INAUGURAL – DISSERTATION

zur

Erlangung der Doktorwürde

Der

Gesamtfakultät für Mathematik,

Ingenieur- und Naturwissenschaften

der

Ruprecht-Karls-Universität

Heidelberg

Vorgelegt von

Florian Raphael Häge

Tag der mündlichen Prüfung: 08.08.2024

Thema

Design of Structured Adhesion Miniproteins for Tissue Engineering

Gutachter: Jun.-Prof. Dr. Franziska Thomas

Prof. Dr. Milan Kivala

Kurzzusammenfassung

Gewebezüchtung, die Erzeugung, Regeneration oder Modifikation von Geweben, benötigt zelladhäsive Matrizen, um Zellwachstum zu ermöglichen. [1] Materialien auf Basis synthetischer Hydrogele gewinnen dank ihrer steuerbaren Zusammensetzung immer mehr an Beliebtheit, aber müssen in der Regel funktionalisiert werden, um ihren Zweck zu erfüllen. Die zwei gängigsten Strategien sind die Immobilisierung von Zelladhäsionsmotiven oder die Immobilisierung von ganzen Domänen extrazellulärer Matrix-Proteine. Die kurzen Zelladhäsionsmotive sind leicht chemisch modifizierbar, besitzen allerdings keine dreidimensionale Struktur. Die zelladhäsiven Proteindomänen erbringen die nötige Bioaktivität, aber denaturieren oft bei Immobilisierung und sind durch chemische Synthese nicht zugänglich. [2] In dieser Arbeit wird ein Kompromiss der beiden Strategien präsentiert: das Design von Adhäsions-Miniproteinen zur Funktionalisierung von Matrizen für Gewebezüchtung.

Hierzu wurde das Design von calciumbindenden Miniproteinen mit einer mikromolaren Dissoziationskonstante (K_d) zur Zelladhäsion durch Ca(II)-gestützte Kohlenhydratbindung am Beispiel von D-Typ Lektinen zum Ziel gesetzt. [3] Drei literaturbekannte Ansätze wurden auf neun verschiedene Miniproteingerüste angewandt: computergestütztes Design mit der Modellierungs-Software *Rosetta3*, [4] intuitives Design von *de novo* Bindungsstellen aus komplexen Salzbrücken [5] und die Transplantation („Grafting“) funktionaler Domänen. [6, 7] Das computergestützte Design mit *Rosetta3* wurde außerdem auf Eignung zur Installation von Metallbindungsstellen in Miniproteinen geprüft. Hierfür sollte eine Kupferbindung auf Basis des aktiven Zentrums von Cu,Zn-Superoxid-Dismutase in das Gerüst einer SH3-Domäne integriert werden. Modelle wurden mittels Festphasenpeptidsynthese synthetisiert und mit Hinblick auf Struktur, Stabilität und Bindungseigenschaften untersucht.

Das computergestützte Design mittels *Rosetta3* basierte auf der Inkorporation des aktiven Zentrums des gewählten natürlichen Modells in die Sequenz des Gerüstproteins. [4] Die erzeugten Modelle lieferten die Basis für das Design von WWcalm4, einem Miniprotein auf Basis der WW-Domäne, welches seine Konformation in der Präsenz von Ca(II) ändert. Intuitives Design baute auf komplexen Salzbrücken an der Oberfläche des Peptidgerüsts. Positiv geladene Aminosäuren der Salzbrücke wurden zu negativ geladenen mutiert. Somit entstand aus der zirkular permutierten WW-Domäne von Kier et al. das Modell WWturn mit einer negativ geladenen Bindungsstelle. [8] Das Miniprotein faltete in die erwartete Struktur, zeigte eine hohe Thermostabilität mit einer Schmelztemperatur (T_m) von 55 ± 0.5 °C und band Ca(II) mit einer K_d = 7.9 ± 3.4 µM. Der Transplantationsansatz wurde genutzt um funktionale Bindungsmotive in Schleifen und *N*-Termini von Peptidgerüsten einzufügen. [6, 9] EF-Hand 2 von Calmodulin wurde an den *N*-Terminus des Tryptophan-Käfigs („Trp-cage“) von Neidigh et al. im Modell Calmcage transplantiert. [10] Das Fusionsminiprotein faltete in eine zum Trp-cage ähnliche Struktur mit einer erhöhten Thermostabilität von T_m = 50 ± 3 °C. Calmcage band Tb(III) mit einer K_d = 200 ± 40 µM und Ca(II) mit einer K_d = 64 ± 26 mM. Ein weiteres erfolgreiches Transplantationsdesign wurde durch die Inkorporation der EF-Hand 3 von Calmodulin in die Schleife des β-Haarnadel-Systems C4 von Anderson et al. im Modell Calmzip3 erzeugt. [11] Das Fusionsminiprotein faltete in eine dem C4-System ähnliche Struktur und änderte seine Konformation in Anwesenheit von Ca(II) mit einer Schmelztemperatur von T_m = 47.5 ± 1.0 °C im Holo- und T_m = 47 ± 3 °C im Apo-Zustand. Calmzip3 zeigte eine starke Tb(III)-Bindung mit einer K_d = 0.96 ± 0.15 µM und eine Ca(II)-Bindung im selben Bereich wie die schwächste Calmodulin- und die stärkste Laminin-Bindung mit einer K_d = 3.9 ± 1.4 µM.

Das computergestützte Design eines aktiven Zentrums einer Superoxid-Dismutase (SOD) in das Gerüst der SH3-Domäne erzeugte das Modell SO1. Es faltete in eine β-Faltblattstruktur

mit hoher Thermostabilität, die in der Anwesenheit von bivalenten Übergangsmetallionen von T_m = 47 ± 0.5 °C für den Holo- auf T_m = 74 ± 1 °C für den Ni(II)-gebundenen Zustand stieg. Im Cu(II)-gebundenen Zustand zeigte SO1 SOD-Enzymaktivität in derselben Größenordnung wie kleine organische Cu(II)-Komplexe mit einer errechneten Geschwindigkeitskonstante k_{SOD} = 6.14 ± 0.58 × 10^7 M^{-1} s^{-1} Größenordnungen schneller als SOD-Mimetika auf Miniproteinbasis. [12-16] SO1 kann daher nicht nur als erfolgreichen Eignungstest für *Rosetta3* zur Installation von Metallbindungen in SH3-Domänen herangezogen werden, sondern auch als Modellsystem für SOD-Enzyme genutzt werden.

Zusammenfassend stellten sich alle drei Ansätze als geeignet zur Erzeugung funktionaler Miniproteine dar. Es wurden Ca(II)-bindende Miniproteine mit K_d-Werten vom niedrigen mikromolaren bis in den millimolaren Bereich designt. Mit ihren Bindungsaffinitäten im selben Bereich, zeichnen sie sich als ideale Kandidaten für Ca(II)-gestützte Kohlenhydratbindung am Beispiel von D-Typ Lektinen aus.

Abstract

Tissue Engineering – the generation, regeneration or modification of tissue – requires adhesive matrices for cells to grow on. [1] Non-immunogenic, hydrogel-based materials are often cell-repellent or non-adhesive and need to be functionalized. The two commonly used strategies either encompass the immobilization of cell adhesion motifs, or of entire domains of cell-adhesive extracellular matrix proteins. The short cell adhesion motifs lack a three-dimensional structure but can be chemically modified due to their short sequence length. Protein domains on the other hand bring the desired bioactivity but often denature upon immobilization, and are not chemically accessible. [2] This work was aimed to form a compromise between these strategies with the design of structured adhesion miniproteins to functionalize matrices for tissue engineering.

For this, it was aimed to design Ca(II)-binding with a micromolar dissociation constant (K_d) onto β-sheet miniprotein scaffolds for cell adhesion by Ca(II)-dependent carbohydrate recognition as seen in D-type lectins. [3] Three literature-known design approaches were applied on nine scaffolds: computational design using the *Rosetta3* modeling suite, [4] intuitive design of *de novo* binding sites from complex salt bridges, [5] and grafting of functional domains. [6, 7] The computational design of miniproteins with *Rosetta3* was further aimed to be studied by the design of a Cu(II)-binding SH3-domain miniprotein based on the active site of Cu,Zn-Superoxide Dismutase. The models were synthesized by solid-phase peptide synthesis and characterized with regards to structure, stability and binding.

Computational design using *Rosetta3* allowed the incorporation of active sites from natural model proteins into the scaffold sequence. [4] This was used as the basis for the design of WWcalm4, a WW domain-based miniprotein that changed conformation in the presence of Ca(II). Intuitive design relied on the location of complex salt bridges on the surface of miniprotein scaffolds and mutating the positively charged to negatively charged amino acids. This approach successfully installed a binding site of high negative charge on the circular permutated WW domain by Kier et al. in the model WWturn. [8] The miniprotein folded into a similar-to-native structure, was thermostable with a melting temperature (T_m) of 55 ± 0.5 °C and bound Ca(II) with a K_d = 7.9 ± 3.4 µM. The grafting approach was used to incorporate Ca(II)-binding EF-hand domains into the loops and *N*-termini of miniprotein scaffolds. [6, 9] EF-hand 2 of Calmodulin was grafted onto the *N*-terminus of the Trp-cage designed by Neidigh et al. in the Ca(II)-binding model Calmcage. [10] The fusion miniprotein folded into a similar structure to its parent peptide with an increased thermostability T_m = 50 ± 3 °C. Calmcage bound Tb(III) with a K_d = 200 ± 40 µM and Ca(II) with a K_d = 64 ± 26 mM. Another successfully grafted design was created from the EF-hand 3 of Calmodulin and the β-hairpin system C4 by Anderson et al. in the model Calmzip3. [11] It folded into a similar-to-native structure and changed conformation upon Ca(II)-binding with a thermostability of T_m = 47.5 ± 1.0 °C for the holo- and T_m = 47 ± 3 °C for the apo-protein. Calmzip3 revealed strong Tb(III) binding with a K_d = 0.96 ± 0.15 µM and Ca(II) binding in the same order of magnitude as the weakest value for Calmodulin and the strongest value for Laminin with K_d = 3.9 ± 1.4 µM.

The computational design of a Superoxide Dismutase (SOD) site on the scaffold of the SH3 domain resulted in the model SO1. It folded into a β-sheet structure with high thermostability that increased in the presence of bivalent transition metals from T_m = 47 ± 0.5 °C for the holo- to T_m = 74 ± 1 °C for the Ni(II)-bound state. It showed SOD activity in the presence of Cu(II) in the same order of magnitude as small organic Cu(II)-complex mimics, outcompeting all miniprotein SOD enzymes to date with an apparent rate constant $k_{SOD} = 6.14 \pm 0.58 \times 10^7$ M^{-1} s^{-1}. [12-16] It therefore presented itself not only as a proof of concept for computational design but also as a potential model for studying SOD enzymes.

In conclusion, all three approaches delivered functional miniproteins with Ca(II)-binding miniproteins that range from millimolar to low micromolar K_d-values. With binding affinities in the same range, they serve as ideal candidates for Ca(II)-dependent carbohydrate recognition after the example of D-type lectins.

Table of Contents

1 Introduction

Miniproteins offer a middle ground between short unfolded peptides and large protein domains inaccessible by synthesis. Engineering new functions onto these scaffolds opens up possibilities for robust, chemically synthesizable compounds for integration in a functional material, as active enzymes or receptors for specific ligand binding. However, designing miniproteins to bind specific carbohydrates has yet to be achieved and could be used to mimic extracellular matrix proteins and functionalize materials for cell adhesion. This research aims to design, synthesize and characterize miniproteins for Ca(II)-mediated carbohydrate binding based on the active centers of extracellular matrix (ECM) proteins such as laminin. This chapter will provide context for the presented research by giving a brief background, followed by the research problem, the aims and boundaries of the study, its significance and limitations.

Tissue Engineering (TE) describes the generation, regeneration or modification of tissue with three basic tools: Cells, matrices and growth factors. Unlike cells that require careful preparation for seeding and growth factors that rapidly increase the complexity of the system when applied, matrices can be designed or functionalized to meet the needs of the experiment. [1] Initially used matrices were synthetic materials already approved for medical use. Hydrolytic biodegradation released potentially cytotoxic acids that called for new materials designed for TE. Natural materials based on extracellular matrix components benefitted from a high bioactivity but often caused undesired immune responses. [17] In the case of the most commonly used natural matrix, Matrigel, the composition is not entirely known to date and batch to batch variation is high that it hampers the reproducibility of experiments. [18, 19] Synthetic hydrogels with controlled composition and tunable properties are therefore on the rise. The materials are non-immunogenic and often bio-inert but can be functionalized by a variety of additives. [20]

Functionalization of hydrogels for cell-responsiveness is generally achieved by one of two strategies: the immobilization of cell adhesion motifs recognized by cell receptors, e.g. Arg-Gly-Asp for integrins [21] or of entire domains of cell-adhesive ECM proteins such as collagen or laminin. [22, 23] Cell adhesion motifs can be chemically modified and synthesized but lack a defined three-dimensional structure and thus bioactivity. Protein domains on the other hand bring the required bioactivity but can hardly be modified and often denature upon immobilization on a surface material. [2] This work relies on a compromise that aims to map the bioactivity of protein domains and the excellent chemical modifiability of cell adhesion motifs. With the design of cell-adhesion miniproteins, it aim to engineer functional domains based on well-characterized, small and independently folding β-sheet peptide scaffolds achievable by chemical synthesis.

This study focused on designing cell adhesion by Ca(II)-mediated carbohydrate binding as found in laminin, perlecan, neurexin or agrin. The LG domains of these ECM proteins bind the heteropolysaccharide of the α-Dystroglycan receptor for cell adhesion. [24] Using intuitive or computational design, this function was introduced to independently folded β-sheet peptides such as the Tryptophan zipper, the WW domain and the SH3 domain. [25-27]

The design of adhesion miniproteins for tissue engineering represents a novel approach to material functionalization. Not only have β-sheet peptides and domains only recently gained attention for peptide engineering in general, [27] but the design of carbohydrate recognition remains an unsolved problem to date. [28] A successfully designed specific carbohydrate receptor, also known as a lectin, would therefore not only open up new possibilities for reproducible peptide-functionalized hydrogels, but also broaden the space of designed β-sheet peptides and prove that a lectin function can potentially be designed onto any protein or peptide scaffold.

The research presented in this doctoral thesis was carried out in a computationally driven low-throughput synthesis approach. Computational designs were focused on Ca(II)-binding and Ca(II)-mediated carbohydrate binding in β-sheet peptide scaffolds. Only natural amino acids were used in the designs to ensure compatibility with all design and modelling software: *Rosetta3*, *ColabFold* and *OmegaFold.*

2 State of Research

2.1 Matrices for Tissue Engineering

2.1.1 Commonly used Matrices

Tissue Engineering (TE) refers to the reconstruction or improvement of damaged tissue or the generation of new tissue using three basic components: Cells, matrices – for cell seeding, tissue formation and shaping – and growth factors – for cell proliferation and differentiation. Depending on the type of tissue generated and the conditions – *in vivo* or *in vitro* – they might be used simultaneously or individually. [1] TE has found a variety of uses as a source of tissue or organ transplants e.g. for bone transplantation [29] and skin regeneration, [30] or for fundamental research on cell and tissue behavior in three-dimensional matrices. [31] Research on TE has been ongoing for almost 40 years and is still relevant to this date.

Of the three basic tools for TE, the one scientists can influence most, is the matrix or template material. Cells need to be harvested from a host, cultured and prepared in sufficient quantity for seeding. Growth factors need to be transported to the relevant site and rapidly increase the complexity of the system when multiple ones are used simultaneously. [1] Matrices however, can be designed to serve the needs of the experiment.

Initially, well-known synthetic materials such as polylactic, polyglycolic and polycaprolactone polymers were used as matrix. These were already approved for use in other medical fields. However, the release of potentially cytotoxic acids during biodegradation made them a poor choice for TE. Natural materials based on extracellular matrix components such as gelatin, collagen, silk or hyaluronic acid – sometimes cross-linked to reduce biodegradation – were used extensively as well. Their high bioactivity makes them good candidates but may lead to undesired immune responses. [17] For example decellularized tissue matrices, or Matrigel, are commonly used as natural matrix material. While their high cell responsiveness and complexity makes them attractive for a lot of applications, the composition is partially undefined to date and varies strongly from batch to batch, hindering reproducibility. [18, 19] Synthetic hydrogels with controlled compositions and tunable properties therefore present a viable alternative. These materials are non-immunogenic, often bioinert and FDA-approved and can be functionalized by a variety of additives. [20]

2.1.2 Functionalization of Matrices

To prepare materials for cell adhesion, they are usually coated or co-polymerized with a functional additive. [32] A look at natural cell adhesion protein superfamilies reveals a plethora of different mechanisms. The Ca(II)-dependent Cadherins and cation-independent Integrins form large protein complexes that bind to the same or similar complexes on other cells by large surface contacts. [33] Integrins are widely expressed adhesion proteins that cause cell-cell or cell-ECM adhesion by cation-dependent amino acid sequence recognition. Selectins adhere to endothelial cells by Ca(II)-mediated recognition of specific carbohydrates. [34]

The traditional coating of hydrogels with cell adhesion proteins, requires their isolation from cells and may therefore cause immune responses. The proteolytic degradation of these proteins, if not replenished, limits the functional lifetime of the material. Additionally, coating positions proteins in a stochastic orientation, making them only partially useful. The immobilization can also cause the proteins to change conformation due to hydrophobic surfaces or low wetness. In the worst case, the proteins unfold and become inactive. The alternative approach is the use of small adhesion motifs like Arg-Gly-Asp (RGD) that bind integrins without a defined structure. Peptides that contain these motifs can be readily synthesized and immobilized but suffer from low activity and diversity. [32]

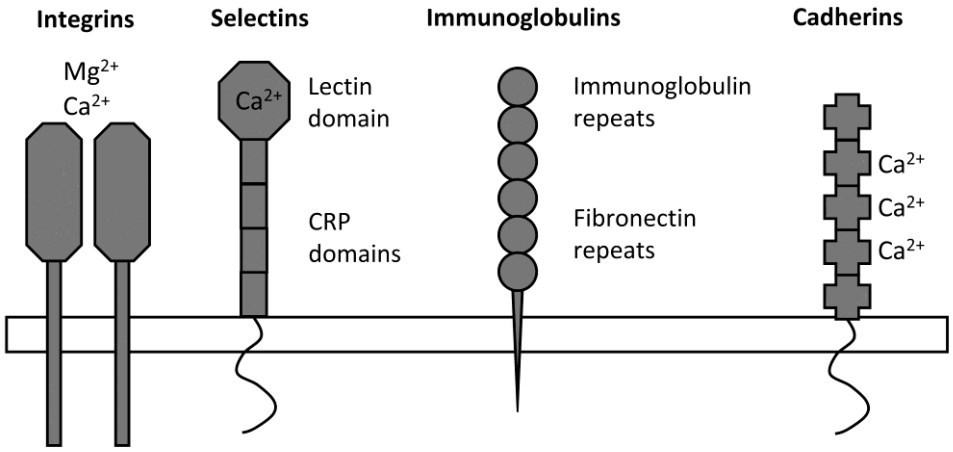


Figure 1. Scheme of cell adhesion receptors with cation-dependent integrins, Ca(II)-dependent Selectins (CRP = complement regulatory protein), Immunoglobulins and Ca(II)-dependent Cadherins (adapted from Buckley et al. Figure 1, page 168) [34]

Both functionalization strategies show unique benefits but fail where the other succeeds. This work proposes a new functionalization strategy using structured adhesion miniproteins. For this, it focuses on the Ca(II)-mediated carbohydrate recognition found in Selectins and other cell adhesion molecules, especially ECM proteins. [3, 34] With a defined three-dimensional structure, they could deliver the bioactivity of proteins at a size achievable by synthesis and mitigate the problems that both peptides and proteins face.

2.1.3 Miniprotein Scaffolds

The term miniprotein serves as a connective link between peptides and proteins. They are well-defined folds with two or more structure elements and a sequestered hydrophobic core that fold cooperatively. [35] The defined three-dimensional structure separates them from peptides, and the smaller size of up to 100 amino acids or 10 kDa that makes them achievable by synthesis, separates them from proteins. [27, 36] With a smaller hydrophobic core, less non-covalent interactions can stabilize the structure. Therefore, many miniproteins are rich in covalent cross-linking such as disulfide bridges in cysteine-knots or fold by metal-binding like in isolated EF-hands or zinc fingers. [35] To be useful for the functionalization of hydrogels, a miniprotein needs to be water-soluble and ideally monomeric.

One of the simplest examples for such a miniprotein are β-hairpins. They are made up of two β-strands connected by a loop. Hairpins made from fragments of natural proteins are however barely folded. To design new β-hairpin miniproteins, it requires a hydrophobic cluster on one side of the hairpin, strong interstrand side chain interactions, multiple charged or aromatic residues and maybe a β-capping motif to increase thermostability. A prime example are the Tryptophan zippers (Trpzip) with their well-folded, highly twisted antiparallel β-sheets. They range from 12 to 16 amino acids in length and are named for their interlocked Trp side chains that stabilize the structure. [25, 35] A more nuanced structure is found in the Tryptophan cage (Trp-cage), a 20-residue miniprotein designed from exendin-4. The protein was truncated and mutated for stabilization. The resulting miniprotein contains an α-helix, a turn and a short proline helix. The two helices interlock, enclosing a Trp side chain in Pro residues. [10, 35]

Larger miniproteins based on β-sheet structures are more difficult to design, due to lower solubility and the tendency to form β-amyloid-like structures. Designed three-stranded antiparallel β-sheet miniproteins are therefore often only partially folded in water or require other solvents, like methanol. [35] Miniproteins based on the 34- to 38-residue WW domains, can be engineered for a variety of functions such as probing carbohydrate-aromatic packing

or binding to DNA, metal ions or organophosphates, while retaining their well-defined fold. [37-40] One of the largest miniproteins at the cusp of accessibility by synthesis is based on an SH3 domain. These 56-residue structures contain five antiparallel β-strands arranged in a β-barrel closed on each side by one of the large distal and RT loops. [26] Both WW domains and SH3 domains fold independently and retain their native function, binding proline-rich sequences, when isolated. [27, 41]

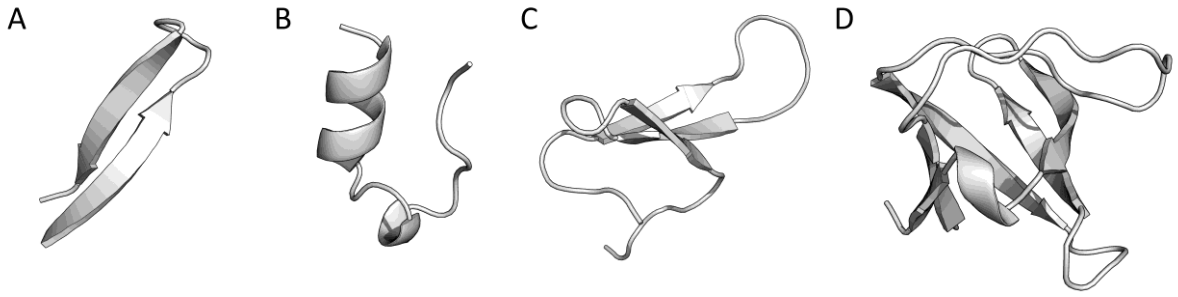


Figure 2. Ribbon representations of miniprotein scaffolds. A) Trpzip 4 (pdb: 1LE3), B) Trp-cage (pdb: 1L2Y), C) WW domain of human pin1 (pdb: 1I6C), D) SH3 domain of c-Crk (pdb: 1CKB).

By engineering one of these scaffolds into a cell adhesion miniprotein, hydrogels can be functionalized with bioactivity without denaturation. First steps towards this goal can be seen in the design of miniprotein-based hydrogels. One example can be found in the β-roll of MIS-38 that was isolated and redesigned by Dooley et al. The RTX motif folds into a β-loop-β structure upon binding Ca(II). The researchers mutated the β-strands to increase aggregation and thereby created a hydrogel that can be easily assembled and disassembled by controlling the Ca(II) concentration. [42] In a similar vein, Nguyen et al. formed a stress-relaxing hydrogel by mutating Trpzip 1. Mutations to control solubility and aggregation tendencies led to an antimicrobial miniprotein-based hydrogel that supports cell growth likely by charge adhesion. [43] Even with new materials based on miniproteins, cell adhesion controlled via a bioactive miniprotein has not yet been achieved. Integrin- or Cadherin-like interactions of large protein surfaces are difficult to miniaturize and cell adhesion motifs already mimic a decent activity towards integrin-like receptors. [34] Ca(II)-mediated carbohydrate recognition on the other hand has not been designed into a miniprotein to date. With a plethora of miniprotein scaffolds available and an increase in computer-aided design and engineering software, the question arises as to why this is the case.

2.1 Challenges in Carbohydrate-binding

2.2.1 Lectins – Carbohydrate-binding proteins

Carbohydrates or glycans form the most abundant class of biomolecules. They come in monomers, oligomers, polymers or linked to other biomolecules, then called glycoconjugates. The most common post-translational modification among proteins is the attachment of a glycan to a protein, creating a glycoprotein. It is therefore no surprise, that carbohydrate-mediated interactions play a vital role in many physiological processes and when disturbed, cause a number of different diseases.

The group of proteins that bind carbohydrates without enzymatically altering them, are called lectins. [44] They are present in all branches of life and show a variety of functions, from immune defense to fertility or cell adhesion and many more. [45] At least 48 different protein scaffolds have been elucidated among them with many different types including galactose-binding galectins or Ca(II)-dependent C-type lectins and D-type lectins. They are often part of larger protein complexes and require high specificity to compete with structurally similar carbohydrates and the main solvent, water. [28, 46]

In general, Carbohydrate-binding sites are generally grooves or shallow pockets at the surface or an oligomer interface. Hydrophobic interactions are essential for binding carbohydrates, especially with aromatic amino acids, that often dictate a specific ring. Most lectins are β-sheet proteins, more precisely β-sandwich proteins. However, mixtures of α-helical and β-sheet elements are also common, especially for C-type lectins. The β-sheet is therefore probably not a prerequisite, but rather a useful tool for forming hydrogen bonds, which are also formed between carbohydrates and lectins as a basis for high affinity. [28, 45]

2.2.2 D-type lectins

Several families of lectins rely on Ca(II)-mediated carbohydrate recognition, notably C-type lectins such as Selectins and the newly identified D-type lectins that bind the Dystroglycan (DG) complex. [3, 28] All entries of the latter feature laminin globular (LG) domains, β-sandwich domains with a Ca(II)-binding site on the rim of the sandwich that mediates the specific recognition of glucuronic acid- (GlcA) xylose (Xyl) disaccharide repeats [-3GlcAβ1,3Xylα1-], termed matriglycan. [3] This matriglycan polysaccharide is glycosylated onto α-DG, a receptor protein expressed in a variety of tissue, allowing for cell adhesion through Ca(II)-mediated carbohydrate recognition.

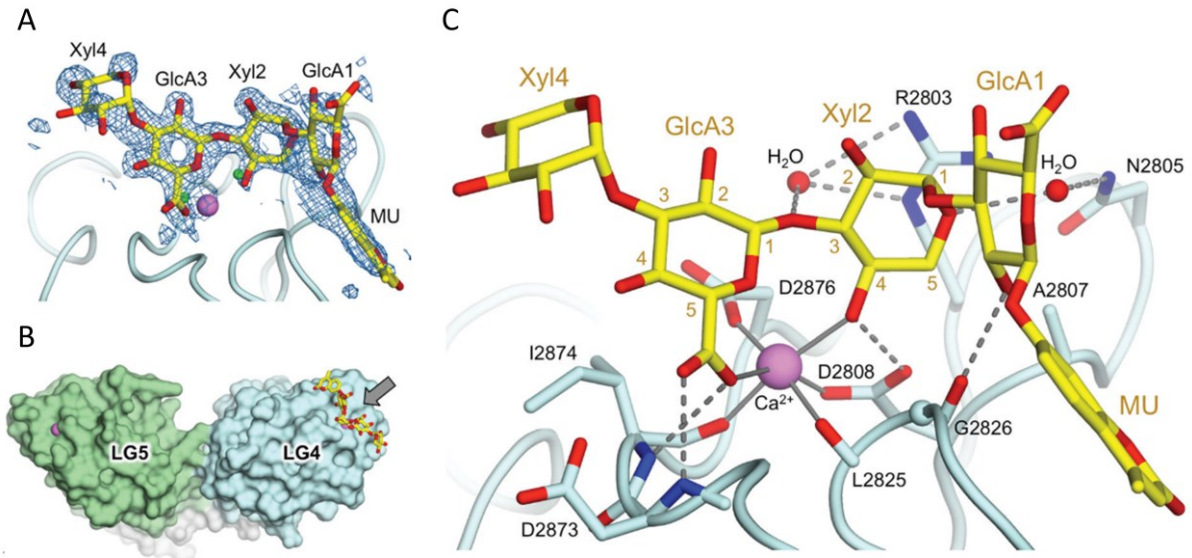


Figure 3. Crystal Structure of Laminin α2 LG4-5 bound to a GlcA-Xyl oligosaccharide. A) Unbiased electron density map contoured on the final model of the oligosaccharide ligand. Sugars beyond Xyl4 are not defined by the electron density and are presumed disordered (green points denote the position of two water molecules in the absence of the ligand); B) Surface representation of Laminin α2 LG4-5 (light blue: LG4, green: LG5, pink: Ca(II)) with the ligand shown in atomic detail and an arrow denoting the viewpoint of panel C; C) Atomic interactions between the GlcA-Xyl oligosaccharide (yellow) and Laminin α2 LG4 (light blue) with the ring carbon atoms of Xyl2 and GlcA3 numbered. Coordination bonds are indicated by solid lines, hydrogen bonds by dashed lines and water molecules are shown as spheres (red). (adapted from Briggs et al. Figure 3, page 16 [24])

Briggs et al. revealed the specific binding of laminin LG domains to matriglycan as a new type of Ca(II)-dependent lectin (see Figure 3). The existing group of C-type lectins coordinate single monosaccharides with one Ca(II) ion, while the newly identified D-type lectins coordinate a disaccharide. [24] As hypothesized by Briggs et al., laminin LG domains were not the exclusive D-type lectins. The same specificity was discovered for other LG domains of related ECM proteins, e.g. agrin, perlecan, neurexin, slit-2 and pikachurin. [3] NMR and X-ray studies revealed the direct coordination of a disaccharide repeat. A hydroxy group of Xyl and the ring oxygen of GlcA coordinate the Ca(II) cation. Additionally, GlcA binds the peptide backbone and Xyl stacks on an Arg residue. A longer saccharide chain leads to a higher affinity due to the increasing number of available binding sites and a rapid rebinding after dissociation. This

leads to an increase in K_d from 10^{-5} M for a disaccharide, to 10^{-7} M for a pentasaccharide and up to 10^{-9} M for the full-length glycan. [3, 24] With their full structure elucidated and a Ca(II)-mediated carbohydrate recognition with a small number of relevant amino acids, D-type lectins present themselves as good candidates for the design of cell adhesion miniproteins.

2.2.3 Lectin Engineering

The design or engineering of lectins is a new field. Nonetheless, there are a number of peptide lectins that either occur naturally or are derived from animal lectins. They can be as small as 6-residue strands that weakly bind to monosaccharides or 18-residue retrocyclins that bind with a higher affinity. Even peptides as large as 30 amino acids binding to oligosaccharides have been discovered. However, none of these peptides form a stable structure without requiring at least three to five disulfide bridges to maintain the active conformation. Engineering new peptide lectins is dominated by high-throughput screening methods of random sequence arrays. This methodology leans on automation and only produces variations of the existing peptide lectins with varying specificity for similar-to-native targets. [28, 45]

Larger lectins are generally engineered from existing lectins in a similar manner to peptides. To create new lectins, point mutations are used to change the preferred targets of lectins or deactivate the enzyme activity of glycosidases, turning them into lectins. Another approach is to dismantle lectin complexes that target multiple different carbohydrates to create lectins specific for a lower number of or a single ligand. All of these approaches feature a vast number of examples but a miniprotein lectin has yet to be published. [28, 45]

All known lectins are making use of the naturally occurring lectin scaffolds. [28] The same holds true for carbohydrate-binding modules in carbohydrate-active enzymes. [47] But, as mentioned in Chapter 2.2.1, there is no distinct structural feature that defines a lectin. Hirabayashi et al. therefore hypothesized that it should be possible to design a lectin on any given scaffold. In this work, I want to test this hypothesis by designing new Ca(II)-dependent lectins from mininprotein scaffolds for cell adhesion

2.3 Design of Ca(II)-binding Peptides and Proteins

2.3.1 Ca(II)-binding in proteins

Calcium is the third most prevalent metal in cells and the main cation used for signaling. With a high phosphate concentration in cells as their main buffer system, the Ca(II) concentration needs to be highly controlled, as calcium phosphate has a very low solubility in water. This requires strong binders and channels to uphold the low concentration in cells (10^{-7} M) and high concentration in the ECM (10^{-3} M). This makes it a good candidate for signaling, as it can be imported in short bursts and quickly sequestered to extracellular space. On top of that, Ca(II) is unique in its binding properties compared to Mg(II). As the bigger cation, Ca(II) has a higher polarizability and often strays from the octahedral coordination preferred by Mg(II). The ideal binding mode for Ca(II) is pentagonal bipyramidal, requiring 6-8 ligands. [48]

Protein	K_d [M]	Ca(II)-binding domain	source
CaM	$5.3 \cdot 10^{-7}$ - $1.18 \cdot 10^{-6}$	EF-hand 1-4	[49]
Laminin	$5 \cdot 10^{-6}$ - $3 \cdot 10^{-4}$	LG3-5	[50]
Slit-2	~ $1 \cdot 10^{-4}$	LG1	[51]
MIS-38	$9.1 \cdot 10^{-4}$	RTX	[42]

Table 1. Ca(II)-binding proteins with their dissociation constant K_d for Ca(II), the Ca(II)-binding domain and the literature source.

Even though there is an abundance of Ca(II)-binding proteins, the number of binding motifs is rather low. The most common and well-studied is shown in Figure 4A, the EF-hand. It has found diverse application in e.g. sensor, signal or buffer proteins. The helix-loop-helix motif contains a 12 amino acid sequence related to binding oriented octahedrally (see Figure 4B). With ligands +X (Asp, Asn) and -X (water molecule) on the x-axis, +Y (Asp, Asn) and -Y (carbonyl oxygen) on the y-axis, and +Z (Asp, Asn, Ser) and -Z (asp, Glu) on the z-axis and other amino acids denoted as o, the sequence reads (+X)o(+Y)o(+Z)o(-Y)o(-X)oo(-Z). This allows for Mg(II) binding in an octahedral coordination and a generally stronger Ca(II) binding with seven ligands in a pentagonal bipyramidal coordination. [9]

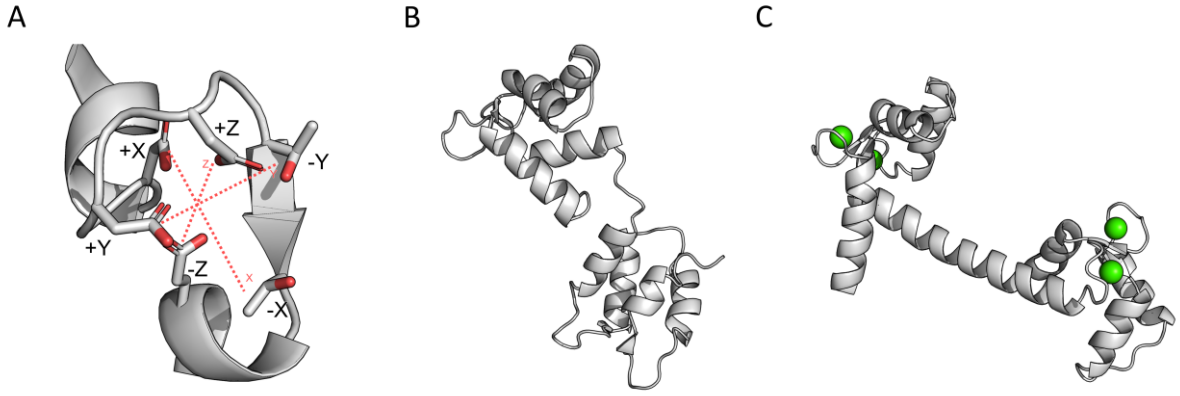


Figure 4. Ribbon representation of CaM EF-hand 1, apo- and holo-Calmodulin. A) EF-hand 1 of rat CaM (pdb: 3CLN, 14-33) with Ca(II) coordinating ligands along the X,,Y,Z axes (red) in positions +X, -X, +Y, -Y, +Z and -Z; B) Ribbon representation of apo-CaM (pdb: 1DMO); C) Ribbon representation of holo-CaM (pdb: 3CLN) with Ca(II) as spheres (green).

