemical Reviews*, 2002, **102**, 4501-4524.
180. R. Ménard and A. C. Storer, 1992, **373**, 393-400.
181. C. G. Whiteley, *Cell Biochemistry and Biophysics*, 2000, **33**, 217-225.
182. H. Jering and H. Tschesche, *Eur. J. Biochem.*, 1976, **61**, 453-463.
183. E. Fioretti, M. Angeletti, G. Lupidi and M. Coletta, *European Journal of Biochemistry*, 1994, **225**, 459-465.
184. H. Tschesche, J. Beckmann, A. Mehlich, E. Schnabel, E. Truscheit and H. R. Wenzel, *Biochim. Biophys. Acta*, 1987, **913**, 97-101.
185. C. P. Sommerhoff, W. Bode, G. Matschiner, A. Bergner and H. Fritz, *Biochim Biophys Acta*, 2000, **1477**, 75-89.
186. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.*, 2000, **28**, 235-242.
187. T. Brinkmann, J. Schafers, L. Gurtler, H. Kido, Y. Niwa, N. Katunuma and H. Tschesche, *J. Protein Chem.*, 1997, **16**, 651-660.
188. D. Belorgey, S. Dirrig, M. Amouric, C. Figarella and J. G. Bieth, *The Biochemical journal*, 1996, **313 (Pt 2)**, 555-560.
189. U. Quast, J. Engel, H. Heumann, G. Krause and E. Steffen, *Biochemistry*, 1974, **13**, 2512-2520.
190. M. Coletta, P. Ascenzi, G. Amiconi, M. Bolognesi, M. Guarneri and E. Menegatti, *Biophys. Chem.*, 1990, **37**, 355-362.
191. M. Coletta, P. Ascenzi, L. Bravin, G. Amiconi, M. Bolognesi, M. Guarneri and E. Menegatti, *Journal of Biomolecular Structure and Dynamics*, 1990, **7**, 959-972.
192. J.-P. Vincent and M. Lazdunski, *FEBS Lett.*, 1976, **63**, 240-244.
193. E. Menegatti, M. Guarneri, M. Bolognesi, P. Ascenzi and G. Amiconi, *J. Mol. Biol.*, 1984, **176**, 425-430.
194. P. Ascenzi, M. Coletta, G. Amiconi, R. de Cristofaro, M. Bolognesi, M. Guarneri and E. Menegatti, *Biochim. Biophys. Acta*, 1988, **956**, 156-161.
195. P. Ascenzi, G. Amiconi, M. Coletta, G. Lupidi, E. Menegatti, S. Onesti and M. Bolognesi, *J. Mol. Biol.*, 1992, **225**, 177-184.
196. E. R. Guinto, J. Ye, B. F. Le Bonniec and C. T. Esmon, *J. Biol. Chem.*, 1994, **269**, 18395-18400.
197. P. Ascenzi, M. Coletta, G. Amiconi, M. Bolognesi, E. Menegatti and M. Guarneri, *Biol. Chem. Hoppe Seyler*, 1990, **371**, 389-393.
198. A. Grzesiak, I. Krokoszynska, D. Krowarsch, O. Buczek, M. Dadlez and J. Otlewski, *J. Biol. Chem.*, 2000, **275**, 33346-33352.
199. P. Ascenzi, G. Amiconi, M. Bolognesi, E. Menegatti and M. Guarneri, *Biochim. Biophys. Acta*, 1990, **1040**, 134-136.
200. E. Menegatti, M. Guarneri, M. Bolognesi, P. Ascenzi and G. Amiconi, *J. Mol. Biol.*, 1986, **191**, 295-297.
201. E. Fioretti, M. Angeletti, M. Coletta, P. Ascenzi, M. Bolognesi, E. Menegatti, M. Rizzi and F. Ascoli, *Journal of Enzyme Inhibition*, 1993, **7**, 57-64.
202. L. Kiczak, M. Kasztura, K. Koscielska-Kasprzak, M. Dadlez and J. Otlewski, *Biochim. Biophys. Acta*, 2001, **1550**, 153-163.
203. M. S. Tallman, C. Nabhan, J. H. Feusner and J. M. Rowe, *Blood*, 2002, **99**, 759-767.
204. C. Figarella, G. A. Negri and O. Guy, *Eur. J. Biochem.*, 1975, **53**, 457-463.
205. K. Ohlsson and A. S. Olsson, *Hoppe Seylers Z. Physiol. Chem.*, 1976, **357**, 1153-1161.

206. S. Schmid, W. Uhl and M. W. Buchler, *Scand. J. Gastroenterol. Suppl.*, 1996, **219**, 47-50.
207. R. Berling, A. Borgstrom and K. Ohlsson, *Int. J. Pancreatol.*, 1998, **24**, 9-17.
208. M. Verstraete, in *Haemostatsis and thrombosis*, eds. A. L. Bloom, C. D. Forbes, D. P. Thomas and E. G. D. Tuddenham, Churchill Livingstone, Edinburgh, 3 edn., 1994, pp. 1057-1073.