EF-hands are found in large protein complexes like myosin and small proteins like calmodulin (CaM, see Figure 4C). The domain usually comes in pairs and often with different affinities for Ca(II) within the same protein as can be seen for the four EF-hands of CaM. The dumbbell-shaped helical protein can be segmented into two subdomains, each containing two paired EF-hands at a lobe. It is highly conserved and has a sequence identity of 100 % among all eukaryotes. In the apo state, the N-terminal subdomain is in a closed conformation, while the C-terminal one is semiopen. In the holo state, both subdomains expose hydrophobic residues for target recognition and signaling. [52]

Ca(II) often include main chain carbonyl interactions. Figure 5 shows the RTX-motif of β-roll domains as found for MIS-38, follow a GGXGXDXUX sequence (X = any amino acid, U = aliphatic amino acid). With less charged carboxylic acid ligands, the binding affinity is considerably lower than for EF-hands but, like for CaM, induces a conformational change required for proper folding of the motif. [42, 53] In the case of D-type lectins, a similar mixture of main chain and side chain coordination can be seen. The Ca(II)-ion is bound by two Asp side chains and two main chain carbonyl oxygens. The incomplete coordination sphere is filled by two water molecules that are displaced upon carbohydrate recognition. When the GlcA-Xyl target is bound, Ca(II) is coordinated by two carbohydrate oxygens to complete the pentagonal bipyramidal coordination. [50, 51]

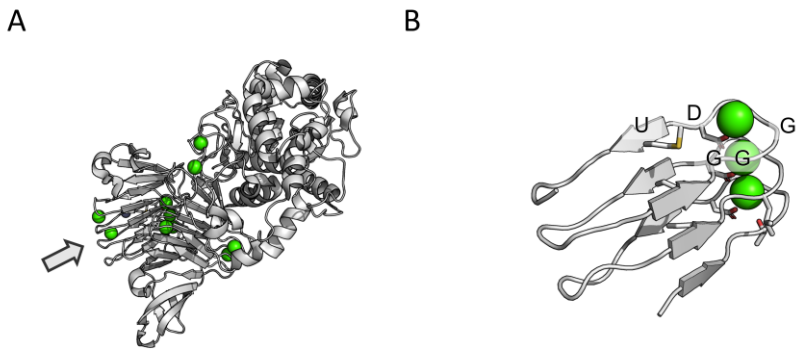


Figure 5. The RTX motif of MIS-38. A) Ribbon representation of extracellular lipase from *Pseudonomas sp.* MIS-38 (pdb: 2Z8X) with Ca(II) as spheres (green) and an arrow denoting the position of the RTX motifs; B) RTX motifs of MIS-38 (pdb: 2Z8X, 489-542) with Ca(II) as spheres (green) and labeled with the standard sequence of an RTX motif (GGXGXDXUX, X = any, U = aliphatic AA).

2.3.2 Design strategies for Ca(II)-binding peptides and proteins

To sidestep the natural Ca(II)-binding proteins that often contain multiple binding sites and fold cooperatively, three approaches to generate new Ca(II)-binders dominate the literature: grafting functional domains, computational design and rational design of new binding sites. [54]

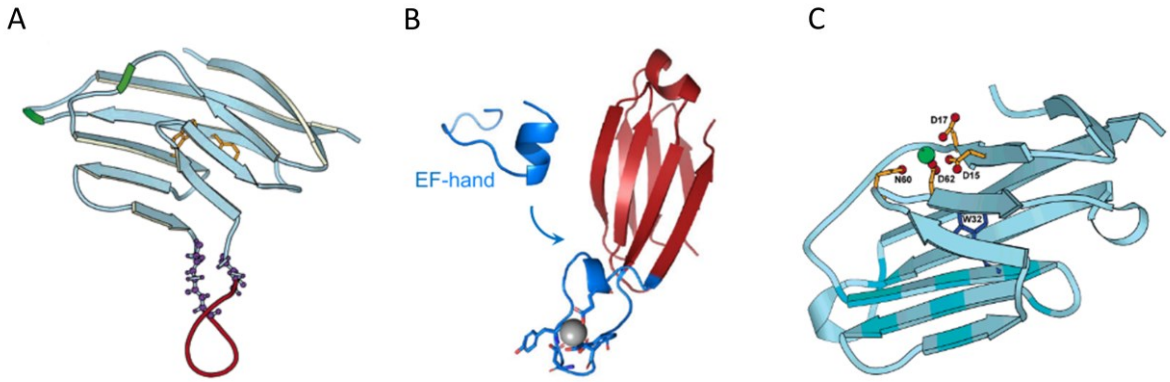


Figure 6. Literature examples of designed Ca(II)-binding proteins. A) CD2-CaM, an EF-hand of CaM (red) grafted into a loop of CD2 (blue); (adapted from Ye et al., Figure 1, page 3744 [6]) B) EF61-dIG8-CC, an EF-hand of CaM (blue) grafted into a loop of a *de novo* designed immunoglobulin-like protein (red); (adapted from Chidyausiku et al., Figure 5, page 7 [7]) C) CD2.Ca, an adhesion protein (CD2, blue) with a *de novo* designed Ca(II) binding site. (adapted from Yang et al., Figure 1, page 2087 [55])

The grafting approach has been applied extensively to study isolated EF-hands of CaM and other sensor proteins. [6, 56] Herein, the functional domain is inserted into the sequence of the scaffold protein e.g. replacing or elongating a loop. Any number of amino acids may be included as spacers before and after the grafted domain to allow for native-like folding. Ye et al. used this method to introduce the four different EF-hands of CaM to domain I of the cell adhesion protein CD2. Each EF-hand was inserted at position 52 in the loop region between two β-sheets of the β-sandwich scaffold protein. The native structure of the CD2 protein was unchanged by the insertion, as revealed by CD, Fluorescence and NMR spectroscopy. This allowed the study of the individual EF-hand's affinity for binding Ca(II). Their findings showed that Ca(II)-binding affinity strongly depends on the charge of its coordinating amino acids and weakens by up to two orders of magnitude at high ionic strength. [6] More recently, Chidyausiku et al. developed a *de novo* Immunoglobulin-like β-sheet fold and added function by grafting an EF-hand 3 of CaM into the loop region between two β-sheets using Rosetta

folding simulations. Even though they used the same approach and added spacers chosen by computational modelling, they achieved Ca(II) binding at least one order of magnitude weaker than their predecessors. [7] Siedlecka et al. used a similar approach to create a minimal EF-hand peptide. They isolated the functional domain and added Ala spacers to initiate an alpha-helical fold. The final peptide folded into a helical conformation but retained only partial functionality. [57] This suggests that the folding of the grafted domain can be greatly influenced by the scaffold protein. An incomplete folding of the EF-hand motif leads to a lowered affinity. Nonetheless, grafting domains has proven to be a sensible approach to designing Ca(II)-binding proteins. Based on these findings, domain grafting presents a promising approach for the design of Ca(II)-binding miniproteins.

Computational design of proteins and peptides relies on a modelling suite to generate new binding sites on a known independently-folding scaffolds. To date, only *de novo* binding sites based on an idealized, pentagonal bipyramidal Ca(II) geometry have been published. [55] However, with the rise of computational modeling suites like Rosetta, models based on a site found in an active Ca(II)-binding protein would also be possible. For this, the binding site is defined by the geometry of the binding amino acids and their ligand. The peptide backbone is assumed to be unchanged by the mutations that create the active site. The modeling suite inserts the desired site in a suitable scaffold and can also be used to optimize the geometry for efficient binding or enzyme activity. [4] However, designing new sites on large proteins has proven to be more successful than designing on peptides or miniproteins. The Modeling suite uses a multitude of approximations and omissions to generate its energy function that can be detrimental when simulating smaller scaffolds. [58]

Protein	K_d [M]	Ca(II)-binding domain	source
CD2-CaM	$1.25 \pm 0.15 \cdot 10^{-4}$	EF-hand 1	[6]
CD2-CaM	$4.00 \pm 0.20 \cdot 10^{-4}$	EF-hand 2	[6]
CD2-CaM	$4.35 \pm 0.30 \cdot 10^{-4}$	EF-hand 3	[6]
CD2-CaM	$1.06 \pm 0.20 \cdot 10^{-3}$	EF-hand 4	[6]
EF61_dIG8-CC	$> 10^{-3}$	EF-hand 3	[7]
CD2.Ca	$1.40 \pm 0.40 \cdot 10^{-3}$	*de novo*	[55]
CD2.trigger	$9.0 \pm 0.25 \cdot 10^{-5}$	*de novo*	[5]
Leu beta roll	$8.7 \pm 0.2 \cdot 10^{-4}$	β-roll motif	[42]

Table 2. Designed Ca(II)-binding proteins with their dissociation constant K_d for Ca(II), the Ca(II)-binding domain and the literature source.

Protein design does not require a heavy emphasis on a modelling suite. Li et al. have shown in their design of a conformation switchable Ca(II)-binding protein that using rational design supported by computational modeling can lead to equally active proteins. Their approach was based on three considerations from past works in the field: an unaltered peptide backbone, at least four charged residues in the binding site for high binding affinity, and binding sites located in regions of natively high electrostatic potential. The resulting protein CD2.trigger bound Ca(II) with a K_d in the range of an isolated EF-hand or laminin LG domains and changed its conformation from a largely disordered structure to the native fold upon binding. [5] Using similar considerations, one should be able to generate Ca(II)-binding miniproteins from any

robust scaffold. Both rational and computational design approaches can lead to active binders with high control over the composition of the active site. Unlike the grafting approach, a successfully folded domain does not guarantee an active binding site, as structure predictions rely on models that might not sufficiently reflect real behavior in solution. [4, 58]

3 Aim

In this doctoral thesis it is aimed to design miniproteins with a folded three-dimensional structure for Ca(II)-mediated carbohydrate recognition as a cell adhesion additive for the functionalization of Tissue Engineering matrices.

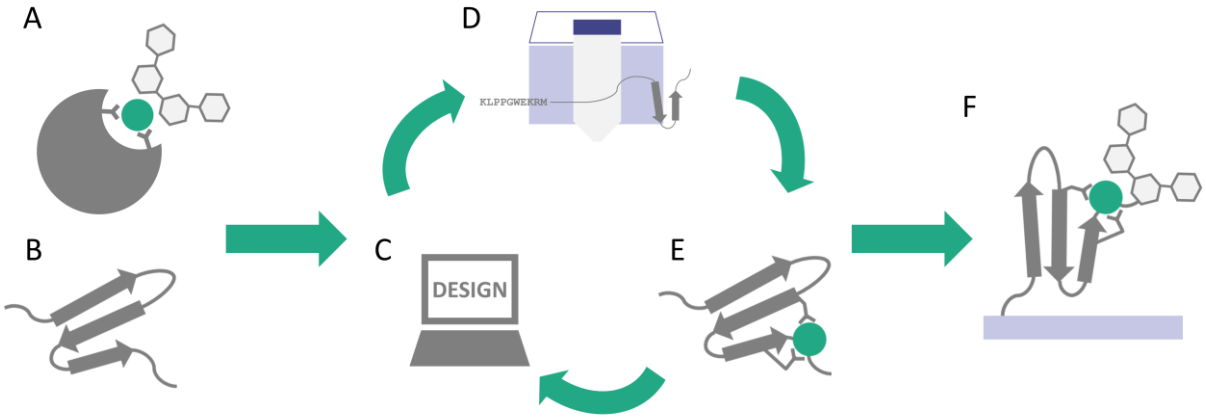


Figure 7. Scheme for the Design of Adhesion miniproteins for Tissue Engineering. A) D-type lectin binding matriglycan; B) Miniprotein scaffold; C) *In-silico* design; D) Solid phase peptide synthesis; E) Ca(II)-binding adhesion miniprotein; F) Adhesion miniprotein immobilized on a hydrogel.

Using literature-known approaches to peptide and protein design, such as computational design with Rosetta, [4] computer-aided rational design, [5] and functional domain grafting, [6] Ca(II)-binding miniproteins are designed on the basis of independently folding β-sheet scaffolds. Promising models are synthesized by solid phase peptide synthesis, purified by high performance liquid chromatography and identified by matrix-assisted laser-disorption ionization time-of-flight mass spectrometry. Successfully synthesized models are characterized for thermostability, structure and function and used to improve the computational models. The design aims to lead to a stable and functional miniprotein that binds Ca(II) in the first instance and can be optimized to recognize carbohydrates. An ideal candidate is thermostable, folded, recognizes matriglycan – the native ligand of D-type lectins – and remains functional at 37 °C. [3]

Additionally, minienzymes based on the c-Crk SH3 domain scaffold and the active site of Cu,Zn-Superoxide Dismutase are designed to test the applicability of computational design using Rosetta for miniprotein scaffolds. The models are synthesized, purified and characterized regarding thermostability, structure and function. An ideal candidate strongly binds Cu(II) – the metal ion responsible for Superoxide Dismutase activity – and shows significant enzyme activity. [59]

4. Results and Discussion

4.1 Computational Design of Ca(II)-binding Miniproteins

Computational design describes the automated search for protein or peptide sequences that fold into a desired three-dimensional structure. This process is guided by an energy function based on a mixture of physical terms, such as van-der-Waals forces and electrostatic interactions, and knowledge-based terms, such as sidechain rotamer libraries and geometric parameters. [60-62] The most commonly used modelling suite to perform computational design at the time this work started was *Rosetta3*. The software contains an extensive Enzyme Design application for *de novo* design or engineering of active enzymes. The design process can be separated into four stages: (1) choice and minimalization of active site, (2) identification of a set of sites on a given scaffold to realize the active site, (3) optimization of surrounding residues for stabilizing transition state interactions, and (4) evaluation and ranking of results. [4]

The first stage describes the choice of a natural model, a blueprint active site the design is based on. One needed to choose either a crystal structure of a known enzyme for a natural reaction or a computationally derived transition state for a reaction that does not occur in nature. [63] From this, a minimalist description of the active site or transition state consisting of catalytic residues, their distance, angles and dihedrals with respect to the target is compiled. [64] The second stage relies on the *Rosetta3 Match* application that searches the peptide backbone of a chosen scaffold one by one for positions where the catalytic residues can be mutated in the defined orientation. Once a set of positions is identified where the complete minimalist site can be realized without clashing with the backbone, a model is saved. [65] *Rosetta3 Design* mutates the surrounding amino acids to ensure a tight packing and optimize transition state binding affinity and catalysis in the third stage. [4] The results are evaluated by the *Rosetta3 Relax* or *Score* algorithms in the last stage. *Relax* mutates chosen residues and optimizes the geometry by slightly changing the amino acid sidechain geometries for a better Rosetta energy value. The *Score* algorithm is used to apply different Rosetta energy functions, e.g. to find aspartimide-promoting sequences. The resulting energy values are then used to find the best models for synthesis. [4, 62]

4.1.1 Slam series

Design Strategy

Ca(II)-mediated carbohydrate recognition for cell adhesion is found in C-type or D-type lectins. The key difference between these is in the number of monosaccharide units coordinated by the Ca(II) ion. C-type lectins coordinate monosaccharide units, while the laminin globular (LG) domains found in D-type lectins coordinate a disaccharide unit. [3, 24] LG domains have a sequence length of around 180 amino acids and can be found in a number of members of the laminin family e.g. Laminin, Perlecan or Neurexin. The C-termini of Laminin α1 and α2 each contain tandem repeats of five LG domains. Of these five, only LG4 and LG5 serve as cell adhesion domains with their binding site for Heparin, Sulphatides or α-Dystroglycan. As can be seen in Figure 8, they fold into curved β-sandwich structures built from two sets of seven antiparallel β-sheets. Each β-sandwich has a Ca(II) binding site on one rim of the sheets responsible for cell adhesion. [66, 67] In LG4, Ca(II) is bound by the side chains of Asp 2808 and Asp 2876, and the main chain oxygens of Leu 2825 and Ile 2874 in an incomplete octahedral coordination sphere with a micromolar K_d-value. [50] This Ca(II) ion mediates the binding to a heteropolysaccharide [-glucuronic acid-β1,3-xylose-α1,3-]$_n$ (GlcA-Xyl) of the α-Dystroglycan receptor for cell adhesion. [24]

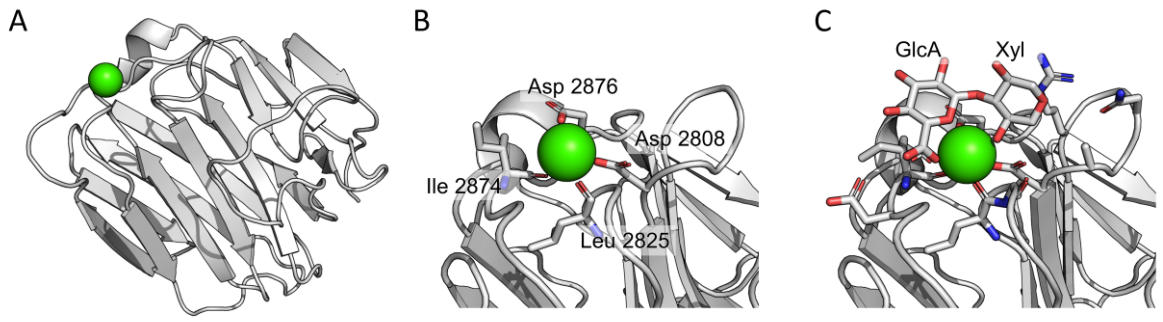


Figure 8. Ribbon representation of Laminin α2 LG4 (pdb: 5IK5). A) Complete LG4 domain with Ca(II) marked in green; B) Ca(II) binding site with coordinating amino acids; C) Ca(II) binding site with the disaccharide ligand (GlcA-Xyl) of α-Dystroglycan. [24]

The scaffold for peptide design was chosen from among the known independently folding β-sheet peptides. The *N*-terminal SH3 domain of c-Crk, shown in Figure 9, (SH3, pdb: 1CKB) was chosen for its structural similarity as a β-barrel domain. [68] Native SH3 domains contain their hydrophobic core on the inside of the β-barrel with amino acids from the overarching highly charged RT Loop and each of the five β-sheets that form the barrel structure. They bind proline-rich sequences on the cleft between the barrel and the RT Loop. With a high melting temperature (T_m) of 67 °C, the 58 amino acid sequence of the c-Crk SH3 domain was known to be thermostable, chemically accessible and studied extensively in terms of folding kinetics. [69]

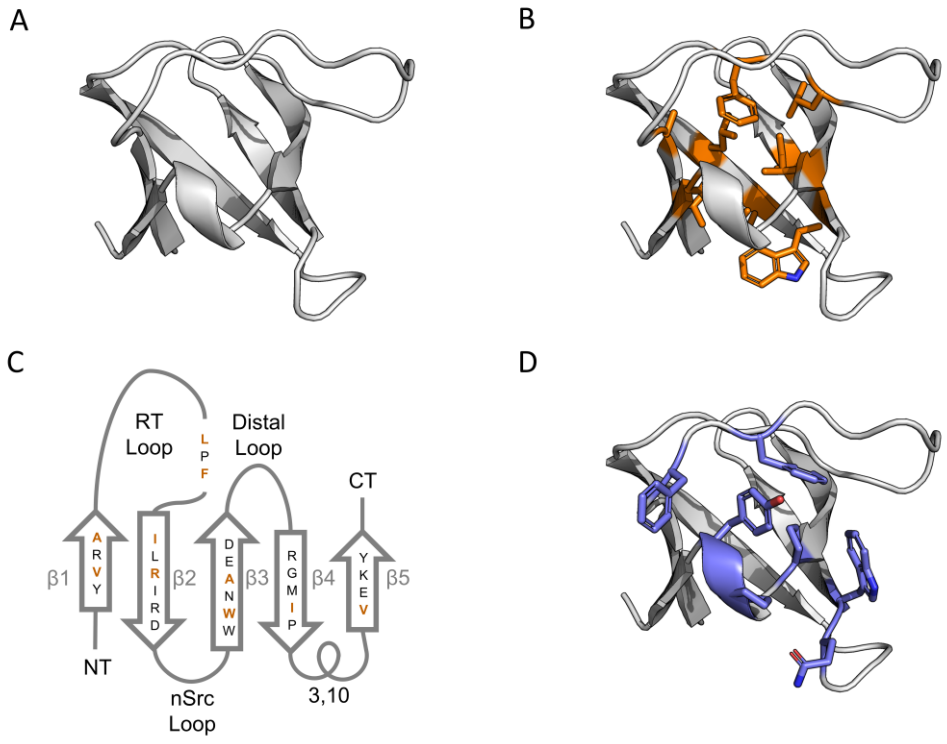
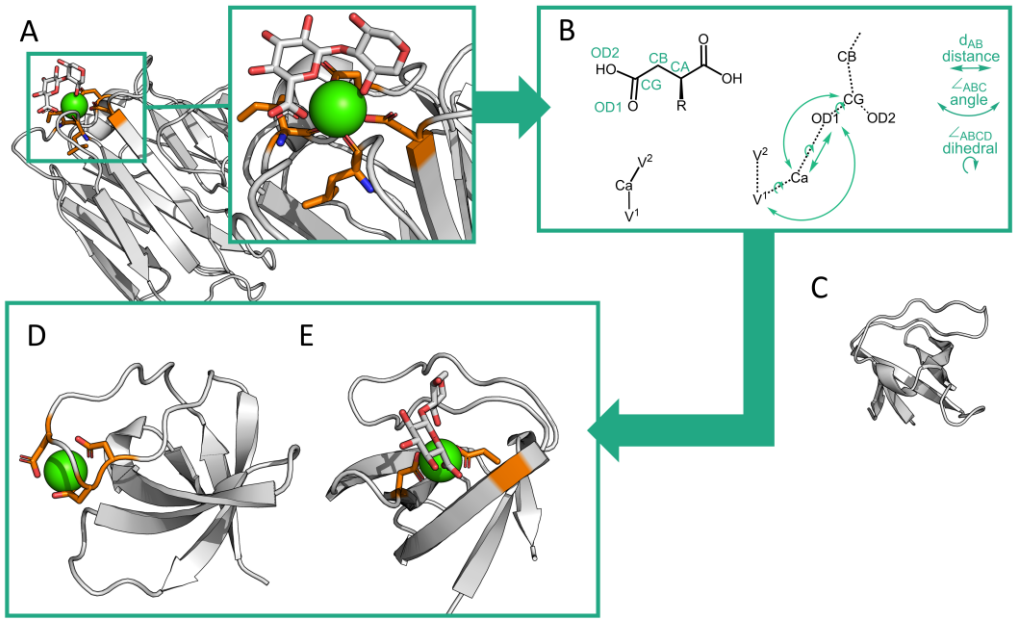


Figure 9. N-terminal SH3 domain of c-Crk (pdb: 1CKB). A) Ribbon representation of the SH3 domain; B) SH3 domain with the amino acids of the hydrophobic core marked in orange; C) schematic representation of the secondary structure motifs with *N*-terminus (NT), β-sheets 1-5 (β), loops, 3_{10}-helix (3,10) and *C*-terminus (CT) and all amino acids of the hydrophobic core marked in orange; D) SH3 domain with the amino acids conserved for polyproline binding marked in blue. [26]

The *Rosetta3* modelling suite was chosen for the computational design of the SH3-Laminin (Slam) miniproteins (see Figure 10). A *PyMol* script read out six parameters for each Ca(II)-

coordinating Asp residue to describe the minimalist active site. Possible positions on the SH3 domain scaffold where then identified by the *Match* algorithm. [4] The positions conserved for the hydrophobic core, and the first and last four positions were omitted. The order in which the different conformations were computed, had no impact on the outcome, but on the total computation time. The higher the flexibility given for each parameter, the longer the computation time. Starting with the most restricted residue, therefore gave results the fastest. *Rosetta3 Design* optimized the binding site for a tight packing. The lowest-scoring model, with the binding site positioned at the highly charged RT Loop, was named Slam1. The synthesis of the sequence was unsuccessful under standard conditions. The model contained a highly aspartimide-promoting Asp repeat that made purification by HPLC not feasible. The sequence was therefore discarded and a new design was prepared.

For the second design, a larger portion of the LG4 active site was chosen. New parameters were taken from the crystal structure to describe the interactions of Ca(II) with all four coordinating side and main chain atoms. *Rosetta3 Match* and *Design* were called to install and optimize a binding site based on the interaction of Asp 2808, Asp 2876 and Leu 2825 or Ile 2874. The main chain coordinating residues were not limited to Leu and Ile, but extended to all non-polar and aromatic residues. The lowest-scoring models all omitted the geometry of Ile 2874. For these models, a second round of *Design* without geometric constraints was called. The lowest-scoring models positioned the binding site at the rim of the β-barrel, reminiscent of the position on the LG4 domain. Slam2 included no mutation to install the binding Asp residues. Instead, the amino acids surrounding two preexisting Asp residues were mutated to accommodate the Ca(II) ion with the disaccharide ligand in the binding site.



F

c-Crk SH3	AE-YVRA-LFDFNGNDEEDLPFKKGD-ILRIRD-KPEEQ-WWNAE-DSEGK-RGMIP-VPY-VEKY
Slam1	AE-YVRA-LFDFNG**T**D**DD**DLPFKKG**E**-ILRIRD-KPEEQ-WWNAE-DSEGK-RG**L**IP-VPY-VEKYG
Slam2	AE-YVRA-LFDFNGNDEEDLPF**V**KGD-ILRIRD-KPEEQ-WWNAE-D**VL**GK-RG**L**IP-VPY-VEKYG
Slam3	AE-YVRA-LFD**YT**G**TVTL**DL**D**FKKG**E**-IL**T**IRD-KPEEQ-WWNAE-DSEGK-RG**L**IP-VPY-VEKY

Figure 10. Computational Design of the Slam series with (A) the crystal structure of Laminin α2 LG4 (pdb: 5IK5) and a zoom of the active center; B) geometric parameters of the Ca(II) center; [4] C) the crystal structure of the c-Crk SH3 domain (pdb: 1CKB); D) Rosetta model of Slam1; E) Rosetta model of Slam2 and (F) sequences of the Slam series with hydrophobic core positions in red and mutations in bold. [4, 26]

Synthesis and Purification

The Slam2 sequence was synthesized by SPPS under standard conditions. The resin mass was halved to compensate for the high swelling of the peptide resin due to the long sequence. The synthesis quality was high, especially for a 58 amino acid sequence. Purification by gradient HPLC (20-60 % B in A) showed a clear main peak containing the desired product. The aspartimide-promoting Asp-Ser sequence of the native SH3 domain was mutated to Asp-Val for binding. In all other Asp positions, suppression of aspartimide formation by the addition of formic acid to the deprotection solution was successful.

Structure and Thermostability

The miniprotein was analyzed by circular dichroism (CD) spectroscopy to study the folding and thermal stability of Slam2. As seen in Figure 11A, the native SH3 domain displays characteristic spectra with two negative bands at ~210 nm and ~230 nm, and a positive band at ~220 nm. The 210 nm band stems from the unordered loops and termini of the domain widened by the 215 nm band of the β-sheets that can appear as a shoulder. [70] Interactions between aromatic residues, especially Trp, can cause exciton couplings with intense CD bands. Typically, one can find a positive band at around 230 nm and a negative band at a lower wavelength between 208-220 nm. In the case of the SH3 domain, the positive 230 nm band has been shown to be a direct result of the interactions of Trp 37 and diminishes upon mutation to Phe. The positive 220 nm band likely results from the high aromatic content and the domain's β-turns, and is not an artifact of two colliding minima. [71]

As seen in Figure 11B, the Slam2 miniprotein showed a similar-to-native folding. Like the wild type SH3 domain, Slam2 displayed a negative band at 210 nm from the β-barrel structure, a positive band at 220 nm, and a second negative band from the exciton coupling of Trp at 228 nm. [72] The thermal denaturation experiments in Figure 11C revealed a melting temperature (T_m) of 50 ± 0.5 °C. [73] The mutations of charged amino acids to hydrophobic ones likely lowered the T_m and increased hydrophobicity. Spectra in the presence or absence of Ca(II) showed no discernible difference and therefore no conformational change upon Ca(II) addition. This was unexpected with the binding site positioned in the stable β-barrel. A comparison of heating and cooling curves in Figure 11D showed a full overlap above 60 °C and a delayed refolding in the cooling curve between 60 °C and 0 °C. The intensity is recovered at 0 °C, making the fold likely fully reversible for a slower cooling.

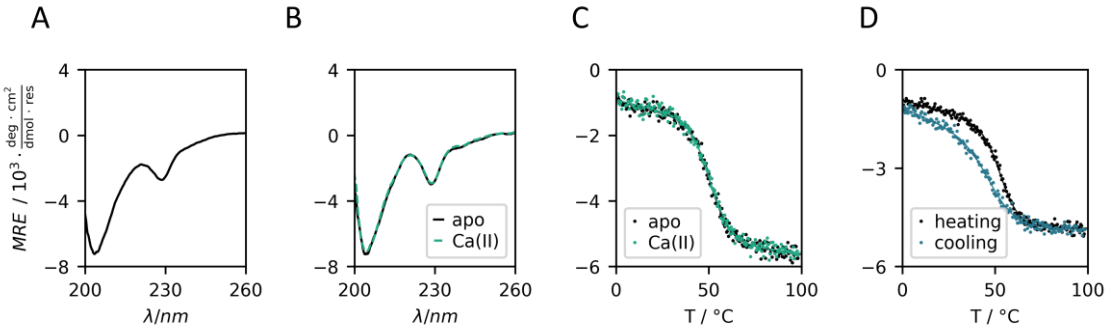


Figure 11. CD spectra and thermal denaturation of Slam2. A) CD spectra at 20 °C of c-Crk SH3 in pH 7.0 buffer containing 10 mM sodium phosphate measured by Pham et al. (page 428, Figure 3F); [74] B) CD spectra at 20 °C, (C) thermal denaturation curves measured from 0 °C to 98 °C in the absence (black) or presence (green) of 2 mM $CaCl_2$ and (D) thermal denaturation curves from 0 °C to 98 °C of while heating (black) and cooling (teal) of Slam2 in pH 7.2 buffer containing 20 mM *N*-ethylmorpholine, 40 mM acetic acid and 100 mM NaCl.

Competitive Binding Assay

To characterize the Ca(II) binding of the designed miniproteins, a reliable and reproducible assay was required. Direct measurements can be used to find the dissociation constant (K_d) if

the Ca(II)-free and -bound forms have different UV/Vis, CD or fluorescence spectra. As discussed above, the CD spectra of our stable designs showed no difference between the Ca(II)-free or -bound forms. Intrinsic Trp fluorescence showed no discernible difference for binding either. According to Linse, Ca(II) binding experiments show the highest accuracy when protein concentration and K_d are about equal. This limits the range where direct measurements are applicable to $K_d > 1$ µM, because of the difficulty to achieve buffers with a Ca(II) concentration lower than 0.5 - 1 µM. A competitive binding experiment, using a Ca(II) chelating dye, would lead to a precise measurement of the K_d even for lower values. [75]

For the competitive binding assay according to a protocol by Linse, a Ca(II)-free buffer is required and using purified water and clean glassware was not sufficient. [75] Ca(II) was found in the buffer salt, pH electrode, glassware and cuvette. Glassware was therefore rinsed with 5 mM EDTA solution and water before use. Ca(II) ions within the buffer were removed with a Chelex® 100 ion exchange resin. The resin has a higher affinity for bivalent than monovalent cations and comes in a sodium-bound form. [76] The buffer was decalcified by different methods to find the most efficient ones: dialysis, column method, batch method. For dialysis, the resin was filled in a dialysis tube and set in the buffer solution. After shaking overnight, the tube was removed. For the column method, the resin was packed in a column and rinsed with buffer solution. The first few column volumes were discarded for equilibration and the remaining ones were collected. For the batch method, the resin was poured in the buffer solution and filtered off after shaking a few hours. To determine the concentration of Ca(II) remaining in the buffer solution, the chelator dye Br_2-BAPTA (B2B) was used. The absorbance was measured for a test solution (A_1), after addition of EDTA as a zero point (A_2) and after the addition of Ca(II) as a saturation point (A_3). According to Equation 1, the Ca(II) concentration was found:

$$c(Ca) = \frac{A_2 - A_1}{A_2 - A_3}$$

Equation 1. Ca(II) concentration after decalcification c(Ca) with absorbance A_1 of the test solution, A_2 after addition of 5 µL EDTA (0.1 M) and A_3 after addition of 10 µL $CaCl_2$ (0.5 M).

Dialysis yielded the highest remaining Ca(II) concentrations with values between 1.3 µM and 3.9 µM and a change in pH of up to 0.5. Extending the duration to one week decreased the value below 1 µM and increased the pH change. The column method gave the quickest results with values between 0.2 µM and 1.2 µM and a pH change of up to 1.5. The more buffer was discarded upon equilibration, the lower the final Ca(II) concentration. The batch method resulted in Ca(II) concentrations between 0.35 µM and 1.9 µM and a change in pH of up to 2. All three methods could be optimized to give a value below 1 µM with a change in pH that called for readjustments. The batch method was chosen as the most efficient one, because it saved time compared to dialysis and material compared to the column method. The pH change was corrected afterwards by measuring small buffer samples and adjusting the main solution without contact to the pH electrode.

For the binding experiments, a small volume of concentrated Ca(II) solution was added to a Ca(II)-free buffered solution of chelator dye and miniprotein until saturation was reached. The results were analyzed using a python script (S1). [77] Figure 12A shows binding curves for different miniprotein K_d-values. For a miniprotein that binds the target stronger than the chelator ($K_d = 1.4 \cdot 10^{-6}$ M), one would expect a sigmoidal curve as seen for K_d = 10 nM. For equal binding strength as seen for K_d = 1.4 µM, the saturation curve is stretched as if the chelator concentration was doubled. For a lower binding strength as seen for K_d = 100 µM, the curve shape near saturation changes as for a secondary binding site. In the case of no competition, a non-binding miniprotein ($K_d = \infty$), the curve is identical to the miniprotein-free titration. [75] Figure 12B shows a titration of Ca(II) against B2B without competition overlaid

with a simulated non-binding curve. Figure 12C shows the competition experiment of Slam2 and B2B for Ca(II) overlaid with the same non-binding simulation. The analysis gave better results, the higher the guess K_d values were and the Slam2 data set overlaid well with both competition-free chelator and the simulated curve. The miniprotein was not competing with the dye and therefore deemed non-binding.

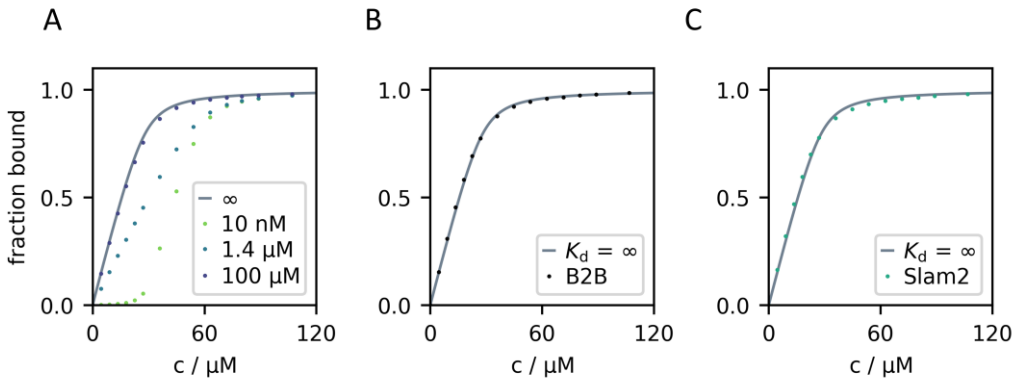


Figure 12. Competitive binding assay of Slam2. All measurements in pH 7.2 buffer containing 20 mM *N*-ethylmorpholine, 40 mM acetic acid and 100 mM NaCl. A) Simulated data for a competition experiment with Br_2-BAPTA (K_d = 1.4 · 10^{-6} M) and an unknown Ca(II) binder with a 10 nM, 1.4 µM or 100 µM K_d and no competition (K_d = ∞); B) titration of Ca(II) to a 28.52 µM Br_2-BAPTA solution (points) and simulated data for Br_2-BAPTA without competition (line); C) titration of Ca(II) to a 28.52 µM Slam2 solution with 28.52 µM Br_2-BAPTA (points) and simulated data for Br_2-BAPTA without competition (line). [75, 77]

Design of Slam3

As Slam2 design was based on improving the geometry of native Asp residues for Ca(II) binding, it is likely that the intended geometry was not achieved experimentally. The design strategy was adjusted to install a new binding site without disturbing the folding or chemical availability. For the new design, the parameters were taken for an interaction of all four coordinating amino acids – either Leu 2825 or Ile 2874 – bound to not only Ca(II) but also GlcA-Xyl. The *Rosetta3 Match* application was called to install the ligand-bound site. The GlcA-Xyl was then translated to a carbohydrate ligand in the *Rosetta3* format. *Design* was called to optimize the Ca(II) binding site with GlcA-Xyl as a separate flexible ligand. The active site was then reintroduced to be included in the active site for a second round of *Design*. The lowest-scoring model installed a binding site in the RT Loop at the same position as a naturally occurring Na(I) binding site of the human Fyn SH3 domain. [78]

Due to the poor results for the other Slam models and the more time-consuming synthesis and purification of the large SH3 domain-based models compared to other scaffolds in this work, Slam3 was not synthesized in the time frame of this work.

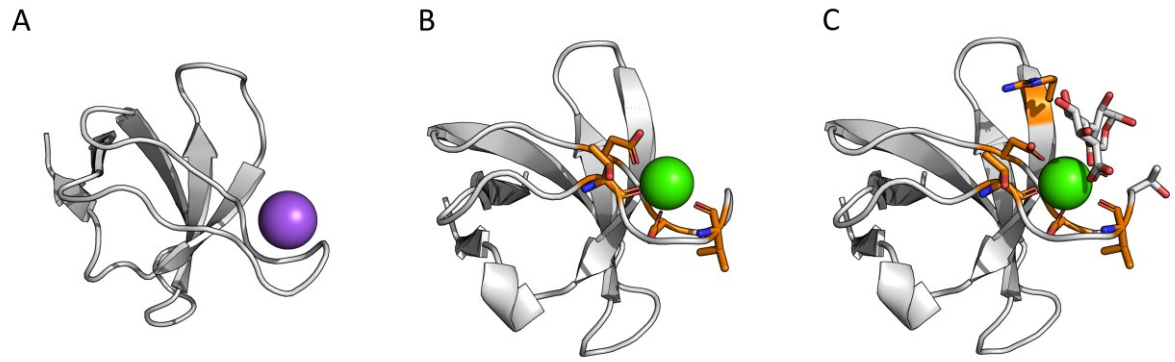


Figure 13. Ribbon representation of Fyn SH3 and Slam3. A) Fyn SH3 domain (pdb: 3UA6) with bound Na(I) colored purple; [78] B) Slam3 with the Ca(II) binding amino acids marked in orange; C) Slam3 with the GlcA-Xyl ligand in grey and interacting amino acids shown.

4.1.2 WWcalm series

Design Strategy

As designed miniproteins tend to bind their targets with a weaker K_d, a tight-binding natural model – mammalian Calmodulin (CaM, pdb: 3CLN) – was chosen. CaM can be divided into two α-helical domains, each of which includes two paired EF-hand motifs. These helix-loop-helix structures bind Ca(II) in a pentagonal-bipyramidal geometry. Ca(II) is coordinated by side chain oxygens of residues Asp 1, Asp 3, Asx 5 and Glu 12 of the 12 amino acid sequence. When no ligand is present, the two domains are connected by a partially unwound extended α-helix. The flexible apo-state has a noticeable propensity to be intrinsically disordered, allowing different conformational states for binding to other peptides. [79]

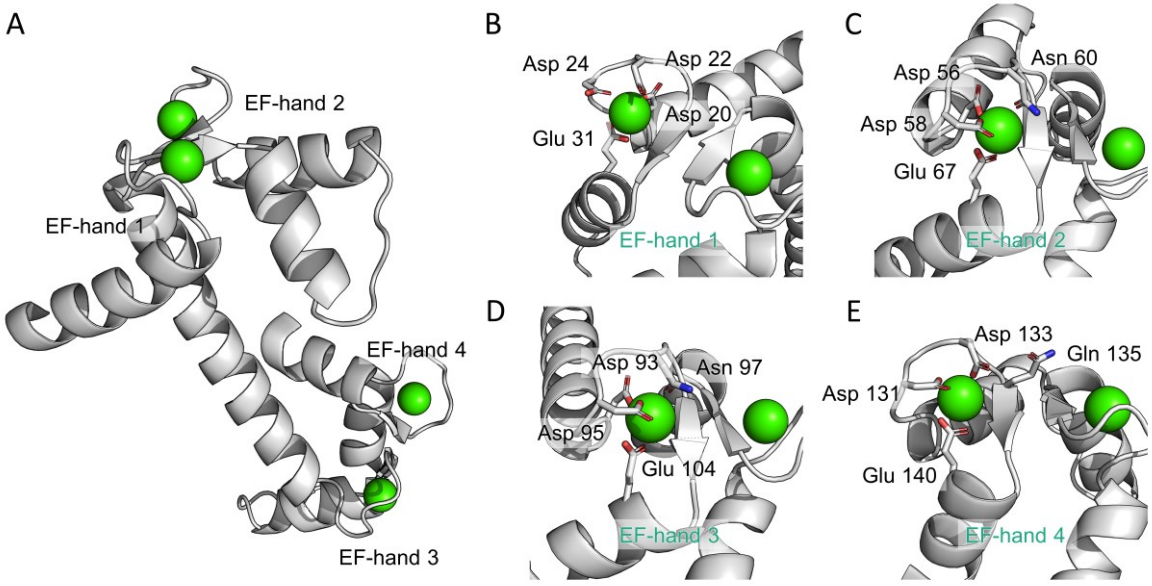


Figure 14. Ribbon presentation of mammalian calmodulin (pdb: 3CLN). A) Complete protein with Ca(II) bound to each EF-hand motif shown in green; B) EF-hand 1, C) EF-hand 2, D) EF-hand 3 and E) EF-hand 4 with binding amino acids shown. [79, 80]

A suitable scaffold for this Design was found in the human Pin1 WW domain (WW domain, pdb: 1I6C). WW domains are β-sheet motifs of 30 to 40 amino acids length and consist of a triple-stranded antiparallel β-sheet (Figure 15). The small hydrophobic core is formed by five amino acids: Leu 2, Pro 3, Trp 6, Tyr 19 and Pro 32 (for hPin1). Due to their high sequence

capacity, they have been used extensively for peptide design. The hPin1 WW domain in particular, was used to study the thermodynamics and folding kinetics of β-sheet structures. This resulted in an abundance of data on it, making it the ideal candidate for design. [27]

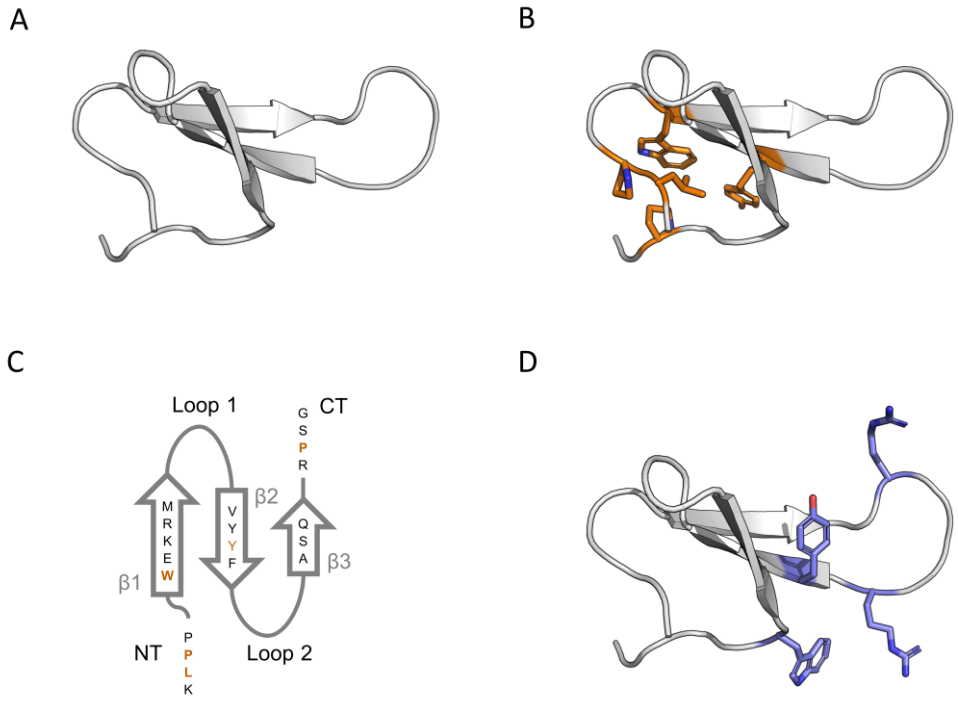


Figure 15. Human Pin1 WW domain (pdb: 1I6C). A) Ribbon representation of the WW domain; B) WW domain with the amino acids of the hydrophobic core marked in orange; C) schematic representation of the secondary structure motifs with *N*-terminus (NT), β-strands 1-3 (β), loops and *C*-terminus (CT), and all amino acids of the hydrophobic core marked in orange; D) WW domain with the amino acids conserved for polyproline binding marked in blue. [27]

The miniproteins designed from a WW domain scaffold with a Calmodulin site, were termed WWcalm. To achieve a micromolar binding site on the WW domain, a partial EF-hand was deemed sufficient, based on two of the four coordinating residues. The binding site was installed with the help of the *Rosetta3* modelling suite. I analyzed the geometry of each Ca(II) binding residue of CaM Loop 1 (Asp 20, Asp 22, Asp 24, Glu 31) in *PyMol* and grouped them in all six non-redundant permutations for *Rosetta3 Match*. The algorithm searched the peptide backbone for suitable positions, where both residues could be accommodated in the expected geometry. The most promising outputs were situated on two adjacent β-strands. *Rosetta3 Design* was then called to complete the coordination sphere of Ca(II) and improve the packing on the WW domain. The resulting models were filtered using *Rosetta3 Relax*. The final sequences featured a binding site on β-strands 2 (Asp 20) and 3 (Glu 27) in WWcalm1 (see Figure 16A) and β-strands 1 (Asp 9) and 2 (Asp 20) in WWcalm2 (Figure 16B).

Based on these binding sites, two more models were designed intuitively. In WWcalm3 (Figure 16C), the binding site on β-strands 1 (Asp 9) and 2 (Thr 20) was extended by an additional coordinating residue in Loop 2 (Glu 22). In WWcalm4 (Figure 16D), the binding site on β-strands 2 (Asp 20) and 3 (Asn 27) was extended by another residue on β-Strand 3 (Asp 25).

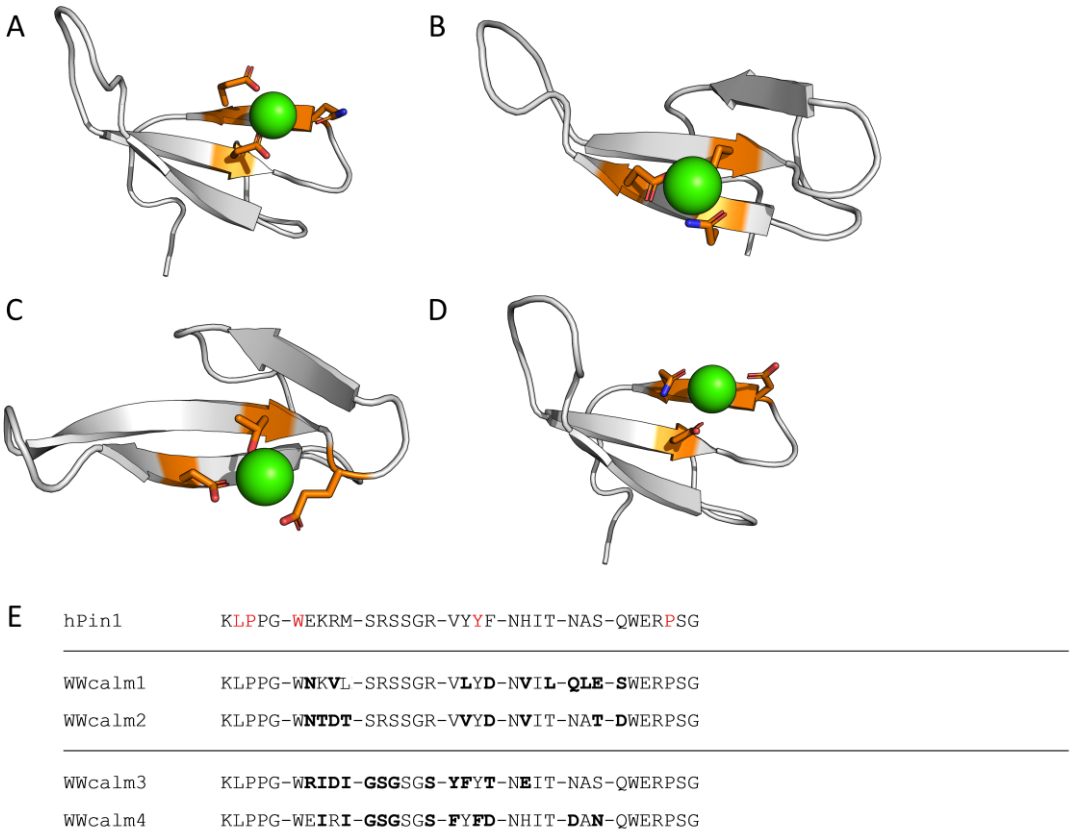


Figure 16. Models and Sequences of the WWcalm series. Models of (A) the computationally designed WWcalm1 and (B) WWcalm2, [4] and (C) intuitively designed WWcalm3 and (D) WWcalm4; E) sequences of the WWcalm series miniproteins and the hPin1 WW domain. Hydrophobic core positions in red, mutations in bold and pseudo-prolines boxed. [27]

Synthesis and Purification

The four sequences were synthesized on H-Gly-HMPB ChemMatrix™ resin with an automated microwave-assisted peptide synthesizer. The desired products were the main products of all syntheses. The synthesis quality of WWcalm3 was considerably lower than for the other sequences (Figure 17C). The masses found for the side products hinted at several deletion sequences during or after the synthesis of Loop 1.

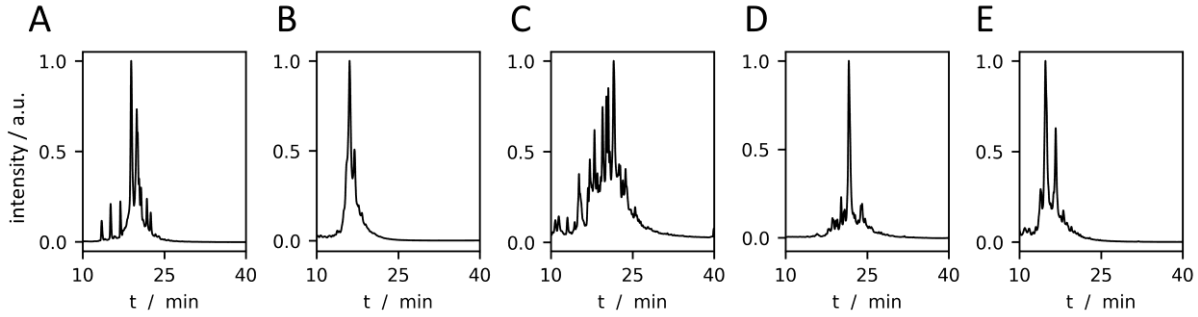


Figure 17. Reverse-Phase HPLC chromatograms of crude WWcalm series miniproteins. (A) WWcalm1 in 20-80 % B in A, (B) WWcalm2 in 20-80 % B in A, (C) WWcalm3 in 20-50 % B in A, (D) WWcalm3 in 20-50 % B in A synthesized with pseudo-proline GS in Loop 1 and (E) WWcalm4 in 25-65 % B in A.

Deletions can occur due to poor solvation of the peptide strand, intermolecular hydrophobic aggregation of the protected peptide chains or the association of secondary structure

elements. One way to reduce these unwanted effects was the incorporation of a cyclic dipeptide building block, a pseudo-proline (Ψ-Pro). Ψ-Pro are known to disrupt aggregation of the peptide chain during synthesis and increasing solvation. [81] The cyclic Gly-Ser dipeptide (Ψ-GS) was incorporated in Loop 1 of WWcalm3, which drastically improved the synthesis quality (Figure 17D).

Structure and Thermostability

The structure and thermal stability of the miniproteins was analyzed by CD spectroscopy (Figure 18). The hPin1 WW domain displays two negative bands at 200 nm and 210 nm and a positive band at 230 nm. The negative band at 200 nm stems from the large disordered termini of the domain. The other negative band stems from the β-sheet structure or aromatic residues. The interactions of Trp residues with other aromatic side chains result in an exciton coupling with a negative band around 215 nm and a positive band around 230 nm. [82]

Both WWcalm1 (Figure 18B) and WWcalm2 (Figure 18C) displayed a negative band at 200 nm and no additional signals in the CD spectra. These spectra indicate a random coil structure for both models. [72] This was likely due to the mutation of Tyr 18 to Leu or Val. Tyr 18 is a highly conserved residue of the hPin1 WW domain. Albeit part of the active site, it also stabilizes the peptide by an aromatic interaction with Trp 29. Mutations of Tyr 18 to non-aromatic residues remove this interaction. [27]

The CD spectra of WWcalm3 (Figure 18D) showed a negative band at 200 nm, a small negative band at 215 nm and a small positive band at 230 nm. The expected strong exciton signal of Trp to Tyr was not present. The spectra suggested a partially folded peptide, also known as molten globule. [83] With the addition of Ca(II), the negative band at 215 nm became more pronounced but the expected conformation was not formed. Thermal denaturation experiments yielded the same results for Ca(II)-free and -containing solutions (see Figure 18E). The denaturation curves showed two transitions, one at around 20 °C and a second one at around 70 °C. An energy barrier preventing the formation of the native structure is typical for molten globule states. [83] This warranted a rational redesign of the miniprotein targeting the flexible Loop 1, the mutation of Tyr 18 and the uninvolved β-strand 3.

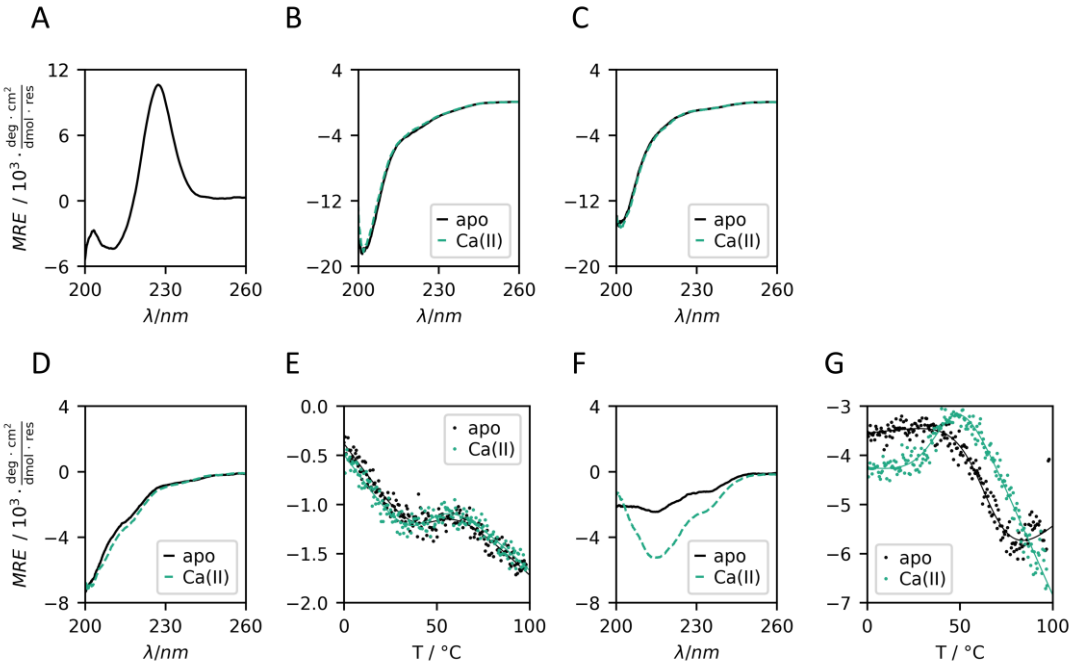


Figure 18. CD spectra and thermal stability of the WWcalm series miniproteins. CD spectra at 20 °C and thermal denaturation curves from 0 °C to 98 °C measured in pH 7.2 buffer containing 10 mM

MOPS and 150 mM NaCl in the absence (black) or presence (green) of 2 mM $CaCl_2$. A) CD spectra of hPin1 WW domain synthesized and measured by Pham et al.; [74] B) CD spectra of WWcalm1; C) CD spectra of WWcalm2; D) CD spectra and (E) thermal denaturation curves of WWcalm3; F) CD spectra and (G) thermal denaturation curves of WWcalm4 (T_m = 61 ± 0.5 °C for the apo-peptide). [73]

WWcalm4 (Figure 18F) displayed three negative bands with a minimum at 215 nm, a shoulder around 220 nm and a small signal at 230 nm. In the presence of Ca(II), the signals intensified. This conformational change upon Ca(II) addition indicated a successful design of a Ca(II) binding site. [54] The CD spectra showed a strong β-sheet signal suggesting a highly ordered structure. The expected Trp exciton coupling was either redshifted with a negative band at 230 nm or missing in the spectra. [72] Thermal denaturation experiments revealed a melting temperature (T_m) of 61.0 ± 0.5 °C in the absence of Ca(II) (see Figure 18G). [73] In the presence of Ca(II), two transitions were observed. One transition at around 40 °C and one at around 70 °C. After the measurement, precipitate was found in the cuvette. This led to the conclusion that WWcalm4 formed β-sheet aggregates leading to the strong negative band at 215 nm. [84] Similar to WWcalm3, the sequence required a redesign for higher stability and solubility.

Sequence optimization for Stability

To increase the thermal stability and improve folding of the designed sequences, a rational design approach was chosen (see Figure 19). A salt bridge between Glu 11 and Arg 16 was installed in Loop 1 to reduce its high flexibility and close the distance between β-strands 1 and 2 in WWcalm3.1. In addition, β-Strand 2 was redesigned to more closely resemble the hPin1 sequence in WWcalm3.2. WWcalm3.3 also included a new β-Strand 3 that was found to increase the thermal stability by Christina Lindner in her work on WW domains. [40] WWcalm3.4 included a mutation of the active site from Thr 20 to Asn.

The new sequences were synthesized with the same procedure as WWcalm3, including the Ψ-GS in Loop 1. The modified miniproteins were synthesized successfully at a high quality and purified without issue.

Unlike the parent miniprotein, WWcalm3.1 showed the CD signature expected for a WW domain (Figure 20A). A positive band at 229 nm was evident of a Trp exciton coupling and a negative band centered around 205 nm of a largely disordered β-sheet peptide. [72] There was no visible difference between the Ca(II)-containing and -free solutions. WWcalm3.1 displayed one transition in thermal denaturation experiments (Figure 20B) with a T_m = 43 ± 1.0 °C. [73] A spectrum measured after heating showed a more pronounced β-sheet minimum at 205 nm. This annealing effect was reproducible, hinting at an energy barrier between the initial folding and the thermodynamic minimum structure. [83] Stabilization of the loop by incorporation of a salt bridge was successful.

WWcalm3.2 CD spectra (Figure 20C) contained a large positive aromatic interaction band at 229 nm and a broad negative band centered around 210 nm. Again, no difference between Ca(II)-free and -containing solution was observed. Thermal denaturation experiments (Figure 20D) revealed one transition at a T_m = 59 ± 0.5 °C. [73] A spectrum measured after thermal denaturation displayed not only a weaker β-signal but also a weaker exciton coupling. The second set of mutations lead to a higher melting temperature but also made folding partially irreversible.

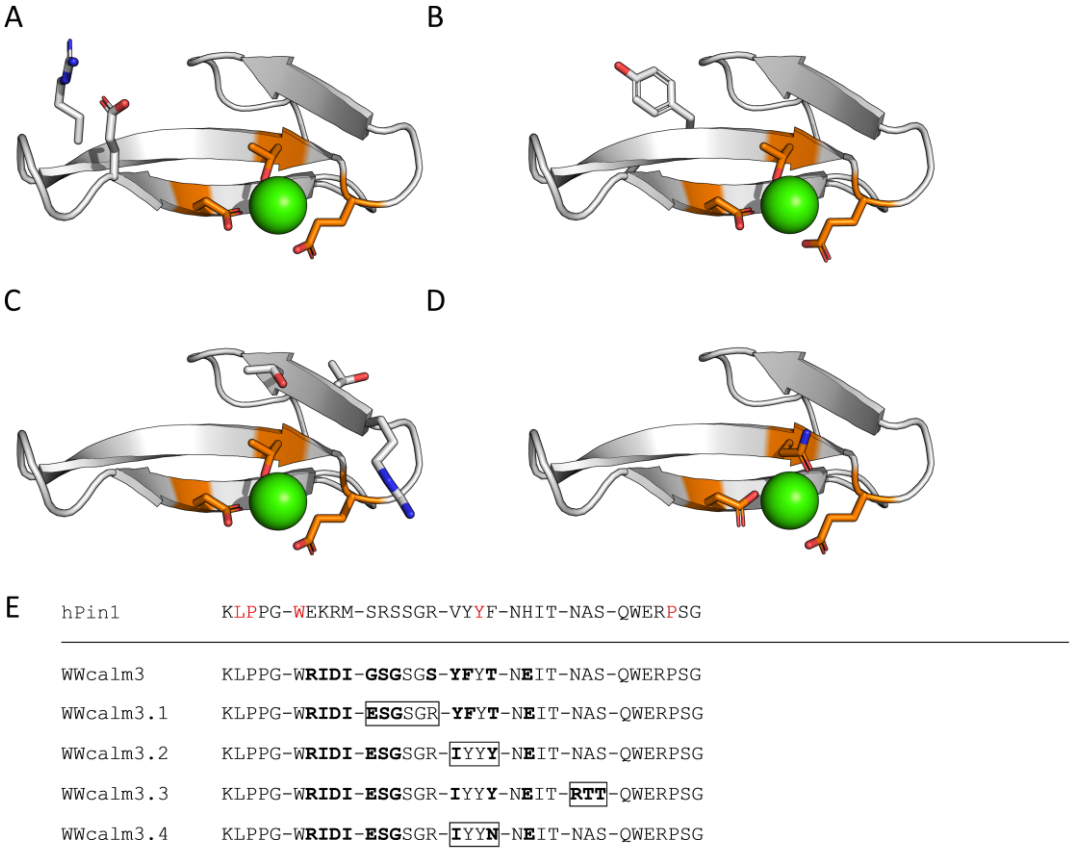


Figure 19. Sequence optimization of WWcalm3. *Rosetta3* models with binding residues (orange) and redesigned residues shown of (A) WWcalm3.1, (B) WWcalm3.2, (C) WWcalm3.3 and (D) WWcalm3.4; (E) Sequences of the WWcalm3 series. Hydrophobic core positions in red and mutations in bold. Newly introduced redesigns are boxed. [27]

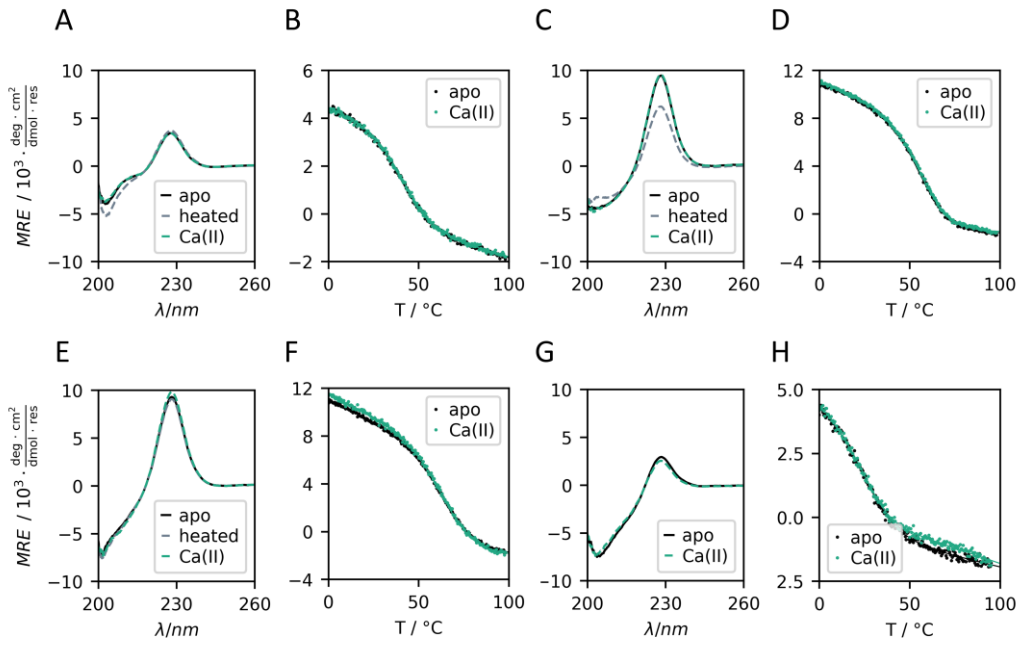


Figure 20. Circular dichroism spectra of the WWcalm3 series miniproteins. CD-spectra at 20 °C or thermal denaturation curves from 0 °C to 98 °C measured in pH 7.2 buffer containing 10 mM MOPS and 150 mM NaCl in the absence (black) or presence (green) of 2 mM $CaCl_2$. A) CD spectra and (B)

thermal denaturation curves of WWcalm3.1; C) CD spectra and (D) thermal denaturation curves of WWcalm3.2; E) CD spectra and (F) thermal denaturation curves of WWcalm3.3; G) CD spectra and (H) thermal denaturation curves of WWcalm3.4.

WWcalm3.3 also folded into a native-like conformation (Figure 20E). The β-sheet minimum was more pronounced than in WWcalm3.2 and centered on the 205 nm band with a shoulder around 215 nm. The Trp exciton signal remained unchanged in the absence of Ca(II). In the presence of Ca(II), the Trp signal was slightly increased. Thermal denaturation experiments (see Figure 20F) showed one transition at T_m = 68 ± 1.0 °C in the presence or absence of Ca(II). [73] A spectrum measured after thermal denaturation revealed a reversible folding-to-unfolding transition.

The spectra of WWcalm3.4 (Figure 20G) displayed a weaker Trp signal as for WWcalm3.1 and a strong negative band centered around 205 nm as for WWcalm3.3. The exciton signal was slightly decreased in the presence of Ca(II). Thermal denaturation experiments (see Figure 20H) also showed a different T_m = 28 ± 2.0 °C in the absence of Ca(II), than the T_m = 31 ± 2.0 °C in its presence. [73] The plateau before the transition was not visible in the experiment and the overall melting temperature was drastically decreased compared to the other redesigns. The tertiary structure might not be fully formed, suggesting a molten globule state for WWcalm3.4. Activity was severely hindered by a stability-function trade-off. [85]

WWcalm4 was also redesigned using the findings on the WWcalm3 series (see Figure 21). WWcalm4.1 included the salt bridge in Loop 1, the hPin1-like β-Strand 2 and a stabilizing mutation in Loop 2. [40] WWcalm4.1 was synthesized with the same protocol as the WWcalm3 series and successfully purified by HPLC.

WWcalm4.1 showed the same aggregating tendencies as WWcalm4 in a solution containing Ca(II). This revealed that the problem was not one of stability or folding. Using the *Protpi Peptide Tool*, [86] the charge of the sequences at different pH values was calculated (Figure 21). WWcalm4 produced a net charge of -1.8 for pH 7.4. When Ca(II) was bound, the net charge was nullified, which lowered the solubility and likely lead to aggregation. [84] A mutation such as Asn 27 to Asp would lead to a lower net charge of -2.8, prevent aggregation in the bound state, and could enable a soluble Ca(II)-binding WW domain.

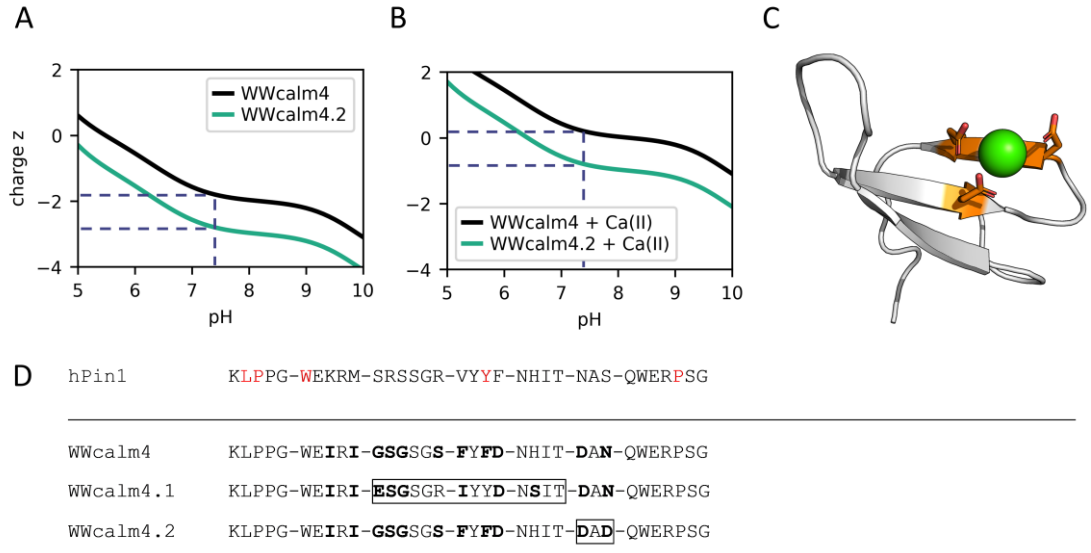


Figure 21. Sequence optimization of WWcalm4. A) *Protpi Peptide Tool* simulation of the net charge of WWcalm4 (black) and WWcalm4.2 (green) at different pH values in the absence and B) in the presence of Ca(II); [86] C) *Rosetta3* model of WWcalm4.2 with net charge at pH 7.4 marked (dashed

line); [4] D) sequences of the WWcalm4 series with mutations printed bold and redesigned elements boxed. [27]

Binding Studies

To characterize the Ca(II) binding of the designed miniproteins, the chelator dye Oregon Green™ 488 BAPTA-5N (BAPTA-5N) was used. The competitive dye binding assay was performed according to a protocol by Linse [75] on a Fluorescence spectrometer. Binding experiments for WWcalm3.1, WWcalm3.2 and WWcalm3.3 were prepared. To a Ca(II)-free buffered solution of chelator dye and miniprotein, a small volume of concentrated Ca(II) solution was added until saturation was reached. The results were analyzed using a python script (S1). [77] Figure 22A shows simulated data points for different K_d-values in competition with BAPTA-5N. The grey curve describes the behavior without competition, a non-binding miniprotein. The data sets found for all three miniproteins fit to the non-binding curve (Figure 22). The analysis gave better results, the higher the guess K_d-values were. The miniproteins were therefore not competing with the dye. The WWcalm3 series was deemed non-binding.

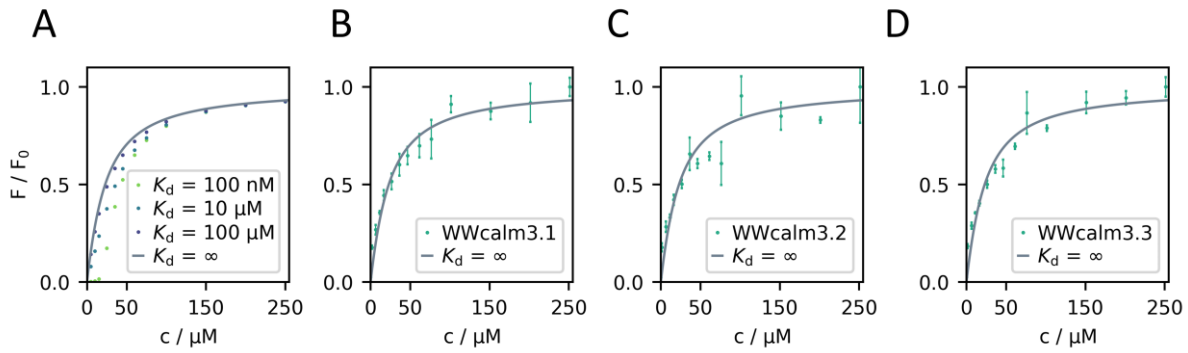


Figure 22. Competitive binding experiments of the WWcalm3 series with BAPTA-5N. Fluorescence intensity at 480 nm excitation and 525 nm emission of 10 µM BAPTA-5N with 20 µM miniprotein and various concentrations c of Ca(II) in pH 7.2 buffer containing 20 mM *N*-ethylmorpholine, 40 mM acetic acid and 100 mM NaCl. A) Simulated data sets for 100 nM, 10 µM, 100 µM and infinitely large K_d; B-D) real data (green) with simulated data for an infinitely large K_d (grey) for (B) WWcalm3.1, (C) WWcalm3.2 and (D) WWcalm3.3. [75, 77]

4.2 Intuitive *de novo* Design of Ca(II)-binding Miniproteins

De novo design aims to create new functionalities, not with a natural model as a blueprint but based on observation of a multitude of similar complexes and an understanding of the underlying principles. [87] Ca(II) binding sites in proteins coordinate the metal ion in pentagonal bipyramidal complexes of high negative charges. As pointed out by Tang et al. for their designed binding sites, a higher negative charge increases the Ca(II) binding affinity. [54] Inspired by their group's designed Ca(II) switch based on the CD2 protein, a new approach at intuitive design was undertaken. [5]

The design of the Ca(II) binding site followed three requirements: (1) a surface position, (2) a high negative charge, and (3) a conformational change upon binding. To fulfill all three requirements, the design was based on charge inversion in complex salt bridges. A salt bridge is a non-covalent bond between oppositely charged residues in close proximity on a protein or peptide. They are generally known to have a small contribution to structural stability. Salt bridges can be exposed on the surface or buried in the hydrophobic core. Most are built from two opposite charges but complex salt bridges form networks of multiple charges. [88] For the design, the peptide surface was searched for a (partially) exposed complex salt bridge. The positively charged contributors, such as Arg, Lys or His, were then mutated to a negatively charged Asp. This replaced the attractive Coulomb forces in the network by repulsive ones. The structure was destabilized and a high negative charge was introduced to the surface of the peptide. A Ca(II) ion attracted to the negative charge could, upon binding, close the gap created by the repulsive charge interactions. If the design was successful, this should lead to a conformational change upon binding to a highly charged *de novo* site on the surface of the protein, fulfilling all requirements set for the design.