209. C. Kluft, in *Blood Use in Cardiac surgery*, eds. N. Friedel, R. Hetzer and D. HRoyston, Springer-Verlag, New York, 1991, pp. 10-15.
210. G. Venturini, M. Colasanti and P. Ascenzi, *Biochem Biophys Res Commun*, 1998, **249**, 263-265.
211. I. Favre, G. W. Moss, D. P. Goldenberg, J. Otlewski and E. Moczydlowski, *Biochemistry*, 2000, **39**, 2001-2012.
212. D. Bojkova, M. Bechtel, K.-M. McLaughlin, J. E. McGreig, K. Klann, C. Bellinghausen, G. Rohde, D. Jonigk, P. Braubach, S. Ciesek, C. Münch, M. N. Wass, M. Michaelis and J. Cinatl, *Cells*, 2020, **9**.
213. J. Leppkes, T. Hohmann and B. Koksch, *Journal of Fluorine Chemistry*, 2020, **232**, 109453.
214. K. Laue, S. Kröger, E. Wegelius and G. Haufe, *European Journal of Organic Chemistry*, 2000, **2000**, 3737-3743.
215. J.-S. Völler, M. Dulic, U. I. M. Gerling-Driessen, H. Biava, T. Baumann, N. Budisa, I. Gruic-Sovulj and B. Koksch, *ACS Central Science*, 2017, **3**, 73-80.
216. B. C. Buer, B. J. Levin and E. N. G. Marsh, *Journal of the American Chemical Society*, 2012, **134**, 13027-13034.
217. U. I. M. Gerling, M. Salwiczek, C. D. Cadicamo, H. Erdbrink, C. Czekelius, S. L. Grage, P. Wadhwani, A. S. Ulrich, M. Behrends, G. Haufe and B. Koksch, *Chemical Science*, 2014, **5**, 819-830.
218. V. Asante, J. Mortier, G. Wolber and B. Koksch, *Amino Acids*, 2014, **46**, 2733-2744.
219. R. Smits and B. Koksch, *Curr. Top. Med. Chem.*, 2006, **6**, 1483-1498.
220. D. W. Kim, Jeong, S. T. Lim, M.-H. Sohn, J. A. Katzenellenbogen and D. Y. Chi, *The Journal of Organic Chemistry*, 2008, **73**, 957-962.
221. D. W. Kim, H.-J. Jeong, S. T. Lim and M.-H. Sohn, *Angew Chem Int Ed Engl*, 2008, **47**, 8404-8406.
222. J. Leppkes, N. Dimos, B. Loll, T. Hohmann, M. Dyrks, A. R. Wieseke, B. G. Keller and B. Koksch, *RSC Chemical Biology*, 2022, DOI: 10.1039/D2CB00018K.
223. E. Moses and H.-J. Hinz, *J. Mol. Biol.*, 1983, **170**, 765-776.
224. J. Smirnovienė, L. Baranauskienė, A. Zubrienė and D. Matulis, *Eur. Biophys. J.*, 2021, **50**, 345-352.
225. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.*, 2000, **28**, 235-242.
226. W. Kabsch, *Acta Crystallogr. D Biol. Crystallogr.*, 2010, **66**, 125-132.
227. H. Czapinska, R. Helland, A. O. Smalås and J. Otlewski, *J. Mol. Biol.*, 2004, **344**, 1005-1020.
228. R. P. Joosten, F. Long, G. N. Murshudov and A. Perrakis, *IUCrJ*, 2014, **1**, 213-220.
229. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read, *J Appl Crystallogr*, 2007, **40**, 658-674.
230. M. D. Winn, M. N. Isupov and G. N. Murshudov, *Acta crystallographica. Section D, Biological crystallography*, 2001, **57**, 122-133.
231. A. Casanal, B. Lohkamp and P. Emsley, *Protein Sci*, 2020, **29**, 1069-1078.
232. C. J. Williams, J. J. Headd, N. W. Moriarty, M. G. Prisant, L. L. Videau, L. N. Deis, V. Verma, D. A. Keedy, B. J. Hintze, V. B. Chen, S. Jain, S. M. Lewis, W. B. Arendall, 3rd,

J. Snoeyink, P. D. Adams, S. C. Lovell, J. S. Richardson and D. C. Richardson, *Protein Sci*, 2018, **27**, 293-315.
233. K. Diederichs and P. A. Karplus, *Nat. Struct. Biol.*, 1997, **4**, 269-275.
234. P. A. Karplus and K. Diederichs, *Science*, 2012, **336**, 1030-1033.
235. P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart, *Acta crystallographica. Section D, Biological crystallography*, 2010, **66**, 213-221.
236. T. Hohmann, M. Dyrks, S. Chowdhary, M. Weber, D. Nguyen, J. Moschner and B. Koksch, *ChemRxiv*, 2022, DOI: 10.26434/chemrxiv-2022-wt95d.