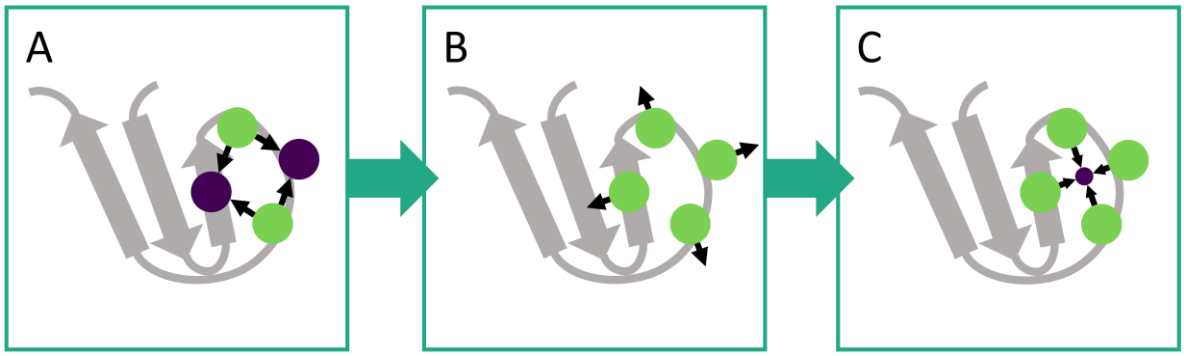


Figure 23. Intuitive Design by salt bridge inversion. Design scheme with (A) a β-sheet peptide with a complex salt bridge, (B) a destabilized β-sheet peptide with a negatively charged site, and (C) a Ca(II)-bound β-sheet peptide. Arrows show electrostatic interaction from negative (green) to positive charge (purple). (Page 4, Figure 3 [54])

4.2.1 ScaX

Design Strategy

According to the design principles set for the *de novo* design, the SH3 domain was searched for an exposed complex salt bridge. The SH3 domain contained several paired salt bridges on the surface of the β-barrel and in the loops, but no complex salt bridge. Two exposed paired salt bridges were directly adjacent and used for the design instead. Arg 29 on β-Sheet 2 binds to Glu 40 on β-Sheet 3 and Arg 46 on β-Sheet 4 binds to Asn 38 on β-Sheet 2. Both, Arg 29 and Arg 46 were mutated to Asp. The salt bridges were lost and the 3D structure destabilized. In the presence of Ca(II), the active site residues were expected to bind the cation like a chelator, reforming the native-like structure. As the design was based on two crossed salt bridges to bind Ca(II) on an SH3 scaffold, it was named ScaX.

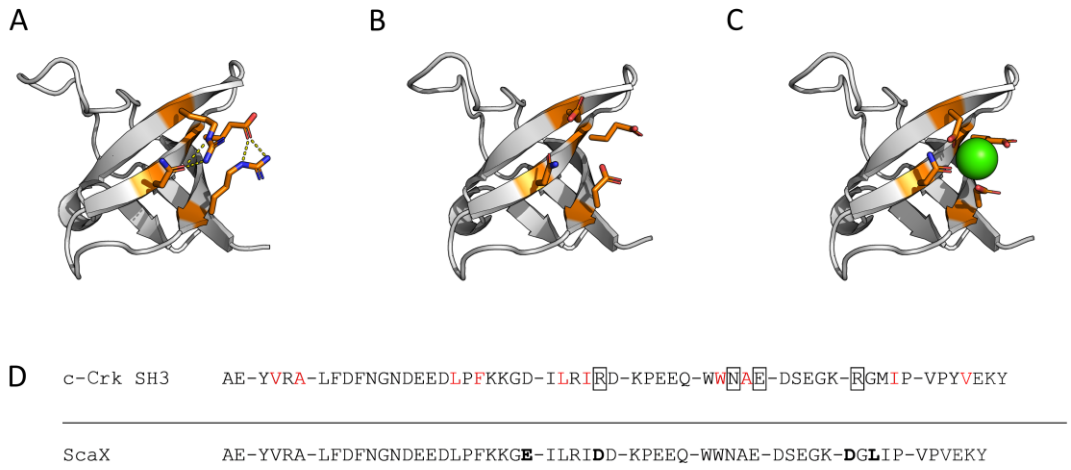


Figure 24. Intuitive Design of ScaX. A) Crystal structure of the c-Crk SH3 domain (pdb: 5IK5) with the residues of the complex salt bridge shown in orange; B) *PyMol* model of ScaX in the absence and (C) in the presence of Ca(II) with the binding residues shown in orange; D) sequences of c-Crk SH3 and ScaX with mutations printed bold, hydrophobic core positions printed red and complex salt bridge boxed. [26]

Synthesis and Purification

With a severely aspartimide-promoting Asp-Gly section in the sequence of ScaX, a new protocol was used for synthesis based on the work by Pham et al. [74] The standard protocol suppresses aspartimide by a bulky OtBu protecting group and formic acid added to the deprotection solution. The cyanosulfurylide (CSY) protocol uses a CSY protecting group to mask the carboxylate by a carbon-carbon bond that is stable at low-temperature deprotection conditions. [74] With the implementation of the first Fmoc-Asp(CSY)-OH in the aspartimide-promoting sections, deprotection steps were conducted at room temperature.

The miniprotein was synthesized at 50 % scale on H-Rink amide ChemMatrix™ resin with Fmoc-Asp(CSY)-OH coupled for Asp 46. Standard cleavage conditions did not deprotect CSY-Asp. The CSY-protected product was purified by RP-HPLC in multiple rounds. The pure CSY-protected miniprotein was deprotected in DMF and purified in a 10-60 % B in A gradient. [74] The final product was separated into aliquots prior to experimentation. The CSY protocol led to a clean product, but at the cost of a longer synthesis time, an additional deprotection step that required careful control of reaction time and temperature and an additional purification step for a clean product.

A Trp 3 variant was synthesized on a H-Rink amide Tentagel resin according to the O-methyl-pentyl (OMpe) protocol. Aspartimide formation for the Asp-Gly sequence was prevented by the use of Fmoc-Asp(OMpe)-OH as an even bulkier substitute for Fmoc-Asp(OtBu)-OH. [89] The miniprotein was purified by RP-HPLC in one run and separated into aliquots for experimentation. This protocol for aspartimide-promoting sequences led to high synthesis quality and ease of purification without the complex deprotection necessary for the CSY-protecting group. The CSY protocol should therefore only be used in extreme cases where aspartimide-formation cannot be suppressed by other means.

Fluorescence Resonance Energy Transfer

Fluorescence Spectroscopy was used to test the new miniproteins for their Ca(II) binding functionality. A direct experiment to determine Ca(II) association by intrinsic Trp Fluorescence would require a strong signal change upon binding. This is usually not the case for Ca(II). The Tb(III) cation has a slightly smaller ionic radius than Ca(II) and carries a higher positive charge. It is an even harder cation that binds to Ca(II) sites similar to or stronger than Ca(II). Tb(III)

forms a strong Fluorescence Resonance Energy Transfer (FRET) with aromatic residues, especially Trp. This can be used to assess the binding of Tb(III) to the miniprotein in a two-step experiment. The Fluorescence of the 5 µM miniprotein was measured in the absence and presence of 2 mM Tb(III). If Tb(III) was bound, the Trp signal (for excitation at 295 nm) around 350 nm was quenched and a new signal at 545 nm was visible. [6]

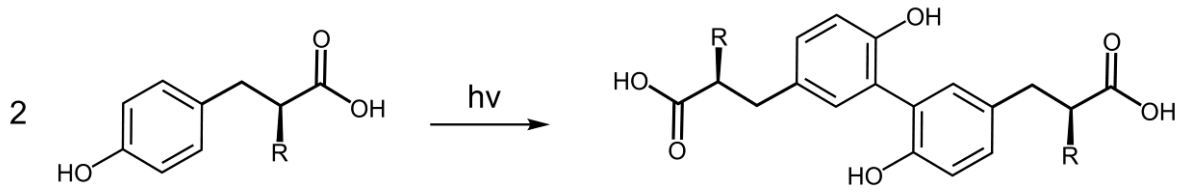


Figure 25. Reaction scheme for the formation of dityrosine (Tyr_2). [90]

The Fluorescence spectra of ScaX (Figure 26) displayed an intensive signal at 402 nm with a shoulder around 350 nm. Trp fluorescence can blue- or redshift depending on the environment. These shifts are usually up to 10 nm and no additional signal should be visible at 350 nm. The two overlapping bands were therefore interpreted as two different species: the expected Trp fluorescence at around 350 nm and another undetermined aromatic species at 402 nm. A search for aromatic fluorescence signals in proteins led to dityrosine (Tyr_2) with a fluorescence emission at around 402 nm. Tyr_2 can form from two neighboring Tyr residues upon excitation by light (Figure 25). [90] It was postulated that, during the work-up of the miniprotein, the Tyr 3 and Tyr 57 residues were excited to form Tyr_2. To validate this theory, a Trp 3 variant was tested as well (Figure 26C). With no neighboring Tyr residues, no Tyr_2 can be formed. The fluorescence spectra displayed only a Trp signal at 350 nm, proving that the large signal at 402 nm was caused by a product formed from Tyr 3.

In the presence of Tb(III), both variants showed a peak at 545 nm, but for ScaX, the intensity was much higher than for its Trp 3 variant. As can be seen in Figure 26, the FRET donor in ScaX is not Trp but Tyr_2, showing a strong quenching of the Tyr_2 signal and an increase of the Tb(III) signal. It was evident, that Tb(III) bound to the miniprotein, making Ca(II) binding highly likely.

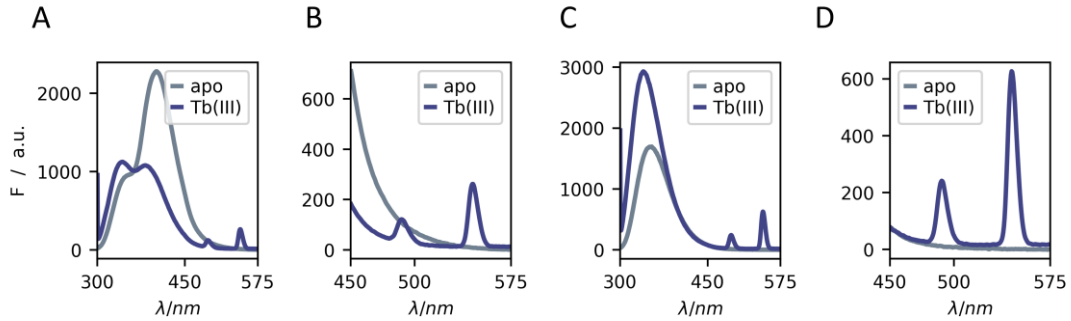


Figure 26. Fluorescence spectra of ScaX and its Trp 3 variant in the absence (grey) and presence (green) of 2 mM Tb(III) measured at 5 µM concentration in pH 7.2 buffer containing 10 mM MOPS and 150 mM NaCl. A) Fluorescence spectra of ScaX with (B) a zoom-in of the Tb(III) signals; C) fluorescence spectra of the Trp 3 variant with (D) a zoom-in of the Tb(III) signals. [91]

Structure and Thermostability

The CD spectra of ScaX (Figure 27B) displayed two negative bands at 200 nm and 228 nm, and a positive band at 220 nm. The negative bands corresponded to a largely disordered conformation and aromatic interactions respectively. [72] The former was blue-shifted compared to the native SH3 domain, pointing at a more disordered structure. The positive

band stemmed from the aromatic residues and β-turns in the barrel structure. [71] Overall, the SH3 CD signature was present, albeit less pronounced than expected. Thermal denaturation experiments (Figure 27D) showed a continuous unfolding of the structure without a clear transition. The results suggest that ScaX was only partially folded in solution. This was enforced by the formation of Tyr_2, which can only occur if the terminal Tyr residues were in proximity, therefore fulfilling the requirement of a compact structure. [83]

The Trp 3 variant showed similar spectra (Figure 27C) and thermal denaturation curves as the original sequence. The negative band at 228 nm was less pronounced and the positive band at 220 nm more pronounced. The 200 nm minimum was slightly red-shifted hinting at a more SH3 domain-like structure. In the presence of Ca(II), this negative band broadened and became more pronounced. Both sequences produced a mainly unfolded structure with an aromatic exciton coupling signal. At least one of the mutated salt bridges was likely required for the full β-barrel to form, as even the presence of 2 mM Ca(II) did not reorganize the superstructure.

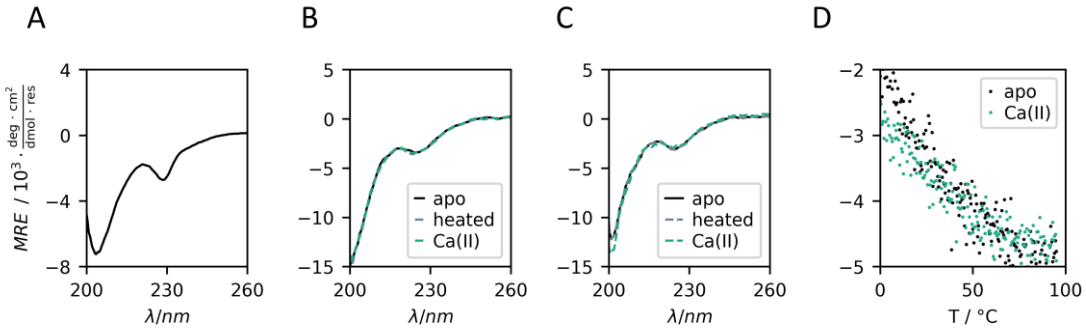


Figure 27. CD spectra and thermal denaturation curves of ScaX in the absence (black) and presence of Ca(II) (green) or after heating (grey); A) CD spectra at 20 °C of c-Crk SH3 in pH 7.0 buffer containing 10 mM sodium phosphate measured by Pham et al.; [74] B) CD spectra of ScaX at 20 °C; C) CD spectra of ScaX Trp 3 at 20 °C; D) thermal denaturation curves from 0 °C to 98 °C of ScaX in pH 7.2 buffer containing 10 mM MOPS and 150 mM NaCl.

4.2.2 WWturn

Design Strategy

To install a *de novo* Ca(II) binding site in a WW domain, the hPin1 WW domain crystal structure (pdb: 1I6C) was searched for an exposed complex salt bridge. The WW domain contained one complex salt bridge (Figure 29A) between the *N*-terminus and *C*-terminus of the domain. The side chains of Glu 30 and Gln 28 of the *C*-terminus bind non-covalently to the side chain and amine group of the *N*-terminal Lys 1. The salt bridge is exposed and complex but cannot be fully inverted without a chemical modification of the *N*-terminus. However, a similar complex salt bridge can be found in the circular permutated variant of the hPin1 WW domain designed by Kier et al. (permutated WW, pdb: 2MDU, see Figure 28). In this variant, the WW domain was cut in Loop 1 and reconnected at the *N*- and *C*-termini. [8]

For the circular permutated WW domain (Figure 29B) the salt bridge is located in the termini-connecting loop. The side chains of Arg 16 and Lys 18 bind the side chain of Glu 15, and Gln 13 binds the main chain oxygen of Glu 15 or Lys 18. The site fulfilled all previously set criteria and was inverted by mutating Arg 16 and Lys 18 to Asp. The resulting mutant was destabilized in the absence of Ca(II) and should reform the loop in its presence. Since the binding site was located at a β-turn of the circular permutated WW domain, it was named WWturn (Figure 29).

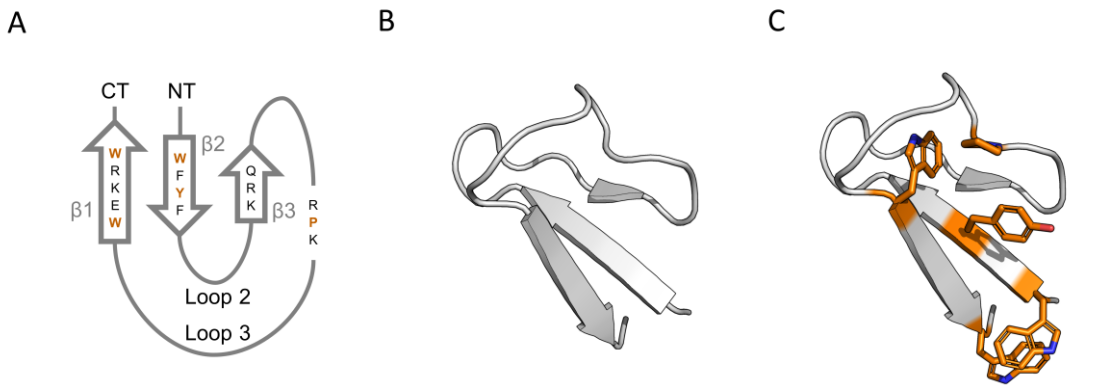


Figure 28. Circular permutated hPin1 WW domain (pdb: 2MDU). A) Schematic representation of the secondary structure motifs with *N*-terminus (NT), β-sheets 1-3 (β), loops and *C*-terminus (CT), and all amino acids of the hydrophobic core marked in orange; B) ribbon representation of the WW domain; C) WW domain with the amino acids of the hydrophobic core marked in orange. [8]

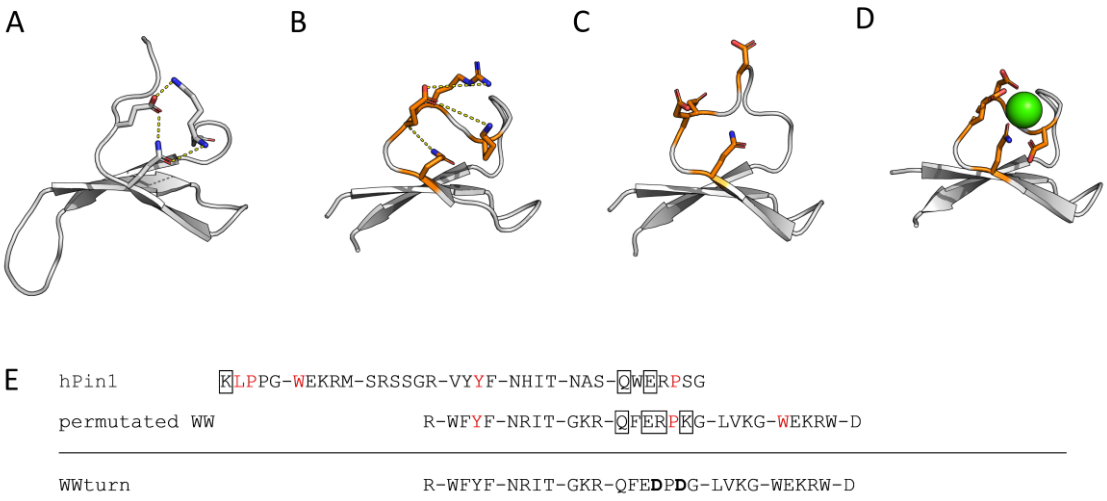


Figure 29. Intuitive design of WWturn. A) Crystal structure of the hPin1 WW domain (pdb: 1I6C) with the residues of the complex salt bridge shown; B) crystal structure of the circular permutant of hPin1 by Kier et al. (pdb: 2MDU) with the residues of the complex salt bridge shown in orange; [8] (C) *PyMol* model of WWturn in the absence and (D) in the presence of Ca(II) with the binding residues shown in orange; (E) sequences of the hPin1 WW, its circular permutant and WWturn with mutations printed bold, hydrophobic core positions printed red and complex salt bridge boxed. [8, 27]

Synthesis and Purification

The miniprotein was synthesized according to CSY protocol on H-Rink amide Polystyrene resin. CSY-protected Asp was used for the coupling of Asp 18 due to the highly aspartimide-promoting Asp-Gly sequence. The CSY-protected product was purified by HPLC with a 10-50 % gradient, as higher starting concentrations of buffer B led to large injection peaks. CSY-deprotection in DMF was successful.

Structure and Thermostability

The structure and stability of WWturn was analyzed by CD spectroscopy. The parent peptide, the circular permutated WW domain (Figure 30A), displayed a positive band at 229 nm from the Trp exciton coupling and two negative bands centered on 202 and 215 nm. It folds like the native hPin1 WW domain with a T_m = 53 °C. The stronger twist of the β-sheet, intensifies the signal at 215 nm. Hence it is more pronounced than the disordered band at 202 nm. [8, 72]

WWturn showed a negative β-sheet band at 215 nm with a small shoulder around 202 nm and a positive Trp exciton coupling band at 229 nm (see Figure 30B). As seen in Figure 30D, the miniprotein folded reversibly in a native-like conformation. The data suggested WWturn to fold into a less disordered structure than the permutated WW domain with a higher Trp exciton coupling and a weaker disordered band. Thermal denaturation curves (Figure 30C) supported this with a T_m = 55.5 ± 0.5 °C, that was slightly higher than that of the parent peptide. [8, 73] There was no discernible difference in the thermal denaturation in the presence or absence of Ca(II). This was not surprising, as the Ca(II) binding site was not situated on the β-sheet or near the aromatic residues that were monitored in CD.

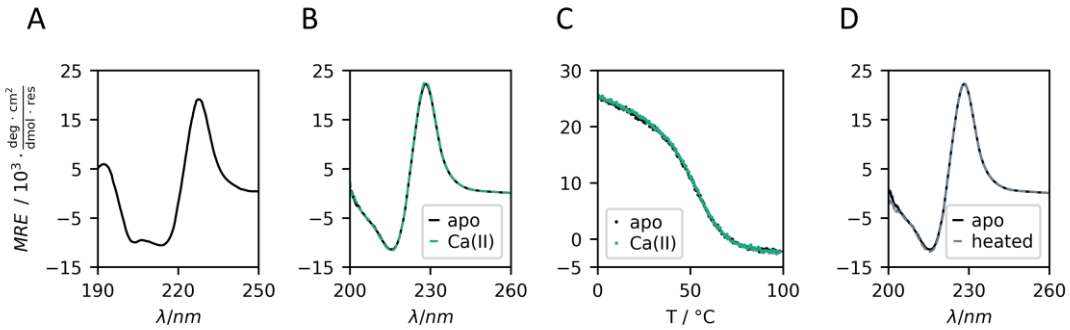
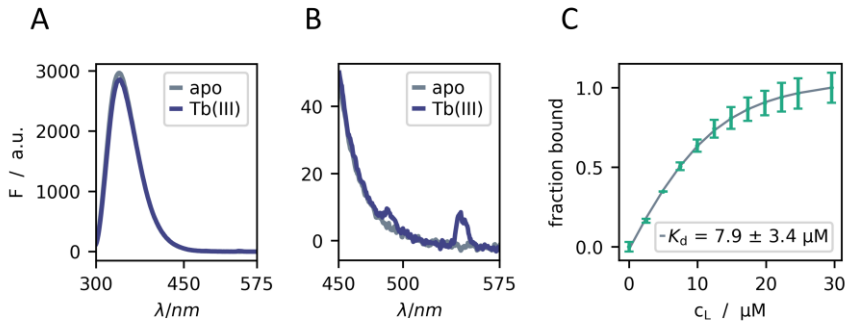


Figure 30. CD spectroscopy and Ca(II) binding study of WWturn. A) CD spectrum at 20 °C of circular permutant of the hPin1 WW domain measured by Kier et al. cp34 (adapted from Figure 3, page 744 [8]); B) CD spectra of WWturn at 20 °C; C) thermal denaturation curves of WWturn from 0 °C to 98 °C; D) CD spectra before (black) and after thermal denaturation (grey) of WWturn at 20 °C in pH 7.2 buffer containing 10 mM MOPS and 150 mM NaCl.

Binding Studies

The Tb(III) binding was investigated by Tb(III) FRET (Figure 31A-B). The fluorescence spectra of WWturn revealed a blue-shifted Trp signal at around 346 nm. The more hydrophobic the environment of Trp, the more blue-shifted the signal becomes. The Trp signal of WWturn therefore signified a buried Trp with no direct solvent contact. [91] In the presence of Tb(III), the signal decreased and a Tb(III) signal with low intensity appeared. This can be explained either by a low population of the Tb(III) complex, meaning a low binding strength, or a large distance between donor and acceptor of the FRET. The FRET signal is proportional to the distance between donor and acceptor to the power of six. A greater distance therefore quickly leads to a signal reduction. [91]

The Ca(II) binding of WWturn (Figure 31C) was analyzed by competitive binding experiments according to a protocol by Linse. [75] The miniprotein was dissolved in a Ca(II)-free buffer solution with an equal amount of the chelator dye Br_2-BAPTA. Ca(II) was titrated to the solution until saturation was reached. The data were analyzed using *DynaFit* with a K_d = 1.4 µM for Br_2-BAPTA. [92] As the mean of three titrations, WWturn displayed a K_d = 7.9 ± 3.4 µM. The *de novo* Ca(II) binding site was successfully designed in the same order of magnitude as the strongest LG domain site and the weakest CaM site (see Figure 31D). [49, 50] The system can now be tested and optimized for specific carbohydrate binding.



D

Protein	WWturn	CaM	Laminin	Slit-2
K_d [μM]	7.9 ± 3.4	0.53-1.18	5-300	100
Ca(II)-binding domain	*De novo*	EF-hand	LG3-5	LG1
source		(Zhang 2012)	(Paulsson 1988)	(Linse 1991)

Figure 31. Binding studies on WWturn in pH 7.2 buffer containing 10 mM MOPS and 150 mM NaCl. A) Fluorescence spectra in the absence (grey) and presence (green) of 2 mM Tb(III) measured at 5 μM concentration with (B) a zoom-in of the Tb(III) signals; C) competitive Ca(II) binding experiment of 10 μM WWturn and Br_2-BAPTA; [75] D) table of protein domains with EF-hands of CaM, [49] LG domains 3-5 of laminin, [50] the LG1 domain of Slit-2 [93] and their respective dissociation constants for Ca(II) compared to WWturn.

4.3 Intuitive Design of Ca(II)-binding Miniproteins by Domain Grafting

A simple approach to intuitive design is the grafting of functional domains onto stable but smaller scaffolds than the native protein. For this, a functional domain, or a part of it, was incorporated in the sequence of a stable peptide scaffold. [54] This was done by either replacing a part of the scaffold with the functional domain or by extending the scaffold by the functional domain. To secure the folding of both the new domain and the remaining scaffold, it was sometimes necessary to add spacers of one or several amino acids between them. The sequences were modelled with *ColabFold* and *OmegaFold*. The results were filtered by the software's confidence score from the local distance difference test (pLDDT) and by the similarity of the predicted and intended structure. [94, 95]

4.3.1 WWfan

Design Strategy

The binding domain was taken from a lectin recently discovered in *Psathyrella Velutina* (propeller lectin, pdb: 2BWR, see Figure 32). The carbohydrate recognition protein folds into a seven-bladed β-propeller structure. The regular blades of the β-propeller fold consist of four-stranded antiparallel β-sheets reminiscent of the integrin fold. The Ca(II) binding site of the propeller lectin is situated in a loop between two β-strands. It was proposed to stabilize the β-propeller structure and aid in ligand binding. [96]

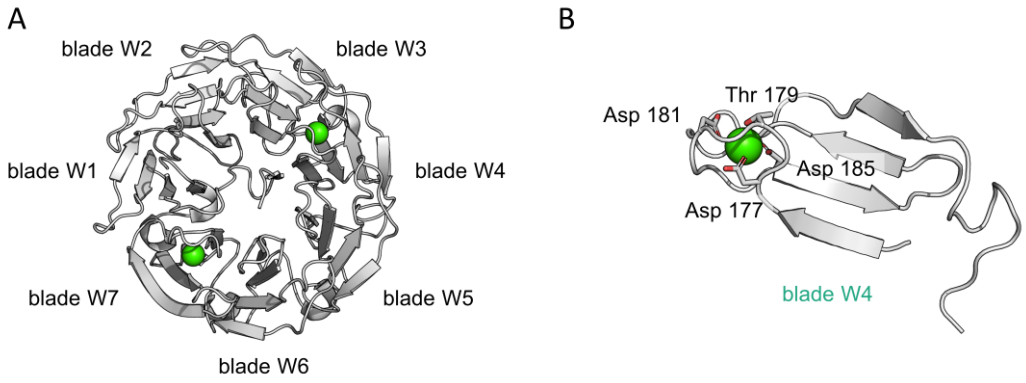


Figure 32. Ribbon presentation of the propeller lectin from *P. Velutina* (pdb: 2BWR). A) Complete protein with Ca(II) bound and each blade named; B) Blade W4 with Ca(II) binding amino acids shown. [96]

The scaffold chosen for this design was the hPin1 WW domain (Figure 33A). The three-stranded antiparallel β-sheet shares structural features with the propeller lectin's blades. The crystal structure of the WW domain was overlaid with blade W4 of the propeller lectin (Figure 33B). The WW domain's β-strands 1 and 2 overlay well with the blade's first and second β-strands. The WW domain was cut between Met 10 and Val 17 to remove Loop 1. This six amino acid sequence was replaced by the nine amino acid sequence of the Ca(II) binding domain of blade 4 (Asp 177 to Asp 185). To allow for a properly aligned β-sheet, the loop was elongated to an even number with the addition of a spacer. Several simulations were performed to find the optimal spacers among Gly, Ala, Arg, Leu, Ile and Val. A mutation of Met 10 to Gly with a neighboring Gly spacer gave the best results. *ColabFold* predicted a three-stranded β-sheet structure with a short loop and two extensive β-strands 1 and 2 in the absence of Ca(II). Albeit not the expected fusion miniprotein structure of the lectin site's and WW domain's natural folds, it could undergo a conformational change upon cation association. The expected structure was predicted by *OmegaFold* (Figure 33D), with a fully formed Ca(II)-binding loop integrated in the WW domain. Both modelling softwares predicted a stable β-sheet miniprotein with a highly charged Loop 1. The model was named WWfan and chosen for synthesis.

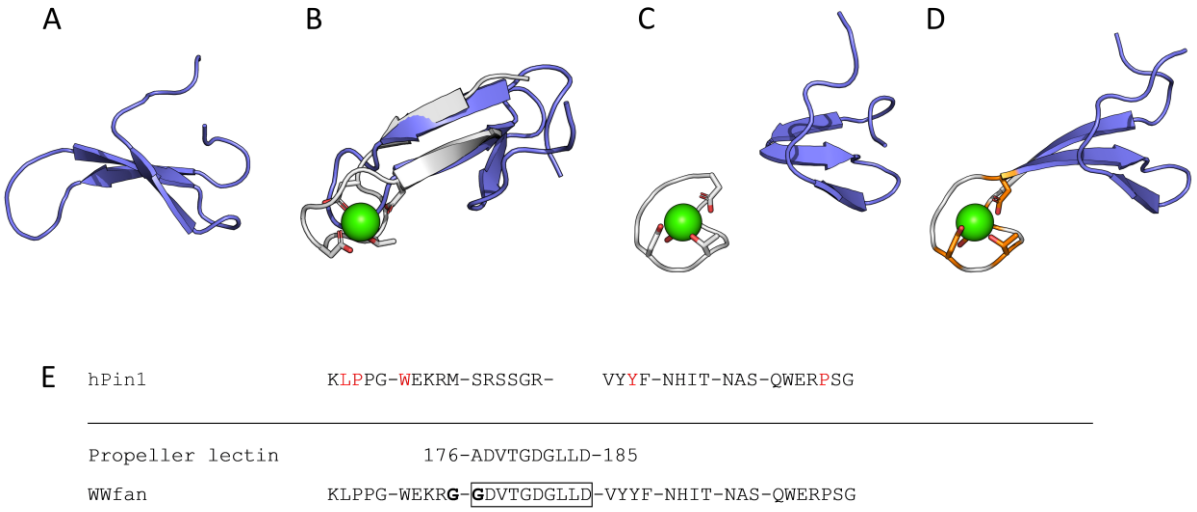


Figure 33. Intuitive Design of WWfan. A) Crystal structure of the hPin1 WW domain (pdb: 1I6C); B) crystal structure of the propeller lectin blade W4 (pdb: 2BWR, 171-202, grey) overlaid with the WW domain (blue); C) model of the grafted Ca(II) binding domain of propeller lectin blade W4 (grey) and the hPin1 WW domain without Loop 1 (blue); D) *OmegaFold* Model of WWfan (blue) with the grafted domain in grey; E) sequences of the hPin1 WW, the Ca(II) binding domain of blade W4 and WWfan with hydrophobic core positions printed red, mutations printed bold and grafted sequences boxed. [27, 96]

Synthesis and Purification

The miniprotein was synthesized according to CSY protocol on H-Gly-HMPB ChemMatrix™ resin, due to the Asp-Gly site at position 16. The final product required an additional purification after CSY-deprotection in DMF.

Structure and Thermostability

The structure and thermal stability of WWfan was analyzed by CD spectroscopy. The CD spectra of the miniprotein (Figure 34A) displayed a negative band centered around 205 nm with a shoulder around 210 nm, and a positive band at 229 nm for the Trp exciton coupling. WWfan folded into a WW domain-like structure with a less pronounced β-sheet than the native WW domain. [82] Thermal denaturation curves (Figure 34B) revealed a T_m = 27 ± 1 °C, in accordance with the weaker folding seen in the CD spectrum. [73] A difference between the absence and presence of Ca(II) was only observed for refolding after thermal denaturation. In its presence, the miniprotein showed almost perfectly reversible folding (Figure 34C). In Ca(II) free solution, the folding was recovered only partially.

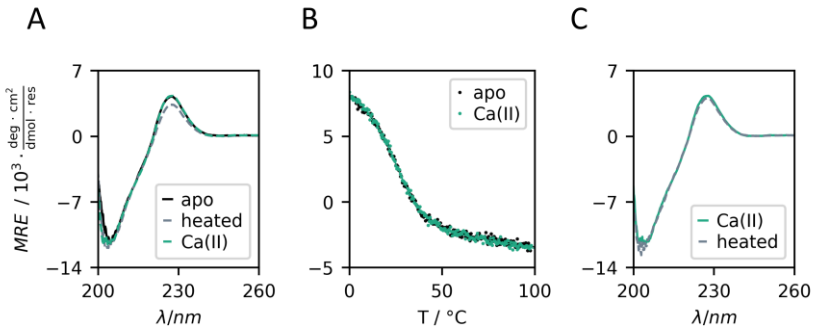


Figure 34. CD Spectroscopy of WWfan at 100 µM concentration in pH 7.2 buffer containing 10 mM MOPS and 150 mM NaCl. A) CD spectra at 20 °C in the absence (black) and presence of Ca(II) (green)

or after thermal denaturation (grey); B) thermal denaturation curves from 0 °C to 98 °C; C) CD spectra at 20 °C in the presence of Ca(II) before (green) and after (grey) thermal denaturation.

The WW domain structure was likely disturbed by the elongation of Loop 1 from six to ten amino acids. Gly 15 of Loop 1 is a nucleation site of the WW domain fold. [27] The extension of the loop positioned Gly 15 in its center instead of the penultimate position. A mutation of Leu 19 of WWfan to Gly could reinstall the nucleation site and improve the folding. Both *ColabFold* and *OmegaFold* predicted a model that aligned better with the native WW domain with a higher pLDDT value.

Binding Studies

The Tb(III) binding was investigated by FRET (Figure 35). The spectra of WWfan revealed a blue-shifted Trp signal at around 346 nm, typical for a buried Trp residue. [91] In the presence of Tb(III), the signal decreased and a low-intensity Tb(III) signal at 545 nm appeared. The predicted distance between the nearest Trp side chain and the binding site was only 7 Å. This short distance should yield a strong FRET signal for a high complex population, suggesting a weak Tb(III) binding for WWfan. A more stable variant may not only improve the folding of the WW domain structure but also the grafted domain and thus, the binding affinity.

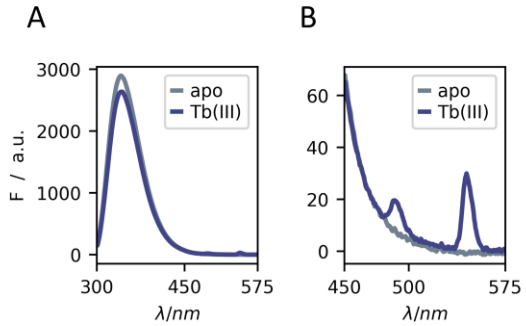


Figure 35. Fluorescence spectroscopy of WWfan in pH 7.2 buffer containing 10 mM MOPS and 150 mM NaCl. A) Fluorescence spectra of 5 µM WWfan in the absence (grey) and presence of Tb(III) (purple) and (B) zoom-in on the Tb(III) signals.

4.3.2 Calmcage

Design Strategy

The Tryptophan cage (Trp-cage, pdb: 1L2Y, see Figure 36) is a 20 amino acid protein folding motif developed by truncation and mutation of exendin-4. The structure is named for its burial of the Trp 6 side chain in a sheath of Pro rings that drives the folding. Aside from mutations to strengthen the Trp-cage motif, the sequence is truncated at the *N*-terminus to remove its non-interacting *N*-terminal helix and create the shortest protein-like construct. [10] By grafting a functional helical domain onto the *N*-terminus of the Trp-cage, a new functional miniprotein was designed.

Figure 37 shows the design process of Calmcage, a fusion miniprotein of CaM EF-hand 2 and Trp-cage. The EF-hand binds Ca(II) in a loop with a *C*-terminal α-helix. The chosen domain was aligned with the *N*-terminal α-helix of the Trp-cage to form one joint elongated α-helix. Predictions of the sequence in *ColabFold* and *OmegaFold* agreed with the expected fusion miniprotein. [94, 95] To allow for higher solubility in aqueous solution, Phe 14 of the EF-hand was mutated to Tyr. The mutant was predicted to fold the same way as the Phe 14 sequence. As a fusion miniprotein of CaM and Trp-cage, the Tyr 14 model was chosen for synthesis.

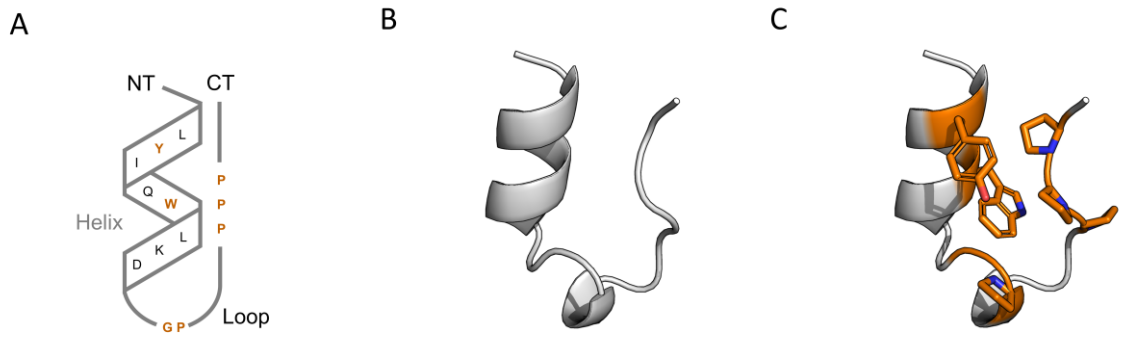


Figure 36. Tryptophan cage by Neidigh et al. (pdb: 1L2Y). A) Schematic representation of the secondary structure motifs with *N*-terminus (NT), α-helix (Helix), loop and *C*-terminus (CT), and all amino acids of the hydrophobic core marked in orange; B) ribbon representation of the Trp-cage; C) Trp-cage with the amino acids of the hydrophobic core shown and marked in orange. [10]

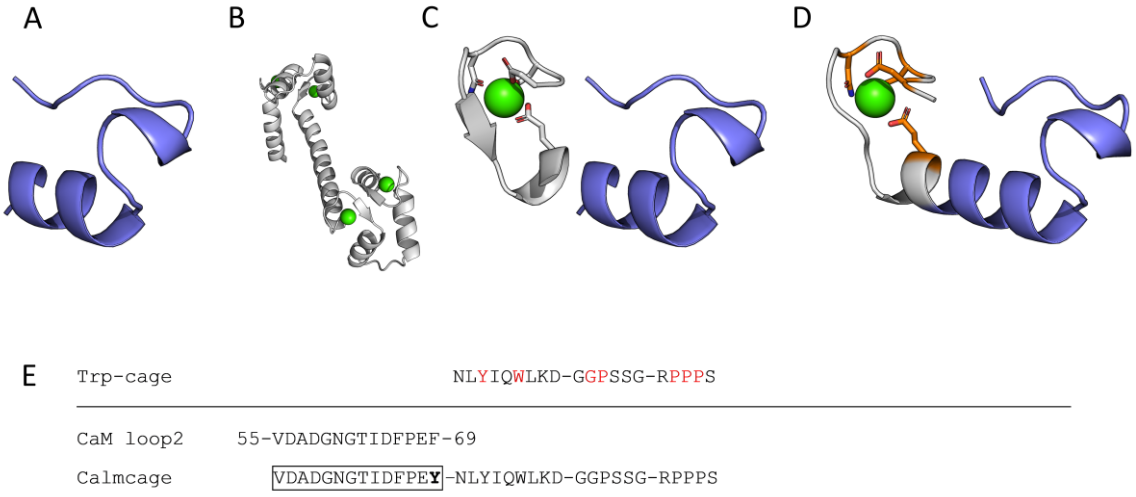


Figure 37. Intuitive Design of Calmcage. A) Crystal structure of the Trp-cage by Neidigh et al. (pdb: 1L2Y); B) crystal structure of rat Calmodulin (pdb: 3CLN); C) model of the grafted EF-hand of Calmodulin Loop 2 (grey) and the Trp-cage (blue); D) *ColabFold* Model of Calmcage (blue) with the grafted domain in grey and binding amino acids in orange; E) sequences of the Trp-cage, the EF-hand of Calmodulin Loop 2 and Calmcage with hydrophobic core positions printed red and grafted sequences boxed. [10, 80]

Synthesis and Purification

The miniprotein was synthesized according to standard protocol on H-Ser-HMPB ChemMatrix™ resin with OMpe-protected Asp coupled for Asp 4 and Asp 24. The product was successfully purified by HPLC.

Structure and Thermostability

The structure and stability of Calmcage was analyzed by CD spectroscopy. The original Trp-cage has only two structural motifs, namely a right-handed α-helix (Leu 2 to Asp 9) and a left-handed P_2-helix (Pro 17 to Pro 19). An α-helix typically shows two negative bands at 222 nm and 208 nm and a P_2-helix one negative band at 206 nm and one positive band at 226 nm. Overlapping both results in the Trp-cage CD spectra (Figure 38A) with two negative bands at 223 nm and 206 nm. It displays cooperative unfolding with a two-state transition at T_m = 42 °C. [10, 72]

The CD spectra of Calmcage (Figure 38B) showed a similar-to-native folding with two negative bands at 223 nm and 206 nm, the latter being more pronounced. The α-helix signal likely intensified due to its elongation by the grafted domain. [72, 97] As seen in Figure 38A, the

native exendin-4, with its longer helix, also displays a more pronounced 206 nm band. [98] Thermal denaturation experiments (Figure 38C) revealed a T_m = 50 ± 3 °C, considerably higher than for the parent peptide. This was not surprising, as the predicted models contained a tight packing of the grafted domain around the α-helix and the *C*-terminus of the Trp-cage. The *N*-terminal Val side chain of CaM Loop 2 was predicted to point directly into the hydrophobic core of the Trp-cage, possibly improving the encapsulation of the Trp side chain. The CD spectra did not change in the presence of Ca(II). A conformational change in the grafted loop would only be visible in CD spectroscopy, if the α-helix would change its conformation upon binding. The main fold of the Trp-cage motif was not affected by Ca(II) and without an aromatic exciton coupling there is no signal to monitor small changes.

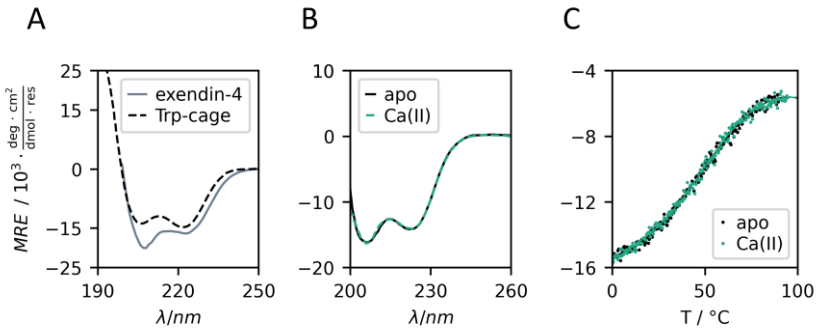


Figure 38. CD spectroscopy of Calmcage at 50 µM concentration in pH 7.2 buffer containing 10 mM MOPS and 150 mM NaCl. A) CD spectra of exendin-4 measured by Wolff et al. (adapted from Figure 5, page 971 [98]) and the exendin-4 derived Trp-cage measured by Neidigh et al. (adapted from Figure 2a, page 427 [10]); B) CD spectra at 20 °C and (C) thermal denaturation curves of Calmcage in the absence (black) and presence of Ca(II) (green).

Binding studies

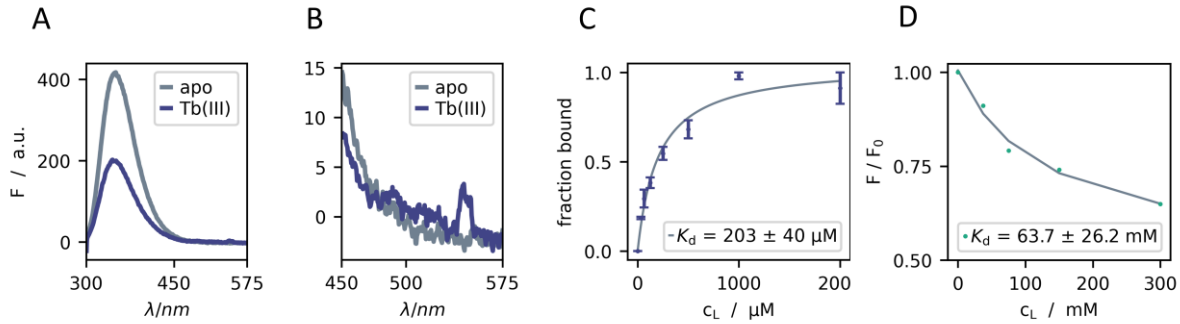
The Tb(III) binding was investigated by FRET. The Fluorescence spectra of Calmcage displayed a small Trp signal at 350 nm (Figure 39). As expected for a peptide with only one Trp sidechain, the intensity of the signal was much lower than for the other designs. The signal intensity was halved by the addition of Tb(III) and produced a small FRET signal at 545 nm. With only Tyr side chains within a radius of 10 Å and the only Trp residue in the chain at a predicted distance of 16 Å, the FRET signal was expected to be low as well. [91] Figure 39C shows a direct titration of Tb(III) to the miniprotein in buffered solution that gave a K_d = 200 ± 40 µM. According to Wallace et al., a fully folded EF-hand lies in the submicromolar range for Tb(III). [99] The lower binding strength suggested a partially folded but functional EF-hand domain.

The Ca(II) binding of calmcage was analyzed by competitive binding experiments. [75] As Tb(III) is usually bound stronger than Ca(II), a high micromolar to millimolar K_d-value was expected. [6, 49] The miniprotein was prepared according to protocol and first analyzed in a competitive experiment with Br_2-BAPTA (K_d = 1.4 µM). The miniprotein was not competing with the dye. A second competitive experiment with BAPTA-5N (K_d = 18 µM) led to the same result.

A new competitive experiment with Tb(III) was started. [7] A buffered solution of 5 µM miniprotein with an excess of 200 µM Tb(III) and varying concentrations of Ca(II) was equilibrated overnight. As shown in Figure 39D, millimolar amounts of Ca(II) in the solution finally decreased the Tb(III) FRET signal. The data were analyzed with *DynaFit* for a K_d = 64 ± 26 mM. [92] The Ca(II) binding site was successfully designed but orders of magnitude weaker than anticipated, even weaker than the laminin site. [49, 50, 54]

Future experiments should test Calmcage for its carbohydrate binding properties in Ca(II)-containing solution and aim to optimize the weak binding site. One issue could be the interface

of the EF-hand's *N*-terminal Val and the Trp-cage's *C*-terminal Ser. The simulation positioned both residues close to but not interacting with each other. A redesign could mutate both for more attractive force between the termini either to similarly polar residues e.g. Thr, or to oppositely charged residues that form a salt bridge.



E

Tb(III)-binding	Calmcage	EF61_dIG8-CC	CaM-CD2	CaM
K_d [M]	$2.0 \pm 0.4 \cdot 10^{-4}$	$2.67 \cdot 10^{-4}$	$2.5 \pm 0.6 \cdot 10^{-5}$	$< 2 \cdot 10^{-8}$
binding domain	EF-hand 2	EF-hand 3	EF-hand 2	EF-hand 1-4
source		(Chidyausiku 2022)	(Ye 2005)	(Wallace 1982)

F

Ca(II)-binding	Calmcage	WWturn	CaM	Laminin
K_d [M]	$6.37 \pm 2.62 \cdot 10^{-2}$	$7.9 \pm 3.4 \cdot 10^{-6}$	$5.3\text{-}11.8 \cdot 10^{-7}$	$5\text{-}300 \cdot 10^{-6}$
binding domain	EF-hand 2	De novo	EF-hand	LG3-5
source		4.2.2	(Zhang 2012)	(Paulsson 1988)

Figure 39. Fluorescence spectroscopy and Tb(III) binding study of Calmcage in pH 7.2 buffer containing 10 mM MOPS and 150 mM NaCl at 5 µM peptide concentration. A) Fluorescence spectra in the absence (grey) and presence of Tb(III) (purple) and (B) zoom-in on the Tb(III) signal; C) titration of Tb(III) with fluorescence measured at 545 nm; D) competitive binding experiment of 100 µM Tb(III) and 5 µM Calmcage with varying concentrations of Ca(II) monitored in fluorescence spectroscopy at 545 nm; E) table of proteins with grafted EF-hand domains with EF61_dIG8-CC by Chidyausiku et al. [7] and CaM-CD2 by Ye et al. [6] and their respective dissociation constants for Tb(III) compared to Calmcage and CaM; [99] F) table of Ca(II)-binding proteins with WWturn (4.2.2), CaM EF-hands [49], Laminin LG domains 3-5 [50] and their respective dissociation constants for Ca(II) compared to Calmcage.

4.3.3 Calmzip

Design Strategy

Tryptophan zippers (Trpzip) are small monomeric β-hairpin motifs that fold cooperatively in water. They are named for their stabilizing rows of indole rings from Trp residues. Trpzips feature two antiparallel β-strands connected by a β-turn. [35] The most thermostable, Trpzip3 (T_m = 79 °C), contains a D-Pro and was therefore not considered as a scaffold. Trpzip2 (T_m = 72 °C) and Trpzip4 (T_m = 70 °C) are of similar thermostability and consist only of L-amino acids. With its longer sequence and more polar turn, Trpzip4 (Figure 40) was chosen for design. [25, 100]

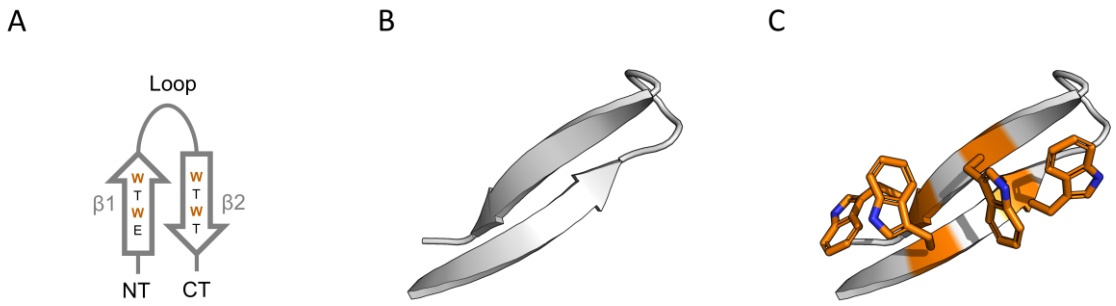
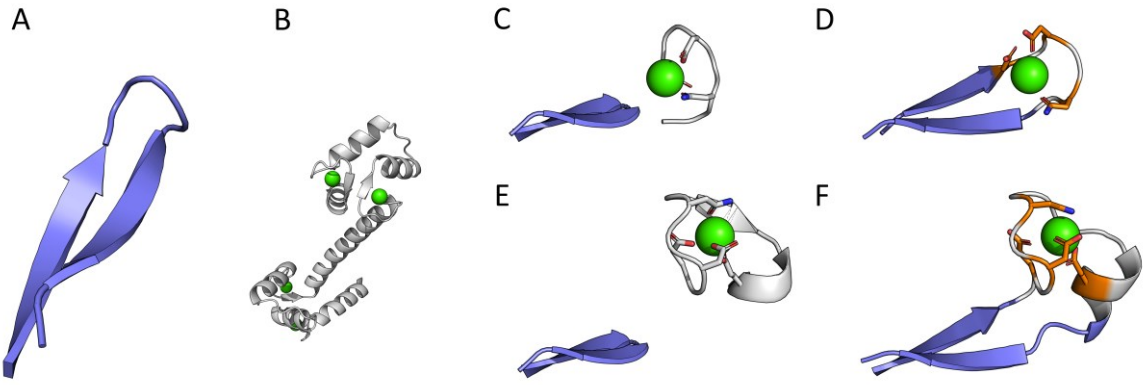


Figure 40. Tryptophan zipper 4 (pdb: 1LE3). A) Schematic representation of the secondary structure motifs with *N*-terminus (NT), β-sheets 1-2 (β), loop and *C*-terminus (CT), and all amino acids of the hydrophobic core marked in orange; B) ribbon representation of Trpzip4; C) Trpzip4 with the amino acids of the hydrophobic core shown and marked in orange. [100]

The domain chosen for grafting was again the EF-hand of CaM Loop 2. As shown in Figure 41, the Calmzip miniproteins were designed using the β-hairpin motif as an end cap to hold the EF-hand together. The loop-helix domain coordinates Ca(II) with three loop side chains and one helix side chain. [49] The design of Calmzip1 (Figure 41C-D) was therefore based on the loop region only. The six amino acid loop of Trpzip4 was replaced by the six amino acid EF-hand loop (Asp 56 to Gly 61). Predictions with an additional Thr spacer at the *C*-terminal end of the loop led to better results with both *ColabFold* and *OmegaFold*. [94, 95]



G

Trpzip4	GEWTW	-DDATKT-	WTWTE
CaM loop2		55-VDADGNGTIDFPEFLT-70	
Calmzip1	G**K**W**K**W	-[DADGNG]T-	W**E**WTE
Calmzip2	G**K**W**K**W	[**V**VDADGNGTID**Y**PE**Y**LT**G**]T-	W**E**WTE

Figure 41. Intuitive Design of Calmzip. A) Crystal structure of Trpzip4 (pdb: 1LE3); B) crystal structure of rat Calmodulin (pdb: 3CLN); C) model of the grafted partial EF-hand 2 of Calmodulin (grey) and Trpzip4 without a loop (blue); D) *ColabFold* Model of Calmzip1 (blue) with the grafted loop in grey and binding amino acids in orange; E) model of the grafted EF-hand 2 of Calmodulin (grey) and Trpzip4 without a loop (blue); F) *ColabFold* model of Calmzip2 (blue) with the grafted domain in grey and binding amino acids in orange; G) sequences of Trpzip4, the EF-hand of Calmodulin Loop 2 and the Calmzip miniproteins with mutations in bold and grafted sequences boxed. [25, 80]

A second design based on the full 12 amino acid EF-hand was attempted for Calmzip2 (Figure 41E-F). Predictions for different loop lengths either led to stable β-sheet or helical models. A combination of a β-sheet miniprotein with a loop-helix motif was only achieved with the addition of several spacers to both sides of the loop. To increase the strength of the β-sheet end cap,

the *N*-terminal β-strand was mutated to a positive charge (Glu 2 to Lys, Thr 4 to Lys) and the *C*-terminal β-strand was mutated to a negative charge (Thr 26 to Glu). The Phe residues of the EF-hand were mutated to Tyr for higher solubility and a possible exciton coupling with the Trp side chains of the β-sheet.

Synthesis and Purification

Calmzip1 was synthesized according to CSY protocol on H-Rink amide polystyrene resin with CSY-protected Asp in position 8 for the Asp-Gly sequence. A solution of the CSY-protected product began to thicken into a gel-like consistency after a few minutes. Higher dilution allowed for purification by HPLC and CSY-deprotection in DMF for the final product.

Calmzip2 was synthesized according to standard protocol on TG XV RAM resin with OMpe-protected Asp in position 10 for the Asp-Gly sequence. The product was purified by HPLC in multiple rounds. When the final product was dissolved in water for aliquotation, the solution became turbid. The solution became clear upon dilution to 20 µM concentration.

Folding and Thermal stability

The parent peptide, Trpzip4 (Figure 42), folds into a β-hairpin motif with a negative band at 213 nm, a positive band at 228 nm and a two-state unfolding at T_m = 74 ± 1 °C. [100] The extreme twisting of the β-sheet intensifies the negative band and the strong exciton coupling with a positive band at 228 nm is caused by an interaction of the four Trp residues. [25]

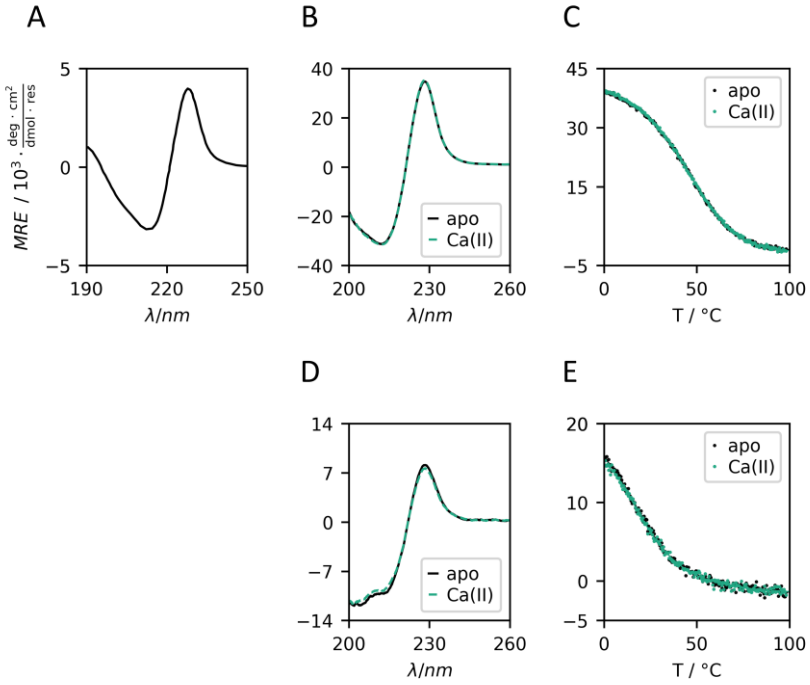


Figure 42. CD spectroscopy of Calmzip1 and Calmzip2 in pH 7.2 buffer containing 10 mM MOPS and 150 mM NaCl in the absence (black) and presence (green) of Ca(II). A) CD spectrum of Trpzip4 measured by Streicher et al. (adapted from Figure 1, page 30 [100]); B) CD spectra at 20 °C and (C) thermal denaturation curves of Calmzip1; D) CD spectra at 20 °C and (E) thermal denaturation curves of Calmzip2.

CD spectra at 20 °C of Calmzip1 (Figure 41B) revealed a reversible similar-to-native folding with a positive band at 229 nm and a negative band at 215 nm. Thermal denaturation experiments (Figure 41C) showed a broad two-state unfolding with a melting temperature of T_m = 52 ± 0.5 °C. [73] The loss in thermal stability can be attributed to the elongation of the loop by one amino acid. This could extend the distance between the β-strands, change the type of β-turn present and weaken the folding. [11] No conformational change was visible upon Ca(II) addition.

The spectra of Calmzip2 (Figure 41D) displayed a positive Trp exciton band at 229 nm and two negative bands at 205 nm and 215 nm. A signal near 200 nm is commonly found for disordered structures, suggesting a partially unfolded peptide as expected due to the long loop. An α-helical contribution with its typical two negative bands at 208 nm and 222 nm was deemed less likely. [72] Thermal denaturation experiments (Figure 41E) supported this conclusion with a T_m = 18.5 ± 1 °C for a broad transition that already set in at 0 °C. [73] The short β-sheet of Trpzip4 seemed unable to support a loop of this size.

Binding studies

As shown in Figure 43A, the Ca(II) binding of Calmzip1 was investigated by Isothermal Titration Calorimetry (ITC). A 25 µM peptide solution was transferred to the sample chamber of a bomb calorimeter and titrated with a 20 mM Ca(II) solution. The heat difference upon addition was recorded and used to estimate the binding strength. The *Malvern* software concluded the peptide to be non-binding. When the data set was fitted as a binding protein, the experiment gave a K_d = 3 ± 5 mM. Given its low binding affinity, or non-binding nature, redesigning Calmzip1 seemed more appropriate than synthesizing a new batch for more detailed analysis.

The binding site of Calmzip2 was investigated by Tb(III) FRET (Figure 41B-C). The large Trp signal at 350 nm was still present above the melting temperature. In the presence of Tb(III), it was reduced and blue-shifted, suggesting a more hydrophobic environment upon binding. The FRET signal at 545 nm was well visible likely due to the short distance between donor and acceptor. [91] Even though the structure was strongly destabilized by the grafted domain, it retained at least partial functionality. This warranted a redesign of the miniprotein for long loop closure.

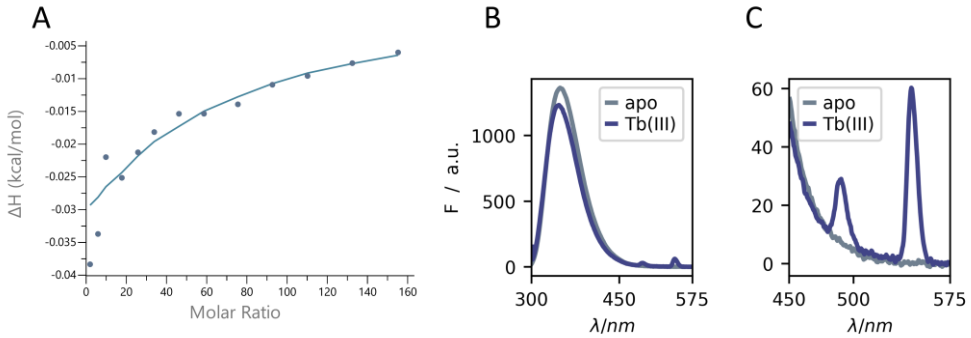


Figure 43. Isothermal titration calorimetry of Calmzip1 and Fluorescence spectroscopy of Calmzip2 in pH 7.2 buffer containing 10 mM MOPS and 150 mM NaCl. (A) ITC spectrum for 25 µM Calmzip1 titrated with 20 mM Ca(II); (B) fluorescence spectra in the absence (grey) and presence of Tb(III) (purple) and (B) zoom-in on the Tb(III) signal for Calmzip2.

Redesign for long loop closure

To install the full EF-hand domain in a β-hairpin structure, a stronger β-sheet cap was required. Anderson et al. designed a β-sheet cap system (C4, Figure 44) that can close 16-mer loops. The β-sheet is stabilized by indole stacking from Trp residues and a salt bridge end cap (Arg 1

and Glu 32). With its eight amino acid β-strands, the β-sheet also includes several Thr and hydrophobic residues with a high β-propensity. [11]

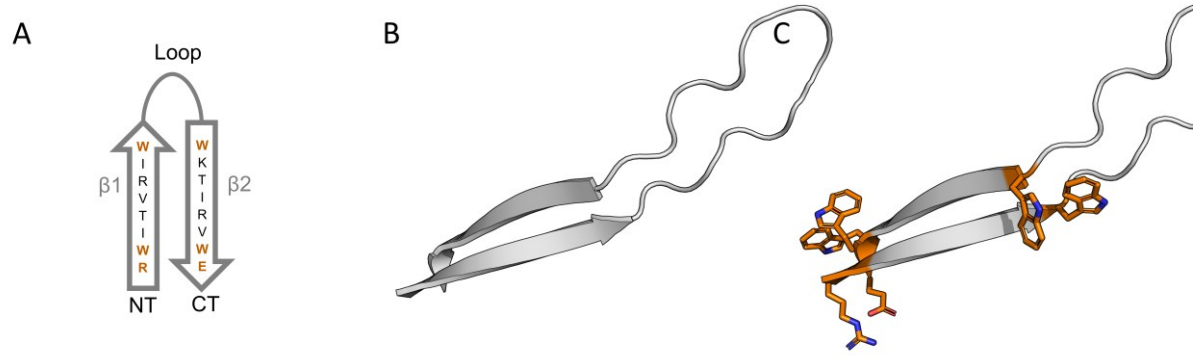
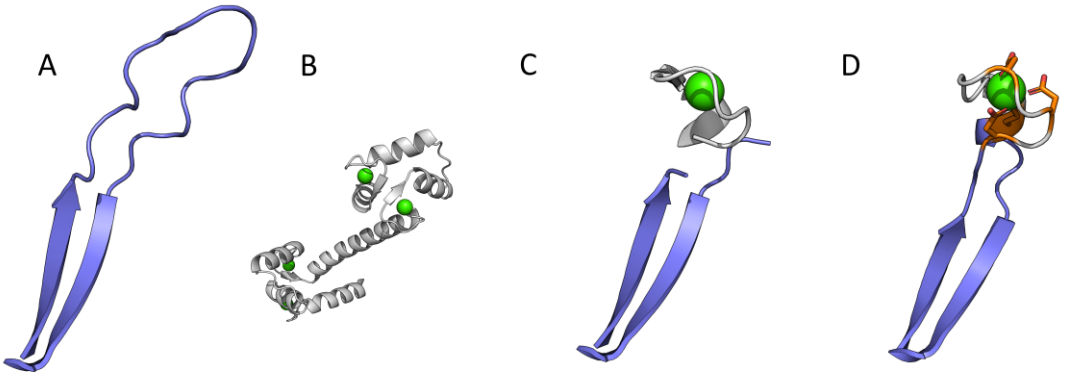


Figure 44. β-Sheet cap system C4 by Anderson et al. (*OmegaFold* model) A) Schematic representation of the secondary structure motifs with *N*-terminus (NT), β-sheets 1-2 (β), loop and *C*-terminus (CT), and all amino acids of the hydrophobic core marked in orange; B) ribbon representation of the C4 system with a 16-mer loop; C) C4 system with the amino acids of the hydrophobic core shown and marked in orange. [11, 95]

Direct replacement of the 16-mer loop by the 12-mer EF-hands of CaM was predicted as an unfolded loop closed by C4 by *ColabFold* and *OmegaFold*. [94, 95] The addition of spacers of varying lengths on both ends of the loop improved the predictions. The best result was predicted for a 16-mer loop with one Gly spacer at the *N*- and three Ala spacers at the *C*-terminus of the EF-hand of CaM Loop 3. The model was named Calmzip3 and prepared for synthesis.



E

C4	RWITVRIW	-GGGGKKGGGGKKGGGG-	WKTIRVWE
CaM loop3		93-DKDGDGYISAAE-104	
Calmzip3	RWITVRIW	-G**DKDGDGYISAAE**AAA-	WKTIRVWE
Calmzip3.1	RWITVRIW	-G**DKDGDGYISAAE**RAA-	WKTIRVWE

Figure 45. Intuitive Design of Calmzip3. A) *OmegaFold* model of the β-sheet cap system C4 with a 16-mer loop by Anderson et al.; [11, 95] B) crystal structure of rat Calmodulin (pdb: 3CLN); C) model of the grafted partial EF-hand of Calmodulin Loop 3 (grey) and the β-sheet cap (blue); D) *ColabFold* model of Calmzip3 (blue) with the grafted loop in grey and binding amino acids in orange; [94] E) sequences of β-sheet-cap C4 with 16-mer loop, the EF-hand of Calmodulin Loop 3 and the Calmzip3 miniproteins with mutations printed bold and grafted sequences boxed. [11, 80]

Synthesis and Purification of Calmzip3

Calmzip3 was synthesized according to standard protocol on TG XV RAM resin with OMpe-protected Asp coupled for the Asp-Gly sequences in positions 12 and 14. Like for Calmzip2,

the solution of the final product in water became turbid after a few minutes. The solution became clear after sonication or mixing by vibration. It instantly aggregated when centrifuged. Calmzip3 likely formed electrostatic aggregates. The design positioned the negative charge in the loop and the positive charge in the β-sheet. With the *C*-terminal amide, the net charge of the miniprotein in neutral solution was around zero (Figure 46). [86] This likely led to peptide assemblies that form aggregates over time. [43, 84] Mutating Ala 22 to Arg would increase the net charge and reduce the negative charge in the loop region. This variant, Calmzip3.1, would likely not aggregate in solution.

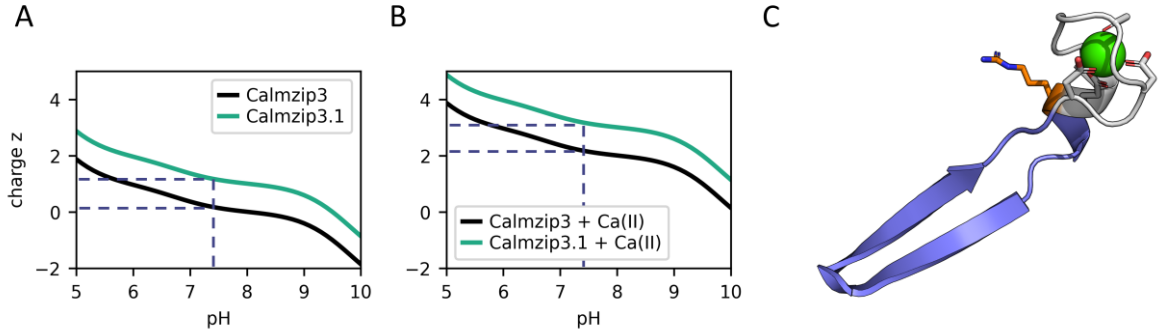


Figure 46. Sequence optimization of Calmzip3. A) *Protpi Peptide Tool* simulation of the net charge of Calmzip3 (black) and Calmzip3.1 (green) at different pH values in the absence and (B) in the presence of Ca(II) with net charge at pH 7.4 marked (dashed line); [86] (C) *ColabFold* model of Calmzip3.1 with Arg 24 shown in orange. [94]

Calmzip3.1 was synthesized according to standard protocol on TG XV RAM resin with OMpe-protected Asp coupled for Asp 12 and Asp14. The product was purified by HPLC for a purity of 90 % with a formylated deletion product as a contaminant of almost identical polarity. The miniprotein was dissolved in pH 7.2 buffer containing 50 mM potassium phosphate and 150 mM sodium chloride without aggregation even after centrifugation. The sample was injected onto a size exclusion chromatography (SEC) column but could not be eluted. It was postulated that the miniprotein was lost on the Sephadex size exclusion resin that contains cross-linked dextran. Other Ca(II) binding peptides have also been lost in this system, hinting at a strong material interaction and carbohydrate binding properties.

Structure and Thermostability

The structure and stability of Calmzip3 was analyzed by CD spectroscopy. For the parent peptide (Figure 47A), spectra similar to Trpzip4 were found, with two negative bands around 215 nm and 200 nm, and a positive band around 230 nm. [11] The spectra of Calmzip3 in Figure 47B revealed a positive Trp exciton coupling band at 229 nm, a positive band below 210 nm and a negative β-sheet signal centered on 215 nm. The broadened negative band could indicate a helical section in the grafted EF-hand. With its negative bands at 222 nm and 208 nm, a contribution from a short α-helix could broaden the 215 nm signal. [72] The positive band below 210 nm was less resolved due to the buffer system but could indicate an ordered loop section. α-Helices and β-sheets alike give positive signals at 193 nm and 198 nm respectively, that are only visible if no disordered section overpowers the signal. [72] In the presence of Ca(II), all bands increased in intensity, suggesting a conformational change upon binding. This fits well with the known behavior of the EF-hand in CaM, where a partially folded flexible α-helix changes conformation to the EF-hand motif upon binding. [49] Given that the β-sheet of Calmzip3 brought the ends of the EF-hand motif together, the conformational change might be less severe than in the native protein.

Thermal denaturation experiments (Figure 47D) revealed a T_m = 47.5 ± 1.0 °C for the holo- and T_m = 47 ± 3 °C for the apo-protein. [73] While the melting temperature was unchanged by

the presence of Ca(II), the denaturation curves did not overlap. The upper asymptote of the denaturation curve of the apo state showed a strong slope beginning at 0 °C. However, in the presence of Ca(II), the denaturation curve is fully sigmoidal with an almost constant upper asymptote. [73] Additionally, all CD signals of the apo-miniprotein increased in intensity after refolding, while the Ca(II)-bound sample remained mostly unchanged. Figure 47C shows only changes in the less resolved bands below 215 nm. The observed unfolding pathway was centered on the Trp signal from the β-sheet structure, not on the flexible helix-loop motif. The very band that refolded to the same intensity in both samples.

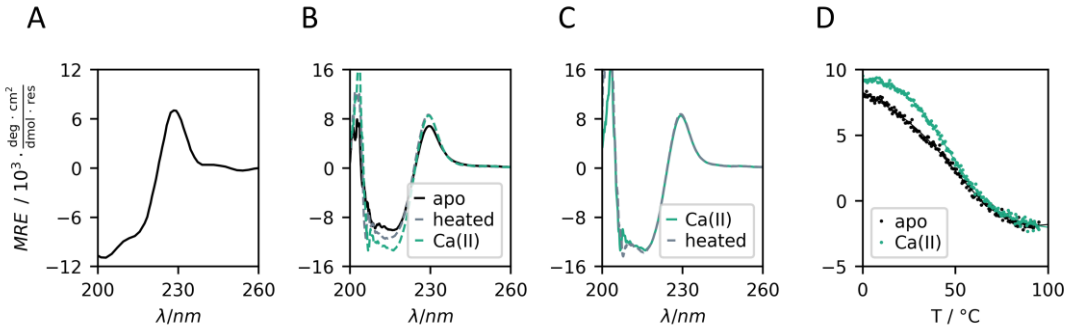
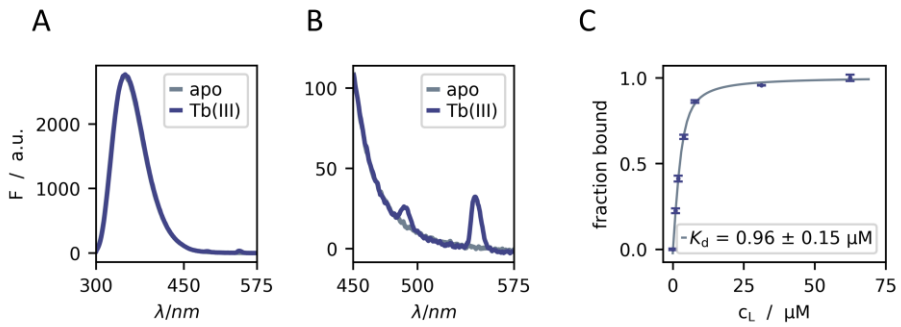


Figure 47. CD spectroscopy of Calmzip3 in pH 7.2 buffer containing 10 mM MOPS and 150 mM NaCl. A) CD spectrum of a 22-mer loop peptide with the C4 β-sheet endcap measured by Anderson et al. (adapted from Figure S3 (12.5 µM), Supplementary Information [11]); B) CD spectra at 20 °C of the apo-miniprotein before (black) and after (grey) heating and in the presence (green) of Ca(II), C) CD spectra at 20 °C in the presence of Ca(II) before (green) and after (grey) heating; D) thermal denaturation curves of Calmzip3 in the absence (black) and presence (green) of Ca(II).

Binding Studies

The Tb(III) binding of Calmzip3 was investigated by FRET (Figure 48). The Fluorescence spectra displayed a strong Trp signal at 350 nm. Upon addition of Tb(III), the signal was only slightly quenched and a small FRET signal appeared. The large Trp signal stemmed from four Trp residues, that are not in contact with the binding site. The weaker FRET signal was likely due to the large distance between donor and acceptor. [91] A direct titration of Tb(III) to the miniprotein in buffered solution gave a K_d = 0.96 ± 0.15 µM. The K_d of each separate EF-hand for Tb(III) is not known but the estimate for CaM with four EF-hands is K_d < 20 nM. [99] Designed single EF-hand CD2 variants by Ye et al. showed a micromolar K_d-value with a trend of site 1 < site 3 < site 2 < site 4. [6] This suggests a fully functional EF-hand for Calmzip3 with a stronger K_d for Tb(III) than any other designed EF-hand protein.

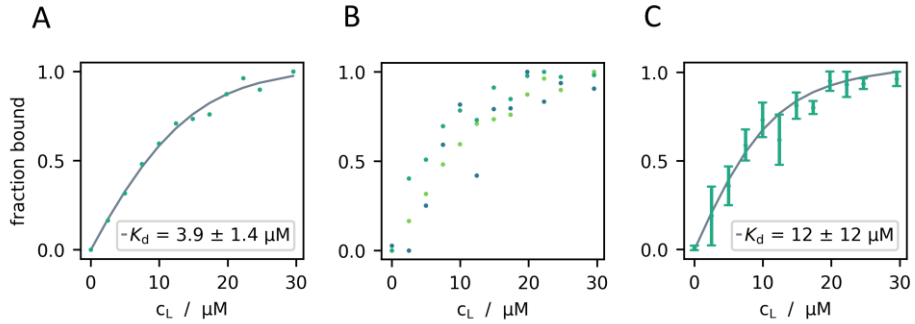
Ca(II) binding was investigated with a competitive dye binding experiment. A solution of Calmzip3 and Br_2-BAPTA was titrated with calcium chloride solution. The experiment was conducted three times at a 10 µM miniprotein concentration. All but one experiment led to data points too scattered for analysis. Figure 49 shows the data points of all titrations overlaid and the analysis of the best data set. The experiment gave a K_d = 3.9 ± 1.4 µM. When all three titrations are averaged, a fit lead to no reliable K_d-values. The design was successful and led to a stronger binding site than WWturn (see 4.2.2) with a K_d in the desired region. Further experiments to characterize the binding to carbohydrates will likely face the same obstacles as the Ca(II) binding analysis. The miniprotein aggregated over time, especially on the glass surface of the cuvette, increasing turbidity and scattering the data. A thorough study of the improved, soluble Calmzip3.1 should be conducted instead.



D

Protein	Calmzip3	Calmcage	CaM-CD2	CaM
K_d [µM]	0.96 ± 0.15	200 ± 40	20 ± 6	< 0.02
Tb(III)-binding domain	EF-hand 3	EF-hand 2	EF-hand 3	EF-hands
source		4.3.2	(Ye 2005)	(Wallace 1982)

Figure 48. Fluorescence spectroscopy and Tb(III) binding study of Calmzip3 in pH 7.2 buffer containing 10 mM MOPS and 150 mM NaCl at 5 µM peptide concentration. A) Fluorescence spectra in the absence (grey) and presence of Tb(III) (purple) and B) zoom-in on the Tb(III) signal; C) titration of Tb(III) with fluorescence monitored at 545 nm; D) table of EF-hand proteins with Calmcage (4.3.2), CaM-CD2, [6] Calmodulin [99] and their respective dissociation constants for Tb(III) compared to Calmzip3.



D

Protein	Calmzip3	WWturn	CaM	Laminin
K_d [µM]	3.9 ± 1.4	7.9 ± 3.4	0.53-1.18	5-300
Ca(II)-binding domain	EF-hand 2	*De novo*	EF-hand	LG3-5
source		4.2.2	(Zhang 2012)	(Paulsson 1988)

Figure 49. Ca(II) binding study of Calmzip3 in pH 7.2 buffer containing 10 mM MOPS and 150 mM NaCl. A) Competitive binding experiment of 10 µM Br_2-BAPTA and Calmzip3 for Ca(II) (green) fitted with *DynaFit* (grey); [92] B) repetitions of (A) shown in different colors; C) mean of three titrations with standard deviation (green) with the fit for dataset (A); D) table of Ca(II)-binding proteins and their respective dissociation constants for Ca(II) with Calmodulin, [49] Laminin, [50] and WWturn (4.2.2) compared to Calmzip3.

4.4 Designed active Ca(II)-binding Miniproteins

In three different approaches: computational design with *Rosetta3*, [4] intuitive *de novo* design, [5] and functional domain grafting, [6] I aimed to design Ca(II)-binding miniproteins with an affinity in the same range as D-type lectins. Laminin, the ECM protein titular for the active LG domains found in D-type lectins, was found to bind Ca(II) with a micromolar K_d-value of K_d = 5-300 µM. [3, 50]

The first approach used the *Rosetta3* modelling suite to incorporate the active site of Laminin LG4 in the sequence of the c-Crk SH3 domain (Slam) and the active site of the Ca(II) sensor protein Calmodulin in the sequence of the hPin1 WW domain (WWcalm). Due to approximations and omissions in the underlying Rosetta energy function, the software does not fare well with small miniproteins or peptides. [58] Models based on the WW domain were particularly poor and warranted a rational redesign on the basis of stability studies conducted within the group. [40] These redesigns lead to a thermostable but inactive WWcalm3 series and the active but aggregating WWcalm4. The aggregating tendencies of WWcalm4 in the presence of Ca(II) likely stemmed from its net charge and should be removed by a mutation to a higher charge in WWcalm4.2. [84, 86] The larger SH3 domain showed a higher compatibility with the modelling suite, but initial models suffered from a difficult purification and low synthesis quality for Slam1 or a synthesis-optimized but inactive SH3 domain variant for Slam2. The most recent design, Slam3 focused on the redesign of the native active site and positioned the new Ca(II)-binding center in the highly charged RT Loop, where a native Na(I) site can be found for the Fyn SH3 domain. [78]

The intuitive *de novo* design was inspired by the CD2.trigger protein designed by Li et al. [5] An exposed complex salt bridge on the miniprotein scaffold of the SH3 or WW domains was changed to a new Ca(II) binding site. The positively charged residues of the chosen salt bridge were mutated to Asp, thus replacing the attractive Coulomb forces with repulsive ones in the absence of Ca(II). Upon binding, the native conformation was expected to reform. Models were created using *ColabFold* and *OmegaFold*. [94, 95] Two adjacent salt bridges were mutated in the SH3 domain for the ScaX model. The model showed a high sensing capacity with Tyr_2 as a FRET donor for Tb(III) but folded into a molten globule structure with extremely poor thermostability. The only exposed complex salt bridge on the WW domain surface was between the *C*- and *N*-termini. Therefore, the circular permutated WW domain by Kier et al. with a similar complex salt bridge was used for the design of WWturn. [8] It showed high thermostability without a stability-function tradeoff and folded into the expected WW domain-like conformation. [85] The model showed a small Tb(III) FRET likely due to the distance to the buried Trp and Tyr residues as evidenced by the Trp signal position in Fluorescence spectroscopy. [91] Competitive binding studies of WWturn with Br_2-BAPTA revealed a K_d = 7.9 ± 3.4 µM in the desired affinity range of D-type lectins. [3, 50]

The final design approach was the grafting of functional domains onto robust miniprotein scaffolds such as the WW domain, the Trp-cage and β-hairpin peptides. [54] Functional domains such as EF-hand or EF-hand-like motifs were incorporated in the loops or termini of the miniproteins and optimized using *ColabFold* and *OmegaFold*. [94, 95] The EF-hand-like motif of a β-propeller lectin from *Psathyrella Velutina* was grafted into Loop 1 of the WW domain for an active Ca(II) sensor protein with a high-intensity Trp-Tb(III) FRET signal in WWfan. [96] However, CD spectroscopy revealed a native-like conformation but poor thermostability. EF-hand 2 of CaM was grafted onto the *N*-terminal helix of Trp-cage for the design of Calmcage. The model folded into a native-like conformation with an increased thermostability. [10] Binding studies for Tb(III) revealed a low affinity of K_d = 200 ± 40 µM, befitting the low Trp-Tb(III) FRET intensity. In a competition binding study of Tb(III) and Ca(II) for Calmcage, a K_d-value of K_d = 64 ± 26 mM was found. With an affinity much lower than D-type lectins, Calmcage was deemed a weaker candidate for lectin activity but may benefit from

a redesign of the *N*- and *C*-terminal residues to improve the folding of the EF-hand motif. [3, 50]

EF-hands of CaM were also grafted into the loop of Trpzip4. [25, 100] The resulting Calmzip models suffered from high thermostability with little to no activity for the grafting of a partial EF-hand in Calmzip1 or low thermostability with a high-intensity Trp-Tb(III) FRET signal for the grafting of a complete EF-hand in Calmzip2. The Trpzip4 scaffold was deemed too short for the large EF-hand loop and replaced in the design of Calmzip3 by the β-sheet end cap system C4 designed by Anderson et al. specifically for long loop closure. [11] The model showed similar-to-native folding at high thermostability but aggregated at high concentration or upon centrifugation. Binding studies of Calmzip3 by Trp-Tb(III) FRET revealed the lowest K_d-value reported for any grafted EF-hand with K_d = 0.96 ± 0.15 µM for Tb(III). [6, 54] Competitive binding studies of Calmzip3 with Br_2-BAPTA for Ca(II) gave similar results with a K_d-value of K_d = 3.9 ± 1.4 µM within the order of magnitude of both CaM and Laminin LG4.[49, 50, 75] However, repetition of the experiment was hampered by the aggregative tendencies of Calmzip3, highlighting the importance of a full characterization of Calmzip3.1.

I successfully designed three miniproteins with a Ca(II)-binding affinity similar to LG domains of D-type lectins (see Table 3): WWturn – on the basis of the circular permutated WW domain by Kier et al., [8] Calmcage – on the basis of Trp-cage designed by Neidigh et al., [10] and Calmzip3 – on the basis of C4 designed by Anderson et al. [11] All three models showed a high thermostability and native-like conformation, making them applicable at elevated temperatures (37 °C) required for cell studies. [17] With a larger coordination site, the presumably fully folded EF-hand motif of Calmzip3 showed a higher Ca(II) affinity than the *de novo* site of WWturn. The lower affinity of Calmcage likely stemmed from an incomplete folding of the EF-hand motif and could be improved upon if necessary. All three models showed the desired stability and functionality for Ca(II)-binding and should be investigated for lectin activity.

Table 3. Designed Ca(II)-binding miniproteins with melting temperature (T_m) and dissociation constant for Ca(II) ($K_d^{Ca(II)}$).

	scaffold	Ca(II)-binding domain	T_m [°C]	$K_d^{Ca(II)}$ [M]
Calmzip3	C4 β-hairpin [11]	CaM EF-hand 3	52.0 ± 0.5	$3.9 \pm 1.4 \cdot 10^{-6}$
WWturn	cp34 WW [8]	*de novo*	55.5 ± 0.5	$7.9 \pm 3.4 \cdot 10^{-6}$
Calmcage	Trp-cage [10]	CaM EF-hand 2	50 ± 3	$6.4 \pm 2.6 \cdot 10^{-2}$

4.5 Computational Design of a SOD-active Miniprotein

Design Strategy

With the difficulties faced in the computational design of functional Ca(II)-binding miniproteins with *Rosetta3* [4] on the basis of SH3 and WW domains, it was aimed to install a Cu(II)-binding site in the SH3 domain as a proof of concept. Recent advances in our group had already shown that a transition metal binding His_3 site can be intuitively designed on the WW domain. [101] Therefore, the human Cu,Zn superoxide dismutase 1 (SOD1, pdb: 3CQP, see Figure 50,) was chosen with its His_4 active site and structural similarity to the SH3 domain, as our natural model. SOD1 is a homodimer with each subunit folded into an antiparallel eight-stranded β-barrel structure arranged in a Greek key motif. Each subunit features a catalytically active Cu,Zn site between two loops, with each metal ion in a tetrahedral coordination. The structure-relevant Zn(II) ion is complexed by three His side chains (His 63, His 71, His 80) and one Asp (Asp 83) residue. The catalytically active Cu(II) ion is complexed by four His residues (His 46, His 48, His63, His 120). His 63 bridges the metal centers in the oxidized state and binds only to Zn(II) in the reduced state. [102] In a recently investigated fungal Cu-only SOD with a similar four His binding site, the role of Zn(II) was adopted by a neighboring Glu instead. [103] As the name implies, SOD enzymes disproportionate superoxide into molecular oxygen and hydrogen peroxide and work as antioxidants in most living cells. [59]

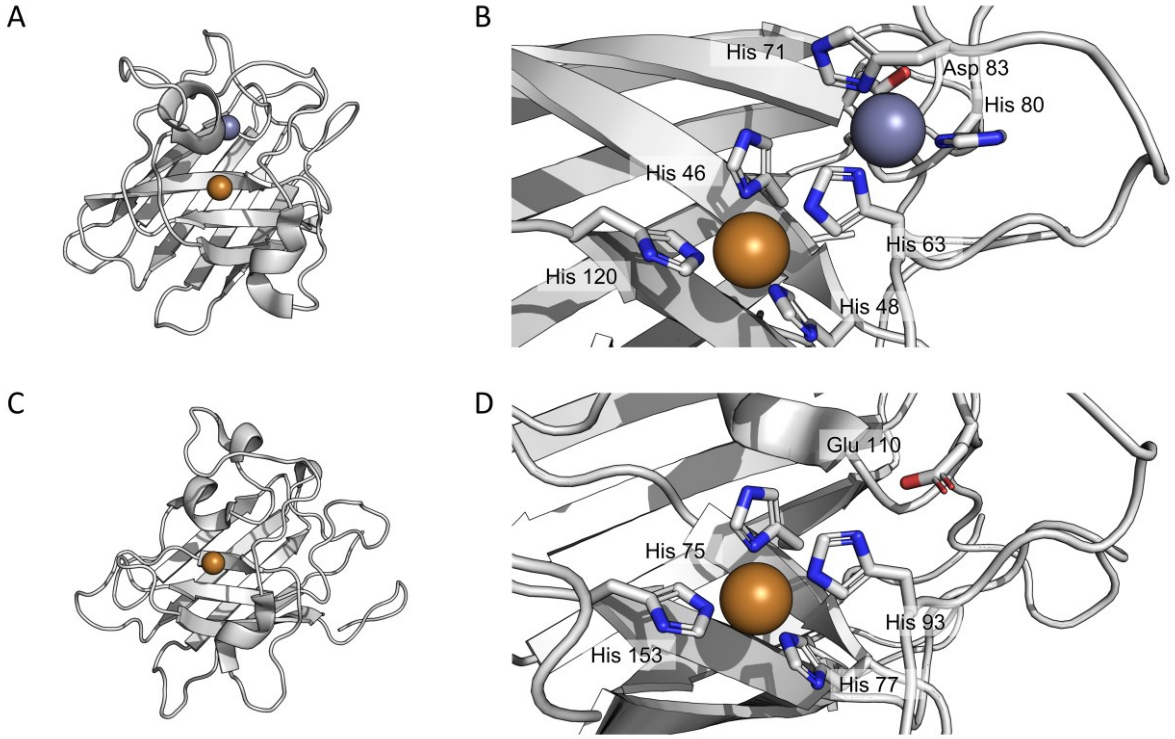
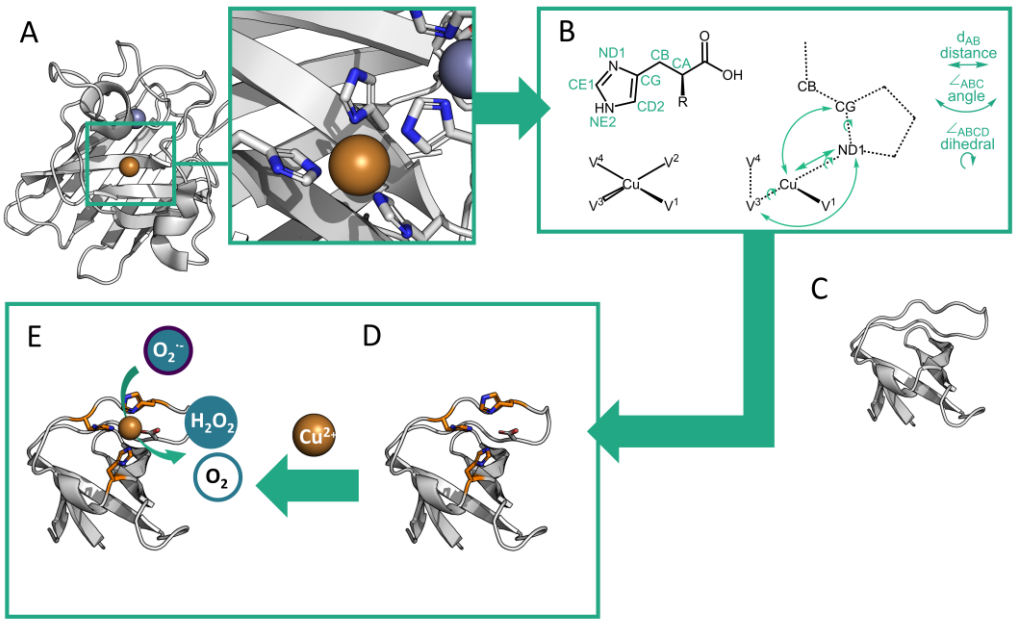


Figure 50. Ribbon presentation of SOD1 (pdb: 3CQP) and SOD5 (pdb: 4N3U). A) Complete SOD1 with Cu(II) (orange) and Zn(II) (grey) bound; B) SOD1 active center with Cu(II)- and Zn(II)-binding amino acids shown; C) complete SOD5 with Cu(II) (orange) bound; D) SOD5 active center with Cu(II)-binding amino acids shown. [103, 104]

Given that Zn(II) is not required for the superoxide dismutase activity, the design presented in Figure 51 was focused on installing Cu(II) binding with a minimalistic His_3 binding site. The geometry was modeled after the three His residues that exclusively bind Cu(II) (His 46, His 48, His 120). For each coordinating residue, six parameters described the distance, angles and dihedral angles of the atoms coordinating Cu(II) and their direct neighbors as shown exemplary for one His in Figure 51B. One-atomic ligands define their orientation by virtual atoms (V1-V4) that have no further impact on the calculation.

Using *Rosetta3 Match*, the active site was realized on the SH3 domain scaffold. One sequence, named ‘match1’, was output multiple times and included mutations of two residues conserved for binding proline-rich sequences, the functionality of native SH3 domains. [69] The replacement of the native function of a peptide scaffold with a new one has already proven successful in the literature. [37] The Cu(II) coordination sphere was completed using *Rosetta3 Design*. The lowest-scoring models were grouped by the position of the active site. Only the designs based on ‘match1’ succeeded in introducing a fourth ligand. Herein, Glu 17 replaced the native SH3 functionality and positioned the active site in a cavity. The lowest-scoring of these sequences, ‘design1’, was optimized for synthesis by mutating Asp in aspartimide-promoting sections to Glu or Asn, and Val 16 to Gln to better resemble the native RT Loop. The resulting sequence, ‘SO1’, was prepared for synthesis.



F

c-Crk SH3	AE-YVRA-LFDFNGNDEEDLPFKKGD-ILRIRD-KPEEQ-WWNAE-DSEGK-RGMIP-VPY-VEKY
M48L	AE-YVRA-LFDFNGNDEEDLPFKKGD-ILRIRD-KPEEQ-WWNAE-DSEGK-RG[L]IP-VPY-VEKY
match1	AE-YVRA-LFD[H]NG[H]DEEDLPFKKGD-ILRIRD-KPEEQ-WWNAE-DSEGK-RG**L**I[H]-VPY-VEKY
design1	AE-YVRA-LFD**H**[T]G**H**DE[VE]LPFKKGD-ILRIRD-KPEEQ-[L]WNAE-DSEGK-RG**L**I**H**-V[VL]-VEKY
SO1	AE-YVRA-LFD**HT**G**H**DE[Q]**E**LPFKKG[E]-ILRIRD-KPEEQ-**L**WNAE-[N]SEGK-RG**L**I**H**-V**VL**-VEKY[G]

Figure 51. Computational Design of SO1 with (A) the crystal structure of SOD1 (pdb: 3CQP) and a zoom of the active center, (B) geometric parameters of the Cu(II) center, (C) the crystal structure of the c-Crk SH3 domain (pdb: 1CKB), (D) Rosetta model of SO1 with amino acids of the active center shown, (E) scheme of the active SO1 enzyme and (F) sequences of the SO1 design process. [4, 26]

Synthesis and Purification

SO1 was first synthesized according to standard protocol (90 °C, 110 s) at 50 % scale to account for additional swelling due to high sequence length. As seen in Figure 52, analytical RP-HPLC revealed a crude product of low purity, making subsequent purification very difficult. According to literature, side reactions like epimerization, aspartimide formation or δ-lactam formation become more prevalent at elevated temperature. [105-111] Based on the work of Palasek et al., a protocol with a lower temperature and extended reaction time was used for coupling the amino acids (50 °C, 600 s). [111] A synthesis according to LT50 protocol significantly improved the product quality leading to high yields of pure product.

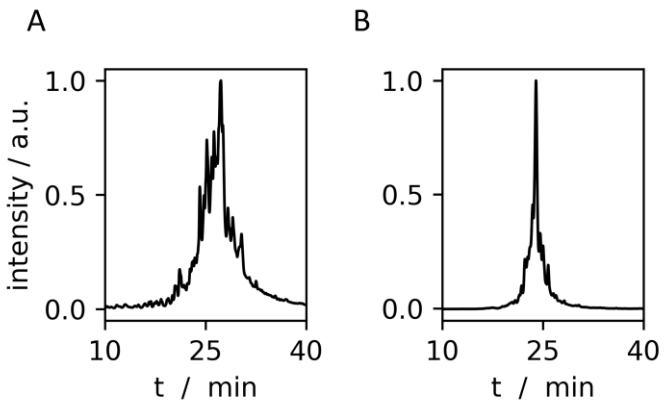


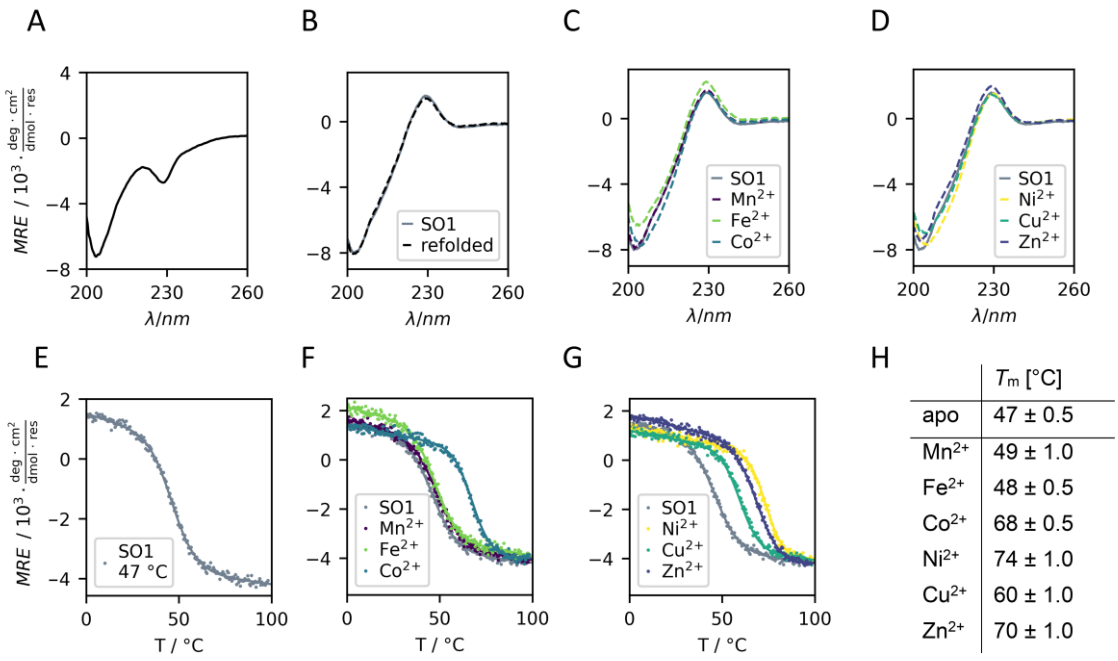
Figure 52. HPLC Chromatograms of SO1 recorded at 280 nm and normalized to the peak of highest intensity with (A) 20-40 % B in A gradient of SO1 synthesized according to standard protocol and (B) 20-50 % B in A gradient of SO1 synthesized according to LT50 protocol.

Structure and Thermostability

The structure and thermal stability of SO1 was analyzed using CD spectroscopy (Figure 53). Unlike the parent protein with its two negative bands at 230 nm and 200 nm, and a positive band at 220 nm, SO1 displayed CD spectra reminiscent of β-hairpin dominant proteins with a negative band at 205 nm and a positive band at 229 nm. [71, 112] The negative band at 205 nm with a shoulder around 215 nm suggested a β-sheet conformation with large disordered sections and rich in β-turns, indicating an SH3 domain structure. [72] The typically negative band at 229 nm resulting from aromatic interactions was replaced by a positive band. This CD signature is not unknown for SH3 domains, as shown by Liu et al. for the human Nck SH3 domain. [112] SO1 contained several mutations of aromatic residues in the RT Loop and 3_{10}-helix of the SH3 domain that interact with Trp 37 to give the exciton coupling band at 229 nm. [113] These residues were conserved for binding – not stability – but may play a part in the tight packing of the RT Loop on the β-barrel. [114]

CD spectra in the presence of first-row transition metal ions may change the intensity of the different bands (Figure 53C-D). In the presence of Mn(II), CD spectra remained unchanged. In the presence of Fe(II) however, the exciton signal increased and the β-sheet signals decreased in intensity. For Co(II), the β-sheet signals broadened and the shoulder around 215 nm became more pronounced. Similar spectra were found in the presence of Ni(II), with a slightly less intense β-sheet band at 205 nm. For Zn(II), spectra similar to Fe(II) were found, with a slightly decreased β-sheet signal and an increased exciton signal intensity. In the presence of Cu(II), the β-sheet band at 205 nm was less pronounced but the remaining spectrum almost perfectly overlapped with the apo-miniprotein's.

Thermal denaturation experiments (Figure 53E-G) revealed a melting temperature T_m = 47 ± 0.5 °C of the apo miniprotein. After denaturation, an additional CD spectrum was recorded, showing reversible and cooperative folding-to-unfolding. [73] The lowered thermal stability compared to the native SH3 domain's T_m = 67 °C may have been a result of a lower packing density around the RT Loop. [69] Binding of bivalent metals in the active centers should increase the packing density and therefore also increase the melting temperature. This behavior was observed in the presence of first-row transition metals. Mn(II) and Fe(II) slightly increased the thermal stability, while Cu(II) brought it to 60 °C. Co(II)- or Zn(II)-bound SO1 showed a thermostability on par with the native SH3 domain and Ni(II)-bound SO1 even increased it to T_m = 74 ± 1.0 °C. Given that the design focused on a tetrahedral coordination sphere for Cu(II), it is no surprise that the metal ions that favor a tetrahedral geometry increase the thermal stability the most. [115]



	T_m [°C]
apo	47 ± 0.5
Mn^{2+}	49 ± 1.0
Fe^{2+}	48 ± 0.5
Co^{2+}	68 ± 0.5
Ni^{2+}	74 ± 1.0
Cu^{2+}	60 ± 1.0
Zn^{2+}	70 ± 1.0

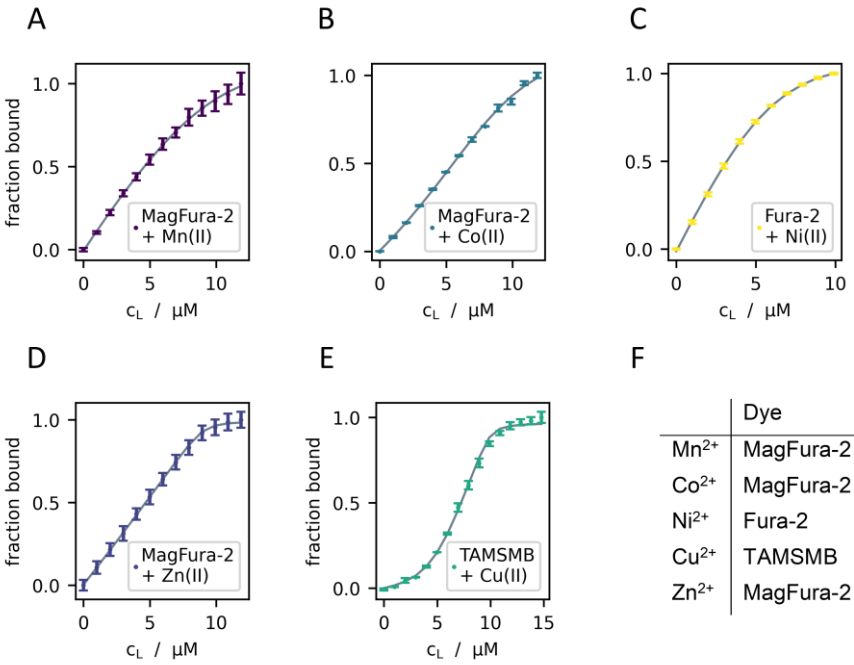
Figure 53. CD spectroscopy of SO1 in pH 7.2 buffer containing 10 mM MOPS and 150 mM NaCl in the absence (grey) and presence of Mn(II) (violet), Fe(II) (leaf green), Co(II) (teal), Ni(II) (yellow), Cu(II) (green), Zn(II) (purple). A) CD spectrum of c-Crk SH3 measured by Pham et al.; [74] B) CD spectra of SO1 at 20 °C before (grey) and after (black) heating; C) CD spectra of SO1 at 20 °C in the absence and presence of Mn(II), Fe(II) and Co(II); D) CD spectra of SO1 at 20 °C in the absence and presence of Ni(II), Cu(II) and Zn(II); E) thermal denaturation curve of apo-SO1; F) thermal denaturation curves of SO1 in the presence of Mn(II), Fe(II) and Co(II); G) thermal denaturation curves of SO1 in the presence of Ni(II), Cu(II) and Zn(II); H) melting temperatures of SO1 in the absence and presence of bivalent transition metals.

Binding studies

The binding affinities of SO1 to Cu(II) and other first-row transition metals was studied in competitive binding experiments (Figure 54). The investigated metal ion was titrated to a buffered solution of SO1 and a metal-binding dye of equal concentration. The change in absorbance was monitored until saturation. [116] The dissociation constant was calculated from the known K_d of the dye as the average of three titration experiments using *DynaFit*. [92]

The binding affinities of SO1 to first-row transition metals (see Figure 54F) follows the Irving-Williams series with each cation binding one order of magnitude stronger than its neighbor in the series: Mn(II) < Co(II) < Ni(II) < Cu(II) > Zn(II). [117] The strongest K_d = 3.60 × 10^{-9} M for Cu^{2+} is evident for the successful design of the envisioned binding site, albeit orders of magnitude weaker than the binding affinity of the natural model SOD1 or the copper-only SOD5 (see Table 4).

As can be seen for SOD5 and Zn(II)-free SOD1, Cu(II)-binding in SODs is much weaker when no secondary cation is present. [118, 119] Native Cu,Zn SOD1 can withstand dialysis with EDTA for days without losing either cation. [120] SOD5 however, loses Cu(II) in EDTA containing solution within 24 hours. [121] The reduced binding affinity of SO1 likely stemmed from two key differences: (1) Zn(II)-free design and (2) a coordination sphere with three His residues instead of four. Even without Zn(II), there is still a difference in binding affinity of six orders of magnitude between SO1 and the Cu-only SOD5. As previously shown by Pham et al., β-sheet peptides can show very different K_d-values with a His_3 site based on scaffold flexibility and geometry of the residues. [101, 122] It is possible that the design missing one of the complexing four His, resulted in a weaker binding while maintaining a similar geometry.



	Dye	K_d [M]
Mn^{2+}	MagFura-2	$1.09 \pm 0.09 \times 10^{-6}$
Co^{2+}	MagFura-2	$4.75 \pm 0.66 \times 10^{-7}$
Ni^{2+}	Fura-2	$4.62 \pm 1.05 \times 10^{-8}$
Cu^{2+}	TAMSMB	$3.60 \pm 0.25 \times 10^{-9}$
Zn^{2+}	MagFura-2	$3.38 \pm 0.05 \times 10^{-8}$

Figure 54. Competitive binding of SO1 and a dye to bivalent transition metal cations in pH 7.2 buffer containing 10 mM MOPS and 150 mM NaCl. Competitive binding experiments of SO1 with (A) Mag-Fura-2 and Mn(II), (B) Mag-Fura-2 and Co(II), (C) Fura-2 and Ni(II), (D) Mag-Fura-2 and Zn(II) and (E) TAMSMB and Cu(II); F) table with cations, dyes and dissociation constants fitted with *DynaFit*. [92]

Table 4. Dissociation constants of SODs for Cu(II).

Protein	conditions	K_d [M]
Human Cu,Zn-SOD1	2M urea	7.3×10^{-17} [118]
Bovine Cu,Zn-SOD1	Zn(II) free	2.5×10^{-16} [119]
Fungal Cu-SOD5		3.2×10^{-15} [121]

SOD activity studies

A successful design of the SOD active center in SO1 should result in an active mini enzyme. To investigate the superoxide dismutase activity of SO1, an indirect assay was employed as shown in Figure 55. Superoxide was generated from Hypoxanthine by a Xanthine - Xanthine Oxidase (XO) system. [59] The superoxide radicals reduced the water-soluble tetrazolium (WST) to a formazan dye monitored by spectroscopy. In the presence of SOD, the dye formation was inhibited by the enzyme activity. The activity of Cu-SO1 was examined in competition with WST-1 by the endpoint of formazan generation after 30 min and in competition with 2,3-bis-(2-Methoxy-4-Nitro-5-Sulfophenyl)-2H-Tetrazolium-5-Carboxanilid (XTT) by the change in the rate of formazan generation within 90 s. Both experiments were conducted in pH 7.8 buffer containing 50 mM potassium phosphate. A control titration in the changed buffer system showed no impact on the K_d of Cu-SO1.

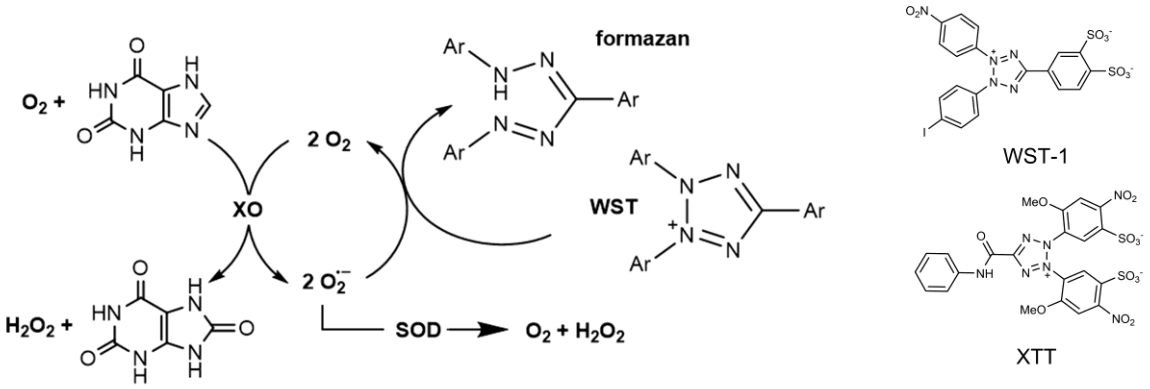


Figure 55. Reaction scheme for an indirect SOD activity assay with structures of the formazan dyes WST-1 and XTT.

The endpoint experiments were based on an assay kit by Merck. The kit was optimized for native SOD1 with a high EDTA concentration in pH 7.8 buffered solution. EDTA was originally added to capture free cations from biological SOD1 samples as it binds Cu(II) with a weaker K_d. [59] The examined Cu-SO1 samples could not withstand EDTA in solution and required an EDTA-free assay. Given that the samples were synthesized chemically, they contained no biological contaminants and free Cu(II) was limited by an excess of miniprotein. The assay was benchmarked with a bovine SOD1 sample. The intensity of WST-1 absorbance increased over time until saturation was reached. The saturation point, seen for all curves in Figure 56B, likely indicates the point at which WST-1 is fully converted. Varying concentrations of Cu-SO1 were assayed and plotted against the ratio (A^0/ A^i) of the final absorbance in the absence of SOD (A^0) over the final absorbance in presence of the sample (A^i). The half maximal inhibitory concentration (IC_{50}) was found at the concentration with a ratio A^0/ A^i = 2. This concentration, $IC_{50,WST\text{-}1} = 3.2 \pm 0.4 \times 10^{-8}$ M, competed equally to one unit bovine SOD1 of a Merck activity assay kit. [59] Control experiments with free Cu(II) at the calculated free concentration (8.4×10^{-9} M) showed a minor but noticeable contribution to the SOD activity. Control experiments with apo-SO1 showed WST-1 absorbance within the error margin of no inhibition. To limit the influence of the known SOD activity of free Cu(II) even further and ensure a representative measurement, rate change experiments were performed at a fourfold excess of SO1 to Cu(II).

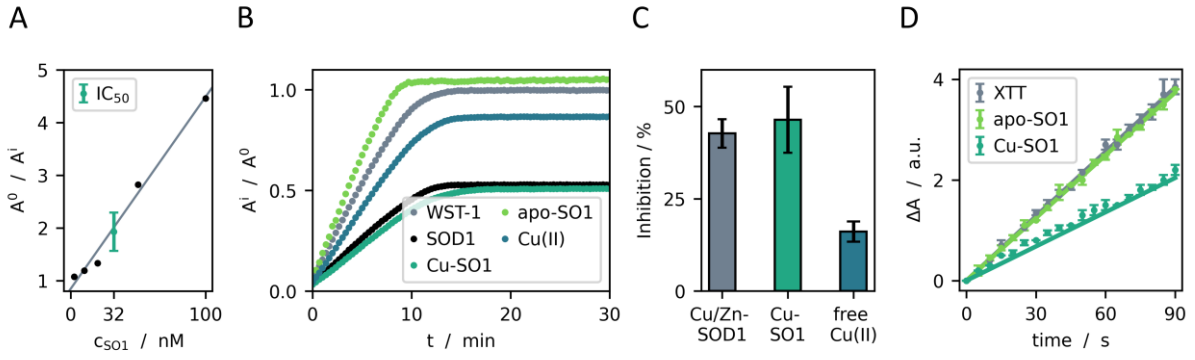


Figure 56. SOD activity plots of SO1. A) Linearized activity rate plot from competitive activity experiments with WST-1 at different concentrations of SO1 (IC_{50} marked in green); B) normalized absorbance chromatograms of WST-1 competing with Cu-SO1 (green), apo-SO1 (leaf green), free Cu(II) (teal) and SOD1 (black); C) bar diagram for the percentual inhibition of WST-1 by 8.4×10^{-9} M free Cu(II) (teal), 1 unit bovine SOD1 (grey) and 3.2×10^{-8} M Cu-SO1 (green); D) absorbance chromatograms of XTT (grey) competing with Cu-SO1 (green) and apo-SO1 (leaf green).

The rate change experiments were conducted according to a protocol by Mathieu et al. in pH 7.8 buffered solution. [15] The rate of formazan generation was monitored for 90 s after the addition of Xanthine oxidase to complete the system. Control experiments with no SOD gave the uninhibited rate of formazan generation. Varying concentrations of Cu-SO1 with a fourfold excess of miniprotein were assayed. The concentration, that halved the rate of formazan generation was the IC_{50}. Expecting a similar reactivity for both WST solutions with a fourfold XTT concentration, the experiments were started at a fourfold $IC_{50,guess} = 1.3 \times 10^{-7}$ M. The final result was found for a slightly higher XTT reactivity, showing an $IC_{50,XTT} = 1.40 \pm 0.22 \times 10^{-7}$ M. Control experiments of apo-SO1 also showed no inhibition.

According to Sutherland et al., Equation 2 can be applied to find the apparent rate constant of a reactant in a competitive experiment. For this, the rate constant of the competitor and its concentration must be known. [123] As the catalytic rate constant of WST-1 was not known, a second experiment with another dye was required. With the catalytic rate constant for XTT with superoxide of $k_{cat} = 8.59 \pm 0.81 \times 10^4$ M^{-1} s^{-1}, [123] an apparent rate constant of $k_{SOD} = 6.14 \pm 0.58 \times 10^7$ M^{-1} s^{-1} was calculated for Cu-SO1.

$$k_{SOD} = \frac{k_{dye} \times c_{dye}}{IC_{50}}$$

Equation 2. Apparent rate constant k_{SOD} of the SOD enzyme with superoxide with catalytic rate constant k_{dye} of the WST with superoxide, dye concentration c_{dye} and half maximal inhibitory concentration IC_{50} of the SOD enzyme.

This marks SO1 not only as the first designed β-sheet miniprotein with SOD activity but also the SOD active miniprotein with the highest activity to date.

Table 5 shows the rate constants of SOD enzymes and the apparent rate constants of SOD mimics, peptides and miniproteins. SO1 showed catalytic activity in the same order of magnitude as OCP1 by Vincent et al. [124] and the Cu(II) complexes presented by Csire et al. [12] with four His coordinating Cu(II). Unlike these examples, SO1 was based on a miniprotein scaffold with a defined three-dimensional structure. It outcompeted the larger helical miniproteins designed by Singh et al. [16] and Mathieu et al. [15] and the Thioredoxin mutant by Benson et al. [14] by one and two orders of magnitude. All of these miniproteins bind the metal center with three His residues.

Only SO1, with an activity that surpasses any SOD mimic of similar sequence length, was designed to replicate the geometry of native SODs. Although native SODs still operate in a league of their own with diffusion limited, 20 to 40 times higher rate constants than SO1. [121, 125] With its high activity and native-like active center, SO1 could serve as a model system to further investigate the working mechanisms of SODs.

Table 5. Apparent rate constants of SOD active complexes and catalytic rate constants of SOD enzymes. SOD active compounds with peptide sequence length, metal-coordinating ligand (β-GCD = mono6-deoxy-6-guanidinocycloheptaamylose, β-CD = cyclodextrin) and the catalytic or apparent rate constant k.

	length	ligand	k [$M^{-1}s^{-1}$]
Bovine Cu,Zn-SOD1	$(152)_2$	His_4	2.3×10^9 [125]
Fungal Cu-SOD5	228	His_4	1.1×10^9 [121]
[Cu(II)(L)$(H_2O)_2$(β-GCD)]	-	β-GCD	1.8×10^8 [13]
[Cu(II)(L)$(H_2O)_2$(β-CD)]	-	β-CD	9.9×10^7 [13]
Ac–S3H4–NH_2 $[Cu^{II}L11]^{2+}$	7	His_4	6.5×10^7 [12]
Cu-SO1	**58**	**His_3**	**6.1×10^7**
OCP1 (Cu^{2+})	8	His_4	2.4×10^7 [124]
Cu(II)/$(HKHGPG)_2$/Zn(II)	12	His_4	8.2×10^6 [126]
S2-Thioredoxin (Fe^{2+})	109	His_3	6.2×10^6 [14]
GRα$_3$D H_3 (Cu^{2+})	98	His_3	3.0×10^6 [15]
MHB (Mn^{2+})	63	His_3	3.7×10^5 [16]

5 Conclusion

Tissue Engineering requires functional materials for cells to grow on. [1] Currently, these matrices are either based on natural materials and suffer from immunogenicity or reproducibility issues, or based on synthetic materials that require functionalization for cell adhesion. The functionalization of hydrogels with cell adhesion proteins shares the problems of natural material-based matrices and the proteins may even denature upon immobilization. [17] The available alternative is the use of peptides with cell adhesion motifs that address integrin-like receptors. They are accessible by chemical synthesis but, due to their missing three-dimensional structure, show a lower bioactivity. [32] In this work, it was aimed to find a compromise between these approaches by designing adhesion miniproteins, that are both achievable by synthesis and have a folded three-dimensional structure. On the basis of independently folding miniprotein scaffolds, it was planned to design new cell adhesion molecules that use Ca(II)-mediated carbohydrate recognition to mimic D-type lectins. [3]

To design a Ca(II)-dependent D-type lectin, the design of a Ca(II)-binding miniprotein was required. It ought to be thermostable, folded and functional above 37 °C to work with cells. It needed to be accessible by chemical synthesis and bind Ca(II) with a micromolar K_d similar to LG domains. [3] To reach this goal, three different approaches were used on four different scaffolds: computational design with *Rosetta3*, intuitive *de novo* design, and functional domain grafting were applied to the SH3 domain, WW domain, Trp-cage and Trpzip. For most designs, the LG4 domain of Laminin α2 and the EF-hands of CaM served as a natural model.

Using computational design with *Rosetta3*, [4] model sequences were designed on the basis of the laminin LG4 active site and the SH3 domain scaffold. These Slam miniproteins proved difficult to design. The first model, Slam1, suffered from multiple side reactions and could not be purified by HPLC. The refined model, Slam2, was optimized for synthesis and folded into a native-like conformation as confirmed by CD spectroscopy. However, no Ca(II)-binding activity could be determined. The latest model, Slam3, was designed step-by-step with a geometry-optimized Ca(II)-site that could accommodate the GlcA-Xyl ligand of D-type lectins. [3] Due to time constraints, the model has not been realized, but displays a Ca(II) binding site at the same position as the Na(I) binding site of the Fyn SH3 domain. [78]

Using the same approach, the EF-hand of CaM was implemented in the hPin1 WW domain in the WWcalm series. [4] The two best-scoring models, WWcalm1 and WWcalm2, were easily achievable by synthesis but showed a random coil conformation in CD spectroscopy. [58] The models were discarded but the placement of the active sites between two β-strands, either between strands 1 and 2 or strands 2 and 3, were used to intuitively design WWcalm3 and WWcalm4. Both were synthesized without issue and improved in quality by using ψ-Pro dipeptides in Loop 1. [81] WWcalm3 was partially folded and changed conformation twice upon heating, hinting at a molten globule state. [83] WWcalm4 folded similar-to-native in the apo state but changed conformation and aggregated in the presence of Ca(II). Attempts to stabilize the designs by rational redesign led to more thermostable and well-folded but inactive variants of WWcalm3. For WWcalm4, the same redesigns did not change the aggregation behavior. A net charge analysis revealed a charge of -2, that could be nullified by binding Ca(II) thus leading to aggregation. [84] The redesigned WWcalm4.2 sequence included only one mutation to increase the negative charge and should lead to an active, folded and non-aggregating variant when synthesized.

The intuitive design of *de novo* Ca(II) binding sites followed the example of CD2.trigger by Li et al. and designed from electrostatic hotspots, namely complex salt bridges. [5] The positively charged amino acids of two adjacent salt bridges on the SH3 domain surface were mutated to Asp to disrupt the native conformation and create a highly negatively charged pocket. Binding Ca(II) was expected to reinstall the conformation. The model was successfully synthesized

and purified. It showed extraordinarily high Fluorescence and Tb(III) FRET signals, due to the formation of Tyr_2. [90] However, the mutations destabilized the β-barrel. The partially folded miniprotein showed very little thermostability and did not fold upon Ca(II) binding.

As no complex salt bridge could be found on the WW domain surface, the circular permutated WW domain cp34 by Kier et al was used. [8] This variant contained a complex salt bridge at the natural domain's termini. The positive charges were mutated to Asp and the model was successfully synthesized. The product was purified at a low yield, calling for optimization of the synthesis. The miniprotein was well-folded and thermostable above 37 °C. It showed a low Tb(III) FRET signal when excited at 280 nm, likely due to the distance between the excited Trp and the binding site. [91] Future experiments might benefit from direct Tb(III) fluorescence excitation at 222 nm. [99] Competitive dye binding assays revealed a Ca(II) binding affinity in the same range as the laminin LG domains with a K_d = 7.9 ± 3.4 µM. [3]

The grafting of functional domains, namely EF-hand or EF-hand-like motifs, was the final method used to design Ca(II)-binding onto new scaffolds. The sequence of the chosen domain was inserted into the sequence of the scaffold, either in a loop or a terminus. [6] The EF-hand-like Ca(II)-binding motif of blade W4 of a lectin found in *Psathyrella Velutina* [96] was grafted into Loop 1 of the hPin1 WW domain. The resulting WWfan model was synthesized and purified with high yield and showed well visible Tb(III) binding activity in FRET. CD spectroscopy revealed a native-like fold but low thermostability. The inserted sequence likely hindered a nucleation site of the WW domain, thus destabilizing the folding. A mutation to reintroduce Gly at the expected position might improve the stability and present a viable option for redesign.

EF-hand 2 of CaM was grafted onto the *N*-terminus of the Trp-cage miniprotein. The scaffold was designed by truncating exendin-4 and this grafting approach aimed to elongate the shortened α-helix. [10] The resulting model Calmcage was synthesized and purified with a good yield. CD spectroscopy revealed a native-like conformation with an increased thermostability, likely due to the elongation of the helix. Fluorescence spectroscopy and FRET showed micromolar Tb(III) binding and millimolar Ca(II) binding K_d-values. Even with a lower binding affinity, Calmcage should be tested for carbohydrate binding. A mutation of the *N*- and *C*-termini to oppositely charged residues could lead to a less flexible EF-hand and might improve the folding of the grafted domain and its binding strength.

EF-hand 2 of CaM was also grafted into the loop of Trpzip4. The β-hairpin models were named Calmzip1, with a partial EF-hand, and Calmzip2, with a complete EF-hand. Both were synthesized and purified at lower concentration due to the slow gelation of the crude material. CD spectroscopy revealed Calmzip1 as a well-folded and thermostable miniprotein but it was found to be non-binding in ITC experiments. In contrast, Calmzip2 showed a strong Tb(III) FRET signal and a native-like conformation, but very low thermostability. The large EF-hand domain was deemed too demanding for the Trpzip4 loop. A redesign with a scaffold for long loop closure by Anderson et al. and the EF-hand 3 of CaM resulted in the model Calmzip3. [11] The model was synthesized and purified but showed an even higher tendency for aggregation. A sequence analysis revealed a net charge close to zero that was changed to a positive with the mutation of an Ala spacer to Lys in Calmzip3.1 which was not fully studied in the course of this work. [84] Further characterization of Calmzip3 revealed strong Tb(III) binding with a K_d = 0.96 ± 0.15 µM and Ca(II) binding in the desired affinity range with a K_d = 3.9 ± 1.4 µM. CD spectroscopy of Calmzip3 at low concentration showed a well-folded β-hairpin conformation with high thermostability.

To prove *Rosetta3* design as a viable option for the SH3 domain scaffold, a SOD1 active site was designed into the scaffold. After thorough synthesis optimization of the model named SO1 by low temperature coupling, the synthesis yielded high amounts of product that were easily

purified. [111] CD spectroscopy showed a spectrum expected for β-Hairpin structures unlike the expected native conformation of SH3 domains. [72] The thermostability of the apo-state was lower than for the native domain but increased when bivalent transition metals were bound. The designed His_3 site successfully bound the target ion Cu(II) with a K_d = 3.60 × 10^{-9} M. With an eight orders of magnitude weaker affinity than the parent protein SOD1, the model featured a smaller binding site with one His replaced by Glu. [119] The SOD activity was shown in two indirect assays where the miniprotein competed for superoxide either with WST-1 or XTT and the apparent rate constant of Cu-SO1 k_{SOD} = 6.14 ± 0.58 × 10^7 M^{-1} s^{-1} was calculated from the known rate constant of XTT. [123] The successful design of SO1 not only proved the compatibility of the SH3 scaffold with *Rosetta3*, but also led to the miniprotein with the highest SOD activity to date, [14-16] that can compete with peptide-based mimics [12, 124] and may be used as a model to study SOD proteins in the future.

Three models fulfilled all criteria set for a Ca(II)-binding miniprotein and should be tested for lectin activity: WWturn, Calmcage and Calmzip3. The syntheses of these miniproteins should be further optimized and in the case of Calmzip3, the solubility-optimized variant Calmzip3.1 requires a full characterization as well. These models were designed using the two approaches that were successfully applied by other groups to design Ca(II)-binding proteins: intuitive *de novo* design and functional domain grafting. [54] The computational design, with its more time-consuming design phase failed to deliver a functional miniprotein, especially for the smaller WW domain scaffold. [58] However, the design of SO1 has shown that computational design of a SH3 domain scaffold with *Rosetta3* is viable, making Slam3 a promising candidate for future experimental characterization. The intuitive *de novo* design approach delivered functional miniproteins but may destabilize the folding to the point of no return, as seen for ScaX. The results suggested that the approach should work better in flexible regions like loops or loop-adjacent β-sheets and may require multiple attempts at different positions to produce results for a chosen scaffold. The grafting approach was the easiest to apply and always delivered an active miniprotein when complete domains were grafted. It worked well for flexible scaffolds, but may require rational redesign for thermostability or solubility to produce a final product. All in all, each method showed promise when combined with a fitting scaffold.

The results showed that no one scaffold was superior to the others. The SH3 domain may be the least robust scaffold used in this work, but it was the only one that produced viable results with *Rosetta3* in SO1. The well-studied and robust WW domain was not suitable for intuitive *de novo* design due to the lack of an accessible complex salt bridge, and lost its thermostability upon domain grafting. With the successful grafting on a β-hairpin peptide and Trp-cage, one could also show that larger scaffolds do not necessarily lead to better results. The often-observed stability-function trade-off was not found for the active models of this work. Apo-SO1 lost thermostability compared to the native domain but partially regained it upon binding Cu(II) and increased it upon binding Ni(II). [26] Calmcage was more stable than the Trp-cage, [10] the Calmzip series less stable than Trpzip4, [100] and WWturn was about equally stable compared to cp34. [8] The more prevalent inactive models however showed that designing new functions on miniproteins without the loss of thermostability, no matter the approach, remains a difficult endeavor.

Hirabayashi postulated that it should be possible to design lectin functionality on any scaffold. [28] Even though I could neither prove nor disprove this statement, this work paved the way for Ca(II)-mediated carbohydrate-binding miniproteins with the active models WWturn, Calmzip3 and Calmcage. Additionally, with the improved designs WWcalm4.2 and Slam3, possible lectin mimics for every scaffold used in this work were presented. It was shown that SH3 domains can be designed with *Rosetta3* to work as mini-enzymes and presented the miniprotein with the highest SOD activity to date in SO1. Future experiments should characterize the carbohydrate affinity of the successfully designed Ca(II)-binders and use the

knowledge on lectin engineering to turn them into structured adhesion miniproteins for Tissue Engineering.

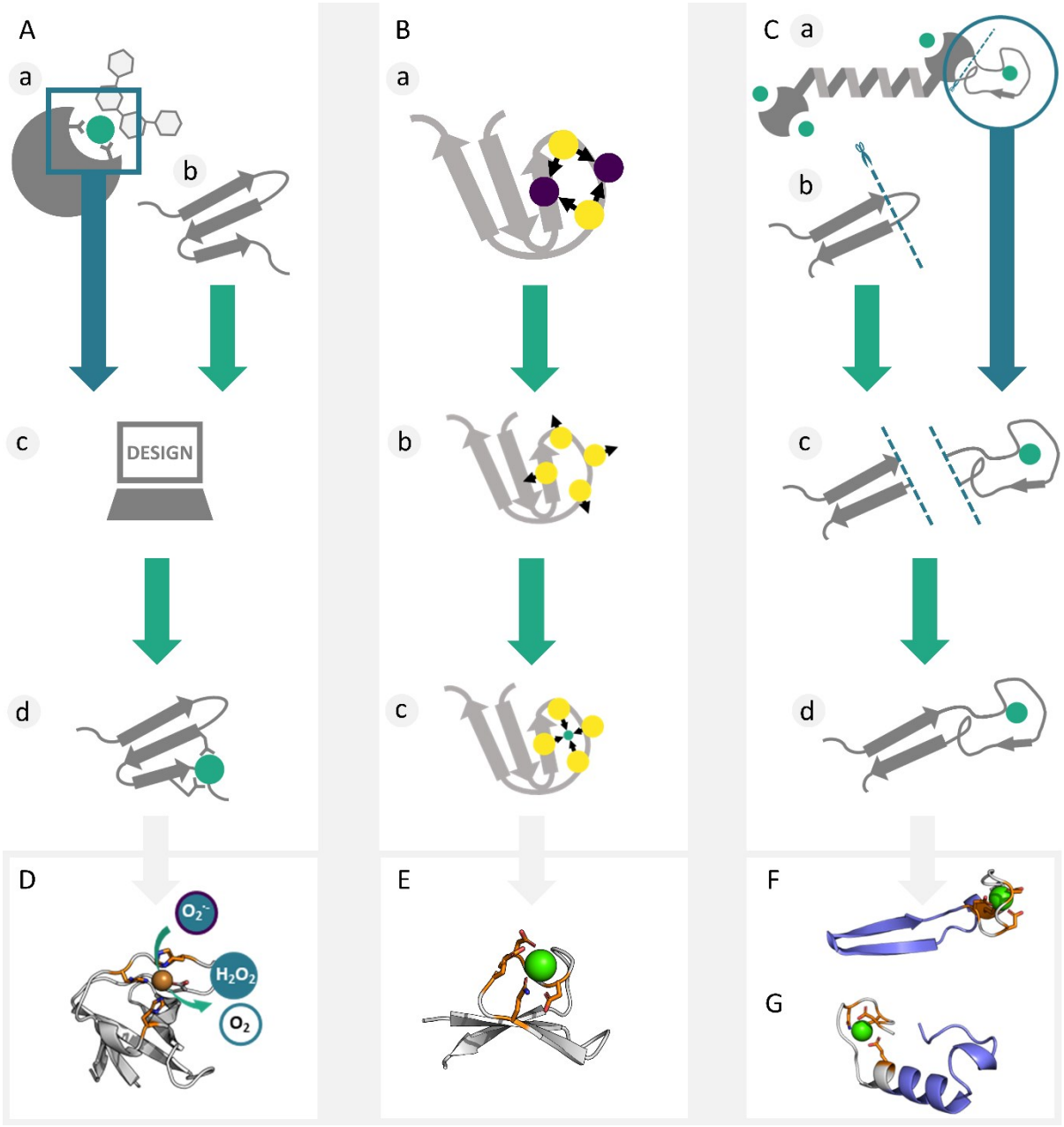


Figure 57. Design of Ca(II)-binding miniproteins with one of three approaches. A) Computational design using *Rosetta3* with (a) the natural model with the active site boxed (teal), (b) the miniprotein scaffold, (c) computational design that incorporates the active site of the natural model in the sequence of the scaffold and (d) the designed miniprotein; [4] B) intuitive *de novo* design from (a) a complex salt bridge on a miniprotein scaffold by (b) mutating the positively charged amino acids to Asp, with (c) the resulting hard cation binding site; [5] C) grafting of (a) a functional domain onto (b) a miniprotein scaffold by (c) incorporating the functional domain at a loop of the scaffold, resulting in (d) a functional miniprotein; [6] D) the computationally designed Cu(II)-binding miniprotein SO1 with superoxide dismutase activity based on the natural model SOD1 and the SH3 scaffold; E) the *de novo* designed Ca(II)-binding miniprotein WWturn based on the circular permutated WW domain by Kier et al.; [8] F) the Ca(II)-binding miniprotein Calmzip3 designed by grafting EF-hand 3 of Calmodulin (grey) onto the loop of the C4 β-sheet-cap by Anderson et al. (blue); [11] G) the Ca(II)-binding miniprotein Calmcage designed by grafting EF-hand 2 of Calmodulin (grey) onto the α-helix of the Trp-cage designed by Neidigh et al. (blue). [10]

6 Experiments

6.1 Peptide Design

6.1.1 Slam series

The model was designed based on the active site of laminin α2 LG4 (LG4, pdb: 5IK5) the N-terminal SH3 domain of c-crk (SH3, pdb: 1CKB) scaffold. The geometric parameters (distance, angles, dihedrals) of all Ca(II) binding amino acids (Asp 2808, Asp 2876, Leu 2825, Ile 2874) were taken from the crystal structure of LG4 in PyMOL. Rosetta3 was used according to a guide provided by Richter et al. [4]

Computational Design of Slam1

A bash script (Script 1) called Rosetta3 Match with additional options in a flag file (Script 2), allowed positions listed in a positions file (Script 3) and a geometric constraint file (Script 4) describing the binding of the two Asp residues of the LG4 active site (Asp 2808, Asp 2876). A bash script (Script 5) with a flag file (Script 6) called Rosetta3 Design for all Match outputs four times. The results were sorted by score of the apo- and holo-model. The best-scoring model was named Slam1.

Computational Design of Slam2

A parameter file (Script 7) for Ca(II) in an octahedral coordination sphere was created. A bash script (Script 1) called Rosetta3 Match with additional options in a flag file (Script 2), allowed positions listed in a positions file (Script 3) and a geometric constraint file (Script 9) describing the binding of the active site residues of the LG4 active site. No outputs were created for four CST blocks. The lowest-scoring outputs were produced when the CST block of Ile 2874 was omitted. A bash script (Script 10) called Rosetta3 Design for all Match outputs with constraints (Script 9) and additional options in a flag file (Script 11). The results were designed a second time without constraints (Script 5) with a flag file (Script 6). The best-scoring model was named Slam2.

Computational Design of Slam3

A parameter file (Script 12) for Ca(II)-bound GlcA-Xyl was created. The design was split into four phases, each initiated by a Python script. The first stage (Script 13) called the Rosetta3 Match application with flag options (Script 2) for CST blocks in any combination of the four binding amino acids of LG4 (Script 15). The second stage (Script 14) initiated Rosetta3 Design with flag options (Script 6) for the same CST block on all Match outputs. The models were ranked in the third stage (Script 16) and the lowest-scoring were relaxed in the final stage (Script 17). The final model omitted the CST block of Ile 2874 and was named Slam3.

Table 6. Sequences of c-crk SH3 domain and the Slam series. Mutations are printed bold.

c-Crk SH3	AE-YVRA-LFDFNGNDEEDLPFKKGD-ILRIRD-KPEEQ-WWNAE-DSEGK-RGMIP-VPY-VEKY
Slam1	AE-YVRA-LFDFNG**T**D**DD**DLPFKKG**E**-ILRIRD-KPEEQ-WWNAE-DSEGK-RG**L**IP-VPY-VEKYG
Slam2	AE-YVRA-LFDFNGNDEEDLPFKKGD-ILRIRD-KPEEQ-WWNAE-D**VL**GK-RG**L**IP-VPY-VEKYG
Slam3	AE-YVRA-LFD**YT**G**TVTL**DL**D**FKKG**E**-IL**T**IRD-KPEEQ-WWNAE-DSEGK-RG**L**IP-VPY-VEKY

6.1.2 WWcalm series

Computational Design of WWcalm1 and WWcalm2

The models were designed based on the active site of EF-hand 1 of calmodulin (CaM, pdb: 3CLN) and the human Pin1 WW domain (WW, pdb: 1I6C) scaffold. The geometric parameters (distance, angles, dihedrals) of each Ca(II) binding residue (Asp 20, Asp 22, Asp 24, Glu 31) were taken from the crystal structure of CaM in PyMOL. Rosetta3 was used according to a guide provided by Richter et al. [4]

A bash script (Script 18) called Rosetta3 Match with additional options in a flag file (Script 2) and a geometric constraint file (CST, Script 19) describing the binding of two of the four residues in all six non-redundant permutations. A bash script called Rosetta3 Design for all Match outputs. The results were sorted by score of the apo- and holo-model. The two best-scoring models were called WWcalm1 and WWcalm2.

Table 7. Sequences of hPin1 WW domain, WWcalm1 and WWcalm2. Mutations are printed bold.

hPin1	KLPPG-WEKRM-SRSSGR-VYYF-NHIT-NAS-QWERPSG
WWcalm1	KLPPG-W**N**K**VL**-SRSSGR-V**L**Y**D**-N**V**I**L**-**QLE**-**S**WERPSG
WWcalm2	KLPPG-W**NTDT**-SRSSGR-V**V**Y**D**-N**V**IT-NA**T**-**D**WERPSG

Intuitive Design of WWcalm3 and WWcalm4

Based on the binding sites of WWcalm1 and WWcalm2, new designs were developed intuitively with PyMol. The binding residues of WWcalm2 (Asp 9, Thr 20) were mutated on the WW domain with the addition of Glu 22 in Loop 2 for WWcalm3. The lost salt bridge between Glu 7 and Arg 9 was reestablished with a mutation to Arg 7. The binding residues of WWcalm1 (Asp 20, Asn 27) were mutated on the WW domain with the addition of Asp 25 on β-Strand 3 for WWcalm4. Loop1 was exchanged for a more flexible Gly-Ser repeat in both designs. β-Strand 2 was mutated to better accommodate the new binding sites.

Table 8. Sequences of hpin1 WW domain, WWcalm3 and WWcalm4. Mutations are printed bold.

hPin1	KLPPG-WEKRM-SRSSGR-VYYF-NHIT-NAS-QWERPSG
WWcalm3	KLPPG-W**RIDI**-**GSG**SG**S**-**YF**Y**T**-N**E**IT-NAS-QWERPSG
WWcalm4	KLPPG-WE**I**R**I**-**GSG**SG**S**-**F**Y**FD**-NHIT-**DA**N-QWERPSG

Rational Redesign of WWcalm3

WWcalm3 was rationally redesigned for a more stable variant. Based on the findings of Christina Lindner for more stable variations of Loop 1 (-ESGSGR-), β-Strand 2 (-IYYY-) and β-Strand 3 (-RTT-), WWcalm3 was iteratively improved in WWcalm3.1, WWcalm3.2 and WWcalm3.3. The binding site was mutated from Thr 20 to Asn for WWcalm3.4.

Table 9. Rational redesign of WWcalm3 with hydrophobic core positions printed red, mutations printed bold and redesigned sections boxed.

hPin1	KLPPG-WEKRM-SRSSGR-VYYF-NHIT-NAS-QWERPSG
WWcalm3	KLPPG-W**RIDI**-**GSG**SG**S**-**YF**Y**T**-N**E**IT-NAS-QWERPSG
WWcalm3.1	KLPPG-W**RIDI**-**ESG**SGR-**YF**Y**T**-N**E**IT-NAS-QWERPSG
WWcalm3.2	KLPPG-W**RIDI**-**ESG**SGR-**IYYY**-N**E**IT-NAS-QWERPSG
WWcalm3.3	KLPPG-W**RIDI**-**ESG**SGR-**IYY**-N**E**IT-**RTT**-QWERPSG
WWcalm3.4	KLPPG-W**RIDI**-**ESG**SGR-**IYYN**-N**E**IT-NAS-QWERPSG

Intuitive Redesign of WWcalm4

WWcalm4 was redesigned for a more stable variant. Based on the findings of Christina Lindner for more stable variations of Loop 1 (-ESGSGR-), β-Strand 2 (-IYYY-) and Loop 2 (-NSIT-) WWcalm4 was mutated in WWcalm4.1. To reduce aggregation and increase solubility, the net charge of the sequence was increased by mutating the binding site from Asn 27 to Asp for WWcalm4.2.

Table 10. Intuitive redesign of WWcalm4 with hydrophobic core positions printed red, mutations printed bold and redesigned sections boxed.

hPin1	KLPPG-WEKRM-SRSSGR-VYYF-NHIT-NAS-QWERPSG
WWcalm4	KLPPG-WE**I**R**I**-**GSG**SG**S**-**FY**F**D**-NHIT-**D**A**N**-QWERPSG
WWcalm4.1	KLPPG-WE**I**R**I**-**ESG**SGR-**IYY**D-N**S**IT-**D**A**N**-QWERPSG
WWcalm4.2	KLPPG-WE**I**R**I**-**GSG**SG**S**-**FY**F**D**-NHIT-**DAD**-QWERPSG

6.1.3 Intuitive Design by Salt Bridge Inversion

The surface of the peptide scaffold was scanned for a complex salt bridge or two adjacent simple salt bridges. The positively charged residues were mutated to Asp. The sequence was modeled using OmegaFold on Python3.10 and ColabFold v1.5.5 (AlphaFold2 using MMseqs2). According to Tang et al., this approach can create a destabilized apo-protein that stabilizes upon binding Ca(II). [54] The permutated WW domain by Kier et al. (cp34) [8] was mutated in positions Arg 16 and Lys 18 to Asp to create the model WWturn. The SH3 domain was mutated in positions Arg 28 and Arg 45 to Asp to create the model ScaX.

Table 11. Intuitive Design of WWturn with sequences of the hPin1 WW domain (pdb: 1I6C), its circular permutant by Kier et al. (cp34, pdb: 2MDU) and WWturn. Hydrophobic core positions are printed red, complex salt bridges are boxed and mutations are printed bold.

hPin1	KLPPG-WEKRM-SRSSGR-VYYF-NHIT-NAS-QWERPSG
cp34	R-WFYF-NRIT-GKR-QFERPKG-LVKG-WEKRW-D
WWturn	R-WFYF-NRIT-GKR-QFE**D**P**D**G-LVKG-WEKRW-D

Table 12. Intuitive Design of ScaX with sequences of the SH3 domain (pdb: 1CKB) and ScaX. Hydrophobic core positions are printed red, complex salt bridges are boxed and mutations are printed bold.

c-Crk SH3	AE-YVRA-LFDFNGNDEEDLPFKKGD-ILRIRD-KPEEQ-WWNAE-DSEGK-RGMIP-VPYVEKY
ScaX	AE-YVRA-LFDFNGNDEEDLPFKKG**E**-ILRI**D**D-KPEEQ-WWNAE-DSEGK-**D**G**L**IP-VPVEKY

6.1.4 Intuitive Design by Domain Grafting

The peptide scaffold was scanned for a part that could be replaced or extended by a Ca(II) binding domain. The Ca(II) binding domain sequence was inserted into the scaffold sequence and modeled using OmegaFold on Python3.10 and ColabFold v1.5.5 (AlphaFold2 using MMseqs2).

Intuitive Design of WWfan

The calcium binding motif of the *Psathyrella Velutina* β-propeller lectin's blade W4 [96] was inserted into the sequence of the WW domain. The new domain (176-ADVTGDGLLD-185) replaced Loop 1 (10-MSRSSGR-16) of the WW domain in the model WWfan.

Table 13. Intuitive Design of WWfan with sequences of the WW domain (pdb: 1I6C), the Ca(II) binding domain of blade W4 of the β-propeller lectin of *P. Velutina* (pdb: 2BWR) and WWfan. Hydrophobic core positions are printed red, grafted domain sections are boxed and mutations are printed bold.

```
hPin1                KLPPG-WEKRM-SRSSGR-     VYYF-NHIT-NAS-QWERPSG

-------------------------------------------------------------------

Propeller lectin              176-ADVTGDGLLD-185

WWfan                KLPPG-WEKRG-GDVTGDGLLD-VYYF-NHIT-NAS-QWERPSG
```

Intuitive Design of Calmcage

An EF-hand of calmodulin was inserted into the sequence of the Trp-cage by Neidigh et al. [10] EF-hand 2 (55-VDADGNGTIDYPEY-68) elongated the N-terminal helix of the Trp-cage in the model calmcage.

Table 14. Intuitive Design of Calmcage with sequences of Trp-cage (pdb: 1L2Y), calmodulin (pdb: 3CLN) and Calmcage. Grafted domain sections are boxed and mutations are printed bold.

```
Trp-cage                              NLYIQWLKD-GGPSSG-RPPPS

-------------------------------------------------------------------

CaM EF-hand 2    55- VDADGNGTIDFPEF-69

Calmcage             VDADGNGTIDFPEY-NLYIQWLKD-GGPSSG-RPPPS
```

Intuitive Design of Calmzip

The EF-hand of calmodulin was inserted into the sequence of a β-sheet hairpin peptide. A part of EF-hand 2 (56-DADGNG-61) replaced the loop (46-DDATK-50) of Trpzip4 in the model calmzip1. The full EF-hand 2 (55-VDADGNGTIDYPEYLT-70) replaced the loop (47-DATK-50) of Trpzip4 in calmzip2. Val and Gly spacers were inserted at the ends of the EF-hand loop to improve folding. EF-hand 3 (93-DKDGDGYISAAE-104) was replaced the loop of β-sheet-cap system C4 by Anderson et al. (9-GGGGKKGGGGKKGGGG-27) [11] in the model calmzip3. Ala and Gly spacers of different lengths and combinations were tested in simulation. The final model included a Gly spacer on the loop's N-terminus and three Ala spacers at the C-terminus. Calmzip3.1 mutated the first Ala spacer to Arg for a positive net charge.

Table 15. Intuitive Design of Calmzip with sequences of Trpzip4 (pdb: 1LE3), calmodulin (CaM, pdb: 3CLN), C4 by Anderson et al. and the Calmzip series. Grafted domain sections are boxed and mutations are printed bold.

Trpzip4	GEWTW	-DDATKT-	WTWTE
CaM EF-hand 2		55-VDADGNGTIDFPEFLT-70	
Calmzip1	GKWKW	-DADGNGT-	WEWTE
Calmzip2	G**K**W**K**W	**V**VDADGNGTID**Y**PE**Y**LT**G**T-	W**E**WTE
C4	RWITVRIW	-GGGGKKGGGGKKGGGG-	WKTIRVWE
CaM EF-hand 3		93-DKDGDGYISAAE-104	
Calmzip3	RWITVRIW	-GDKDGDGYISAAE**AAA**-	WKTIRVWE
Calmzip3.1	RWITVRIW	-GDKDGDGYISAAE**RAA**-	WKTIRVWE

6.1.5 Computational Design of SO1

The models were designed from human Cu/Zn Superoxide Dismutase 1 (SOD1, pdb: 3CQP, chain D) and the SH3 domain (pdb: 1CKB) following a guide by Richter et al. [4] The geometric parameters (distance, angles, dihedrals) of each exclusively Cu(II) binding residue (His 63, His 71, His 80) were taken from the crystal structure of SOD1.

A bash script (Script 21) ran Rosetta3 Match with allowed positions listed in a positions file (Script 3) and a geometric constraint file (Script 22) describing the binding of the three His residues of the SOD1 active site (His 63, His 71, His 80). Rosetta3 Design was called ten times for all Match outputs by a bash script (

Script 23). Rosetta3 Relax was called (Script 24) with residue file options to test for redundant mutations (Script 25) and aspartimide-promoting positions where Asp can be replaced (Script 26). Rosetta3 Score (Script 27) was used to sort the outputs by their Rosetta score with a special score function that penalized aspartimide-promoting sequences (Script 28). The results were sorted by Rosetta score of the apo- and holo-model. The best-scoring model was optimized by intuitive design for synthesis and named SO1.

Table 16. Computational Design of SO1 with the SH3 domain (pdb: 1CKB) and SO1. Hydrophobic core positions are printed red and mutations are printed bold.

c-crk SH3	AE-YVRA-LFDFNGNDEEDLPFKKGD-ILRIRD-KPEEQ-WWNAE-DSEGK-RGMIP-VPY-VEKY
SO1	AE-YVRA-LFD**HT**G**H**DEQ**E**LPFKKGE-ILRIRD-KPEEQ-**L**WNAE-NSEGK-RG**L**I**H**-V**VL**-VEKYG

6.2 Solid Phase Peptide Synthesis

Peptides were synthesized in a microwave-assisted peptide synthesizer (*CEM Liberty Blue*) set to 0.1 mmol scale. The reaction vessel was loaded with a H-Gly-HMPB ChemMatrix® resin (loading density: 0.5 mmol/mg), H-Ser-HMPB ChemMatrix® resin (loading density: 0.5 mmol/mg), H-Rink Amide ChemMatrix® resin (loading density: 0.5 mmol/mg) or TG XV RAM resin (loading density: 0.27 mmol/mg) as solid support. For peptide sequences longer than 50 amino acids, the amount of resin was halved. The resin was allowed to swell in DMF for at least 15 min before the synthesis. The synthesizer coupled solutions of Fmoc-protected amino acids in DMF (0.2 M) to the resin. DIC dissolved in DMF (0.5 M) served as the activator, and a solution of Oxyma® in DMF (1.0 M) as the activator base. A solution of piperidine (20 % *v*/*v*) with formic acid (5 % *v*/*v*) in DMF removed the Fmoc protecting group.

After the synthesis, the peptide was transferred into a 20 mL syringe with a PE frit and washed with DMF (3 x 5 mL) and DCM (10 x 5 mL) and dried under reduced pressure. A mixture of TFA (8.5 mL), TIPS (1 mL) and water (0.5 mL) cleaved the peptide from the resin and detached the protecting groups of the amino acid side chains within 3 h. The collected solution was washed with TFA (2 x 5 mL). TFA evaporated in a stream of nitrogen. The peptide was precipitated from ice-cold Et_2O (25 mL) and isolated by centrifugation. The pellet was washed with Et_2O (2 x 25 mL). The pellet was dissolved in a solution of acetonitrile (20 %) and TFA (0.1%) in water, frozen in liquid N_2 and lyophilized (*Christ Alpha 2-4-LDplus* connected to a *VACUUBRAND RZ 6* pump).

6.2.1 Standard protocol

His was coupled at 50 °C for 10 min. All other amino acids were coupled at 90 °C if not stated otherwise. Amino acids were added by single coupling reactions up to the 24th amino acid, further amino acids by double couplings – two coupling reactions with a set of washes in between. Arg and Gly were always double coupled.

6.2.2 CSY protocol

Standard protocol with the following changes: Fmoc-Asp(CSY)-OH was deprotected at room temperature. The system was cooled down during the last washing step before Fmoc-Asp(CSY)-OH was added for the first time. All further deprotection steps included a cooldown of the system to the reaction at room temperature. Standard cleavage conditions did not remove the CSY-protecting group. The CSY-protected peptide was purified by RP-HPLC, frozen in liquid N_2 and lyophilized prior to CSY-deprotection according to a protocol by Pham et al. [74] 250 nmol of CSY-protected peptide was dissolved in buffer A (1 mL), DMF (800 µL) and HFiP (200 µL). 2.2 equivalents per CSY-group of *N*-chloro succinimide in DMF (4 × 110 µL) were added under constant mixing. Between each addition, the mixture was equilibrated for 5 min. The reaction was quenched with 11.0 equivalents per CSY-group of sodium ascorbate in buffer A (110 µL) and a 1:1 mixture of DMF and buffer A (1450 µL) was added. The mixture was directly injected into an HPLC. All peaks were collected, pooled, frozen in liquid N_2 and lyophilized for purification.

6.2.3 LT75 protocol

Standard protocol with the following changes: His was coupled at 50 °C for 10 min. All other amino acids were coupled and deprotected at 75 °C for 5 min.

6.2.4 LT50 protocol

Standard protocol with the following changes: All amino acids were coupled at 50 °C for 10 min. Deprotection was carried out according to standard protocol.

6.3 High-Performance Liquid Chromatography

The peptides were purified by reversed-phase HPLC at 50 °C. The mobile phase consisted of TFA (0.1%) in water (buffer A) and TFA (0.1%) in acetonitrile (buffer B). The solutions were purified in 30-minute gradients (20-50 % B in A) if not stated otherwise.

6.3.1 Semipreparative HPLC

Semipreparative HPLC was performed using a *Jasco* chromatography system (pump *PU-4180*, column oven *CO-4060* and detector *UV-4070*) or a *Shimadzu* chromatography system (pump *LC-40D*, degasser *DGU-40S*, detector *SPD-M40*, column oven *CTO-40S*). A *VDS optilab VDSpher® PUR 100 C18-SE* (250 mm x 10 mm, 100 Å, 5 µm) column or a *Macherey-Nagel VP 250/10 NUCLEODUR 100-5 C18ec* (250 mm x 10 mm, 100 Å, 5 µm) column was used at a flow rate of 3 mL/min. Semipreparative chromatograms were monitored at 220 nm and 280 nm. The peptides were dissolved in solutions of buffer B (20 %) in buffer A and filtered with H-PTFE syringe filters (0.2 µm pore size) from *Macherey-Nagel.* Fractions containing pure peptide were identified by analytical HPLC and MALDI-TOF MS, pooled and lyophilized. Spectra were visualized using a Python script for data from the *Jasco* (Script 29) or *Shimadzu* system (Script 30).

6.3.2 Analytical HPLC

Analytical HPLC was performed using a *Hitachi Primaide* chromatography system containing a *1110 Pump*, a *1210 auto sampler*, *1310 column oven* and a *1430 diode array detector*. A *VDS optilab VDSpher® PUR 100 C18-SE* (250 mm x 4.6 mm, 100 Å, 5 µm) column was used at a flow rate of 1 mL/min. Analytical chromatograms were monitored from 190 to 400 nm. Samples collected from Semipreparative HPLC were diluted (1:1 v/v) in water. Spectra were visualized using a Python script (Script 31).

6.3.3 Size Exclusion Chromatography

Size exclusion chromatography was performed using *Shimadzu* chromatography system (pump *LC-40D*, degasser *DGU-40S*, detector *SPD-M40*, column oven *CTO-40S*). A *Cytiva Superdex 75 Increase 10/300 GL* column was used at a flow rate of 0.8 mL/min. The peptides were dissolved in pH 7.2 buffer containing 50 mM MOPS and 150 mM NaCl and filtered with H-PTFE syringe filters (0.2 µm pore size) from *Macherey-Nagel.* Fractions containing pure peptide were identified by analytical HPLC and MALDI-TOF MS, pooled and lyophilized.

6.4 Mass Spectrometry

Matrix-assisted laser desorption ionisation time-of-flight mass spectrometry (MALDI-TOF MS) was performed on a *Bruker Autoflex Speed*. 1 µl of the matrix was mixed with 1 µl of the sample, by pipetting onto the target plate and air drying. For peptides based on SH3 domains (Slam, SO1, ScaX) a supersaturated sinapinic acid solution and for lower mass peptides a saturated 2,5-dihydroxybenzoic acid solution was used as matrix. The mass spectra were processed with *mMass* and visualized using a Python script (Script 32). [127]

6.5 Buffer preparation

Concentrated buffer solutions were prepared from a solution of the buffer salt in water. The pH was adjusted for the concentrated solution with NaOH (1M) and HCl (1M) using a *Hanna Instruments* HI2210 pH meter equipped with an HI1131 electrode. The volume was set using a measuring flask. The solution was filtered under reduced pressure and stored in a fridge. Buffers used for Ca(II) binding experiments were filtered with Chelex100 resin.

6.5.1 Decalcification by Dialysis

A cellulose membrane dialysis tube (*Merck*) was rehydrated in boiling water. The dialysis tube was filled with 5 g *Chelex100* resin suspended in the concentrated buffer solution. The dialysis tube was submerged in a 50 mL conical tube filled with 40 mL concentrated buffer solution with both ends pointing outwards. The lid was closed to seal conical tube and the ends of the dialysis tube. The overhang of the dialysis tube was cut and the flask was sealed with wax tape. The conical tube was equilibrated overnight (or for a week) on a shaker plate. The buffer solution was transferred into a new container. The decalcification was controlled according to 6.5.4.

6.5.2 Decalcification by Batch method

The concentrated buffer solution was transferred into a conical tube and 5 g Chelex100 were added. The tube was sealed and equilibrated for 2-3 h on a shaker plate. The suspension was centrifuged, decanted and filtrated to remove the resin. The decalcification was controlled according to 6.5.4.

6.5.3 Decalcification by Column method

A 5 mL BD syringe equipped with a PE frit was filled with 5 g Chelex100 resin and rinsed with at least 20 mL water and 15 mL concentrated buffer solution. The remaining buffer solution was added to the stamp opening of the syringe and the flow-through was collected. The decalcification was controlled according to 6.5.4.

6.5.4 pH readjustment and control of Ca(II) concentration after Decalcification

Small portions of 1 mL were taken out to control the pH and the main solution was adjusted with NaOH (1M) and HCl (1M). The remaining concentration of Ca(II) was controlled by a buffered 25 µM solution of Br_2-BAPTA. The sample (200 µL) was transferred into a cuvette (10 mm) and the absorbance (A_1) was measured on a nanophotometer *IMPLEN NP80* at room temperature (22 to 25 °C). 5 µL of an EDTA solution (0.1 M) were added for a second absorbance measurement (A_2) and 10 µL of a calcium chloride solution (0.5 M) were added for a third absorbance measurement (A_3). The concentration of Ca(II) in buffered solution c(Ca) was then calculated according to Equation 1.

$$c(Ca) = \frac{A_2 - A_1}{A_2 - A_3}$$

Equation 3. Ca(II) concentration after decalcification c(Ca) with absorbance A_1 of the test solution, A_2 after addition of 5 µL EDTA (0.1 M) and A_3 after addition of 10 µL $CaCl_2$ (0.5 M).

6.6 UV/Vis Spectroscopy

UV/VIS spectroscopy was performed on a nanophotometer *IMPLEN NP80* at room temperature (22 to 25 °C). The sample (200 µL) was transferred into a cuvette (10 mm) or added to the nanovolume (1 µL) for measurement.

6.6.1 Concentration determination of peptides

The peptides were dissolved in water and the concentration determined by measuring the absorbance at 280 nm. The extinction coefficient was calculated from the sum of the extinction coefficients of Trp (5500 $M^{-1}cm^{-1}$) and Tyr (1490 $M^{-1}cm^{-1}$) residues. [128] Each sample was measured three times and the mean value was used for concentration calculation.

6.6.2 UV/Vis Spectra measurement

Spectra were recorded from 200 nm to 900 nm in a 10 mm cuvette. The *Spectra measurement* mode of the *IMPLEN* Software with *smoothing on* and *blank correction off* was used.

6.7 Circular Dichroism Spectroscopy

CD spectroscopy was performed on a *Jasco J-1700* CD spectrometer, equipped with a *Jasco PTC-510* Peltier thermostat or a *Jasco J-810* CD spectrometer, equipped with a *Jasco PTC-423S* Peltier thermostat using quartz cuvettes from *Starna*. The measuring chamber was flushed constantly with nitrogen. Settings: data pitch: 0.1 nm; scanning speed: 100 nm/min, sensitivity: low, response: 2 s; bandwidth: 1 nm.

6.7.1 Spectra measurement

CD spectra were recorded from 190 to 260 nm at 20 °C. After placing the cuvette into the holder, the sample was left to equilibrate for 5 min prior to measurement. 10 spectra were recorded and accumulated. The measured ellipticity θ (in mdeg) was converted into mean residue ellipticity MRE (in deg·cm^2·dmol^{-1}·res^{-1}) using Equation 4 where θ^{blank} is the blank signal, c is the peptide concentration in mol/L, l is the pathlength in mm and n is the number of backbone peptide bonds. [70]

$$MRE = \frac{\theta - \theta^{blank}}{c \cdot l \cdot n}$$

Equation 4. Mean residue ellipticity with sample ellipticity $\boldsymbol{\theta}$ (in mdeg), blank ellipticity $\boldsymbol{\theta^{blank}}$ (in mdeg), sample concentration $\boldsymbol{c}$, cuvette pathlength $\boldsymbol{l}$ and number of backbone peptide bonds $\boldsymbol{n}$.

6.7.2 Thermal denaturation measurement

Thermal denaturation curves were recorded at the wavelength of maximal MRE from 0 to 98 °C with a temperature gradient of 1 °C/min. Data points were collected every 0.5 °C and fitted to a two-state model (Equation 5) with molar enthalpy ΔH, universal gas constant R, melting temperature T_m (in K), temperature T (in K), slope of the lower plateau m, slope of the upper plateau n, abscissa of the lower plateau u and abscissa of the upper plateau v . [73]

$$f(T) = (m\,T + u) + \frac{(n\,T + v) - (m\,T + u)}{1 + \exp\left(\frac{\Delta H}{R\,T} - \frac{\Delta H}{R\,T_m}\right)}$$

Equation 5. Two-state transition model for thermal denaturation measurements with molar enthalpy $\boldsymbol{\Delta H}$, universal gas constant $\boldsymbol{R}$, melting temperature $\boldsymbol{T_m}$ (in K), temperature $\boldsymbol{T}$ (in K), slope of the lower plateau $\boldsymbol{m}$, slope of the upper plateau $\boldsymbol{n}$, abscissa of the lower plateau $\boldsymbol{u}$ and abscissa of the upper plateau $\boldsymbol{v}$.

6.8 Fluorescence Spectroscopy

Fluorescence spectroscopy was performed on a *JASCO FP-8300* spectrofluorometer at room temperature (22 to 25 °C). The sample (200 µL) was transferred into a cuvette (10 mm). Trp excitation was carried out at 295 nm for emission spectra from 300-575 nm. Settings: excitation bandwidth 1 nm, emission bandwidth 5 nm. Response 1 s, data interval 0.5 nm, scanning speed 100 nm/min, sensitivity: high.

6.9 Fluorescence Resonance Energy Transfer Spectroscopy

5 µM solutions of peptide were prepared in pH 7.2 buffer containing 50 mM MOPS and 150 mM NaCl. The samples (200 µL) were prepared with and without 2 mM Tb(III). Titrations were performed by preparation of samples at varying concentrations of Tb(III). Fluorescence spectroscopy was performed on a *JASCO FP-8300* spectrofluorometer at room temperature (22 to 25 °C) in a 10 mm cuvette. The cuvette was rinsed with Hellmanex, water and

isopropanol and dried under a stream of nitrogen before each measurement. The data was analyzed using a Python script (Script 33). The Kd was determined as the mean of three titrations.

6.10 Competitive Dye Binding Assay

5 µM solutions of peptide and dye were prepared in pH 7.2 buffer containing 50 mM MOPS and 150 mM NaCl. The sample (2.0 mL) was prepared in a cuvette (10 mm) with a magnetic stirrer. UV/VIS spectroscopy was performed on a nanophotometer *IMPLEN NP80* at room temperature (22 to 25 °C). The absorbance was measured after stirring for 5 minutes at the wavelengths λ_1 and λ_2 of the respective dye for each addition (2-2.5 µL) of 1.0 mM metal ligand solution until saturation was reached. All concentrations were doubled for Ca(II) binding assays. The data was analyzed using DynaFit () [92] or simulated using a Python script using the *pybindingcurve* module (Script 35). The K_d was determined as the mean of three titrations.

Table 17. Metal binding dyes with recorded wavelengths and K_d in aqueous buffer solution.

	Ca^{2+}	Mn^{2+}	Co^{2+}	Ni^{2+}	Cu^{2+}	Zn^{2+}
Dye	Br_2-BAPTA	Mag-Fura-2	Mag-Fura-2	Fura-2	TAMSMB	Mag-Fura-2
$\lambda_{1,2}$ [nm]	263/239.5	324/365	324/365	335/365	586	324
K_d [M]	$1.4 \cdot 10^{-6}$	$8.9 \cdot 10^{-7}$	$9.2 \cdot 10^{-7}$	$6.9 \cdot 10^{-9}$	$4.4 \cdot 10^{-8}$	$3.7 \cdot 10^{-8}$
Source	[75]	[129]	[129]	[116]	[130]	[129]

6.11 Competitive Ligand Binding Assay

5 µM solutions of peptide were prepared in pH 7.2 buffer containing 50 mM MOPS and 150 mM NaCl. The samples (200 µL) were prepared with 100 µM Tb(III) and varying concentrations of Ca(II). Fluorescence spectroscopy was performed on a *JASCO FP-8300* spectrofluorometer at room temperature (22 to 25 °C) in a 10 mm cuvette. The cuvette was rinsed with Hellmanex, water and isopropanol and dried in a stream of nitrogen before each measurement. The data was analyzed using DynaFit (Script 36). [92] The K_d was determined as the mean of two titrations.

6.12 Isothermal Titration Calorimetry

A 25 µM buffered sample solution of the investigated peptide (900 µL) containing 50 mM MOPS and 150 mM NaCl was prepared. A 20 mM calcium chloride solution containing 50 mM MOPS and 150 mM NaCl was prepared from a 500 mM stock solution. The sample cell of a *Malvern Microcal PEAQ-ITC* isothermal titration calorimeter was washed twice sample solution and filled with 250 µL sample solution. The syringe attached to the device was washed twice with calcium chloride solution and filled with 50 µL calcium chloride solution. The experiment was set up for 13 titrations at 25 °C. The data was analyzed with *MicroCal PEAQ-ITC Analysis Software v1.41*.

6.13 Superoxide Dismutase Activity Assay

Xanthine Oxidase (CS0009E) and bovine Superoxide Dismutase (CS0009C) were thawed and diluted 100-fold and 50-fold respectively in dilution buffer (CS0009B). Hypoxanthine was dissolved in formic acid, diluted with water and neutralized with NaOH. Measurements were performed on a nanophotometer *IMPLEN NP80* at room temperature (22 to 25 °C).

6.13.1 Endpoint determination

Endpoint determination was carried out based on a protocol by Merck from a Superoxide Dismutase activity kit adjusted for EDTA-free use with weaker-binding systems. A sample (200 µL) of 25 µM WST-1 (CS0009D) was prepared in pH 7.8 buffer containing 100 mM potassium phosphate with 50 µM hypoxanthine, a defined concentration or activity of SOD sample and 20 µL of Xanthine Oxidase solution to start the reaction. The sample was transferred into a cuvette (10 mm) and the absorbance at 450 nm was monitored over the course of 30 minutes. The mean of the final 10 absorbance values (2 min) of each measurement (A_i) was used to determine the degree of inhibition. The IC_{50} was determined by a linear regression of the linearized rate $L = A_0 / A_i$ versus the peptide concentration at L = 2. Data were analyzed and visualized using a Python script (Script 37).

A negative control was measured without Superoxide Dismutase (A_0), a positive control with 1 unit/mL bovine Superoxide Dismutase, a positive control with free $CuSO_4$ at 8.4 and 30.7 nM concentration and test solution of Cu-SO1 (peptide /Cu(II) 25:24) at 2.5, 10.0, 20.0, 32.0, 50.0 and 100.0 nM concentration.

6.13.2 Rate determination

Rate determination activity assays were carried out according to a protocol by Mathieu et al. [15] A sample (200 µL) of 100 µM XTT was prepared in pH 7.8 buffer containing 100 mM potassium phosphate with 200 µM hypoxanthine, a defined concentration of SOD sample and 20 µL of Xanthine Oxidase solution to start the reaction. The sample was transferred into a cuvette (10 mm) and the absorbance at 470 nm was monitored over the course of 90 s. A linear regression gave the rate of reaction. The degree of inhibition was calculated from the rate in the presence and absence of SOD sample. The IC_{50} was determined by linear regression of the linearized rate $L = v_0 / v_i$ versus the peptide concentration at L = 2.

A negative control was measured without Superoxide Dismutase (v_0), a negative control of apo-SO1 (560.0 nM) and test solutions of Cu-SO1 (Cu(II)/peptide 1:4) at 120.0 nM, 130.0 nM and 140.0 nM concentration. Data were analyzed and visualized using a Python script (Script 38).

7 References

1. Ikada, Y., *Challenges in tissue engineering.* J R Soc Interface, 2006. **3**(10): p. 589-601.
2. Vasita, R., I.K. Shanmugam, and D.S. Katt, *Improved biomaterials for tissue engineering applications: surface modification of polymers.* Curr Top Med Chem, 2008. **8**(4): p. 341-53.
3. Hohenester, E., *Laminin G-like domains: dystroglycan-specific lectins.* Curr Opin Struct Biol, 2019. **56**: p. 56-63.
4. Richter, F., et al., *De novo enzyme design using Rosetta3.* PLoS One, 2011. **6**(5): p. e19230.
5. Li, S., et al., *Rational design of a conformation-switchable Ca2+- and Tb3+-binding protein without the use of multiple coupled metal-binding sites.* FEBS J, 2008. **275**(20): p. 5048-61.
6. Ye, Y., et al., *Probing site-specific calmodulin calcium and lanthanide affinity by grafting.* J Am Chem Soc, 2005. **127**(11): p. 3743-50.
7. Chidyausiku, T.M., et al., *De novo design of immunoglobulin-like domains.* Nature Communications, 2022. **13**(1).
8. Kier, B.L., J.M. Anderson, and N.H. Andersen, *Circular Permutation of a WW Domain: Folding Still Occurs after Excising the Turn of the Folding-Nucleating Hairpin.* Journal of the American Chemical Society, 2014. **136**(2): p. 741-749.
9. Lewit-Bentley, A. and S. Rety, *EF-hand calcium-binding proteins.* Curr Opin Struct Biol, 2000. **10**(6): p. 637-43.
10. Neidigh, J.W., R.M. Fesinmeyer, and N.H. Andersen, *Designing a 20-residue protein.* Nat Struct Biol, 2002. **9**(6): p. 425-30.
11. Anderson, J.M., et al., *Optimization of a beta-sheet-cap for long loop closure.* Biopolymers, 2017. **107**(3).
12. Csire, G., et al., *Coordination, redox properties and SOD activity of Cu(II) complexes of multihistidine peptides.* J Inorg Biochem, 2017. **177**: p. 198-210.
13. Zhou, Y.H., et al., *Synthesis, structure, and activity of supramolecular mimics for the active site and arg141 residue of copper, zinc-superoxide dismutase.* Inorg Chem, 2007. **46**(3): p. 734-9.
14. Benson, D.E., M.S. Wisz, and H.W. Hellinga, *Rational design of nascent metalloenzymes.* Proc Natl Acad Sci U S A, 2000. **97**(12): p. 6292-7.
15. Mathieu, E., et al., *Rational De Novo Design of a Cu Metalloenzyme for Superoxide Dismutation.* Chemistry, 2020. **26**(1): p. 249-258.
16. Singh, U.P., et al., *Design and Synthesis of De Novo Peptide for Manganese Binding.* International Journal of Peptide Research and Therapeutics, 2006. **12**(4): p. 379-385.
17. Berthiaume, F., T.J. Maguire, and M.L. Yarmush, *Tissue engineering and regenerative medicine: history, progress, and challenges.* Annu Rev Chem Biomol Eng, 2011. **2**: p. 403-30.
18. Hughes, C.S., L.M. Postovit, and G.A. Lajoie, *Matrigel: a complex protein mixture required for optimal growth of cell culture.* Proteomics, 2010. **10**(9): p. 1886-90.
19. Patel, R. and A.J. Alahmad, *Growth-factor reduced Matrigel source influences stem cell derived brain microvascular endothelial cell barrier properties.* Fluids Barriers CNS, 2016. **13**: p. 6.
20. Unal, A.Z. and J.L. West, *Synthetic ECM: Bioactive Synthetic Hydrogels for 3D Tissue Engineering.* Bioconjug Chem, 2020. **31**(10): p. 2253-2271.
21. D'Souza, S.E., M.H. Ginsberg, and E.F. Plow, *Arginyl-glycyl-aspartic acid (RGD): a cell adhesion motif.* Trends Biochem Sci, 1991. **16**(7): p. 246-50.
22. Heino, J., *The collagen family members as cell adhesion proteins.* Bioessays, 2007. **29**(10): p. 1001-10.
23. Aumailley, M., *The laminin family.* Cell Adh Migr, 2013. **7**(1): p. 48-55.
24. Briggs, D.C., et al., *Structural basis of laminin binding to the LARGE glycans on dystroglycan.* Nat Chem Biol, 2016. **12**(10): p. 810-4.
25. Cochran, A.G., N.J. Skelton, and M.A. Starovasnik, *Tryptophan zippers: stable, monomeric beta -hairpins.* Proc Natl Acad Sci U S A, 2001. **98**(10): p. 5578-83.

26. Muralidharan, V., et al., *Solution structure and folding characteristics of the C-terminal SH3 domain of c-Crk-II.* Biochemistry, 2006. **45**(29): p. 8874-84.
27. Pham, T.L. and F. Thomas, *Design of Functional Globular beta-Sheet Miniproteins.* Chembiochem, 2024: p. e202300745.
28. Hirabayashi, J. and R. Arai, *Lectin engineering: the possible and the actual.* Interface Focus, 2019. **9**(2): p. 20180068.
29. Xue, N., et al., *Bone Tissue Engineering in the Treatment of Bone Defects.* Pharmaceuticals (Basel), 2022. **15**(7).
30. Vig, K., et al., *Advances in Skin Regeneration Using Tissue Engineering.* Int J Mol Sci, 2017. **18**(4).
31. Andreas, K., et al., *Key regulatory molecules of cartilage destruction in rheumatoid arthritis: an in vitro study.* Arthritis Res Ther, 2008. **10**(1): p. R9.
32. Hersel, U., C. Dahmen, and H. Kessler, *RGD modified polymers: biomaterials for stimulated cell adhesion and beyond.* Biomaterials, 2003. **24**(24): p. 4385-4415.
33. Chothia, C. and E.Y. Jones, *The molecular structure of cell adhesion molecules.* Annu Rev Biochem, 1997. **66**: p. 823-62.
34. Buckley, C.D., et al., *Cell adhesion: more than just glue (review).* Mol Membr Biol, 1998. **15**(4): p. 167-76.
35. Baker, E.G., et al., *Miniprotein Design: Past, Present, and Prospects.* Acc Chem Res, 2017. **50**(9): p. 2085-2092.
36. Ozga, K. and L. Berlicki, *Design and Engineering of Miniproteins.* ACS Bio Med Chem Au, 2022. **2**(4): p. 316-327.
37. Stewart, A.L., J.H. Park, and M.L. Waters, *Redesign of a WW domain peptide for selective recognition of single-stranded DNA.* Biochemistry, 2011. **50**(13): p. 2575-84.
38. Hsu, C.H., et al., *The Dependence of Carbohydrate-Aromatic Interaction Strengths on the Structure of the Carbohydrate.* J Am Chem Soc, 2016. **138**(24): p. 7636-48.
39. Pham, T.L., et al., *Relationship of Thermostability and Binding Affinity in Metal-binding WW-Domain Minireceptors.* Chembiochem, 2024. **25**(4): p. e202300715.
40. Lindner, C., et al., *Thermostable WW-Domain Scaffold to Design Functional beta-Sheet Miniproteins.* J Am Chem Soc, 2024.
41. Mongiovi, A.M., et al., *A novel peptide-SH3 interaction.* EMBO J, 1999. **18**(19): p. 5300-9.
42. Dooley, K., et al., *Engineering of an environmentally responsive beta roll peptide for use as a calcium-dependent cross-linking domain for peptide hydrogel formation.* Biomacromolecules, 2012. **13**(6): p. 1758-64.
43. Nguyen, A.K., et al., *Hierarchical assembly of tryptophan zipper peptides into stress-relaxing bioactive hydrogels.* Nat Commun, 2023. **14**(1): p. 6604.
44. Goldstein, I.J. and C.E. Hayes, *The lectins: carbohydrate-binding proteins of plants and animals.* Adv Carbohydr Chem Biochem, 1978. **35**: p. 127-340.
45. Arnaud, J., A. Audfray, and A. Imberty, *Binding sugars: from natural lectins to synthetic receptors and engineered neolectins.* Chem Soc Rev, 2013. **42**(11): p. 4798-813.
46. Sun, Z., B. Fan, and M.J. Webber, *Molecular Engineering of Carbohydrate Recognition.* ChemSystemsChem, 2023. **5**(5).
47. Armenta, S., et al., *Advances in molecular engineering of carbohydrate-binding modules.* Proteins, 2017. **85**(9): p. 1602-1617.
48. Carafoli, E. and J. Krebs, *Why Calcium? How Calcium Became the Best Communicator.* J Biol Chem, 2016. **291**(40): p. 20849-20857.
49. Zhang, M., et al., *Structural basis for calmodulin as a dynamic calcium sensor.* Structure, 2012. **20**(5): p. 911-23.
50. Paulsson, M., *The role of Ca2+ binding in the self-aggregation of laminin-nidogen complexes.* Journal of Biological Chemistry, 1988. **263**(11): p. 5425-5430.
51. Dempsey, C.E., et al., *Analysis of alpha-Dystroglycan/LG Domain Binding Modes: Investigating Protein Motifs That Regulate the Affinity of Isolated LG Domains.* Front Mol Biosci, 2019. **6**: p. 18.
52. Chin, D. and A.R. Means, *Calmodulin: a prototypical calcium sensor.* Trends Cell Biol, 2000. **10**(8): p. 322-8.

53. Dominguez, D.C., M. Guragain, and M. Patrauchan, *Calcium binding proteins and calcium signaling in prokaryotes.* Cell Calcium, 2015. **57**(3): p. 151-65.
54. Tang, S., et al., *Design of Calcium-Binding Proteins to Sense Calcium.* Molecules, 2020. **25**(9): p. 2148.
55. Yang, W., et al., *Design of a calcium-binding protein with desired structure in a cell adhesion molecule.* J Am Chem Soc, 2005. **127**(7): p. 2085-93.
56. Huang, Y., et al., *A single EF-hand isolated from STIM1 forms dimer in the absence and presence of Ca2+.* FEBS J, 2009. **276**(19): p. 5589-97.
57. Siedlecka, M., et al., *Alpha-helix nucleation by a calcium-binding peptide loop.* Proc Natl Acad Sci U S A, 1999. **96**(3): p. 903-8.
58. Das, R., *Four Small Puzzles That Rosetta Doesn't Solve.* PLoS ONE, 2011. **6**(5): p. e20044.
59. McCord, J.M. and I. Fridovich, *Superoxide Dismutase.* Journal of Biological Chemistry, 1969. **244**(22): p. 6049-6055.
60. Shapovalov, M.V. and R.L. Dunbrack, Jr., *A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions.* Structure, 2011. **19**(6): p. 844-58.
61. Wang, J., et al., *Computational Protein Design with Deep Learning Neural Networks.* Scientific Reports, 2018. **8**(1).
62. Alford, R.F., et al., *The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design.* J Chem Theory Comput, 2017. **13**(6): p. 3031-3048.
63. Siegel, J.B., et al., *Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction.* Science, 2010. **329**(5989): p. 309-13.
64. Tantillo, D.J., J. Chen, and K.N. Houk, *Theozymes and compuzymes: theoretical models for biological catalysis.* Curr Opin Chem Biol, 1998. **2**(6): p. 743-50.
65. Zanghellini, A., et al., *New algorithms and an in silico benchmark for computational enzyme design.* Protein Sci, 2006. **15**(12): p. 2785-94.
66. Tisi, D., et al., *Structure of the C-terminal laminin G-like domain pair of the laminin alpha2 chain harbouring binding sites for alpha-dystroglycan and heparin.* EMBO J, 2000. **19**(7): p. 1432-40.
67. Rudenko, G., E. Hohenester, and Y.A. Muller, *LG/LNS domains: multiple functions -- one business end?* Trends Biochem Sci, 2001. **26**(6): p. 363-8.
68. Kurochkina, N. and U. Guha, *SH3 domains: modules of protein-protein interactions.* Biophys Rev, 2013. **5**(1): p. 29-39.
69. Borreguero, J.M., et al., *Multiple folding pathways of the SH3 domain.* Biophys J, 2004. **87**(1): p. 521-33.
70. Greenfield, N.J., *Using circular dichroism spectra to estimate protein secondary structure.* Nat Protoc, 2006. **1**(6): p. 2876-90.
71. Maxwell, K.L. and A.R. Davidson, *Mutagenesis of a buried polar interaction in an SH3 domain: sequence conservation provides the best prediction of stability effects.* Biochemistry, 1998. **37**(46): p. 16172-82.
72. Sreerama, N. and R.W. Woody, *Computation and analysis of protein circular dichroism spectra.* Methods Enzymol, 2004. **383**: p. 318-51.
73. Greenfield, N.J., *Analysis of the kinetics of folding of proteins and peptides using circular dichroism.* Nat Protoc, 2006. **1**(6): p. 2891-9.
74. Pham, T.L., et al., *The CSY-protecting group in the microwave-assisted synthesis of aggregation-prone peptides.* RSC Chem Biol, 2022. **3**(4): p. 426-430.
75. Linse, S., *Calcium binding to proteins studied via competition with chromophoric chelators.* Methods Mol Biol, 2002. **173**: p. 15-24.
76. Lehto, J., et al., *Hydrolysis and exchange by Chelex 100 chelating resin.* Reactive Polymers, 1994. **23**(2-3): p. 135-140.
77. Shave, S., et al., *PyBindingCurve, Simulation, and Curve Fitting to Complex Binding Systems at Equilibrium.* J Chem Inf Model, 2021. **61**(6): p. 2911-2915.
78. Martin-Garcia, J.M., et al., *The promiscuous binding of the Fyn SH3 domain to a peptide from the NS5A protein.* Acta Crystallogr D Biol Crystallogr, 2012. **68**(Pt 8): p. 1030-40.

79. Andrews, C., et al., *Structural Aspects and Prediction of Calmodulin-Binding Proteins.* Int J Mol Sci, 2020. **22**(1).
80. Babu, Y.S., C.E. Bugg, and W.J. Cook, *Structure of calmodulin refined at 2.2 A resolution.* J Mol Biol, 1988. **204**(1): p. 191-204.
81. Wöhr, T., et al., *Pseudo-Prolines as a Solubilizing, Structure-Disrupting Protection Technique in Peptide Synthesis.* Journal of the American Chemical Society, 1996. **118**(39): p. 9218-9227.
82. Davis, C.M. and R.B. Dyer, *WW domain folding complexity revealed by infrared spectroscopy.* Biochemistry, 2014. **53**(34): p. 5476-84.
83. Uversky, V.N. and A.V. Finkelstein, *Life in Phases: Intra- and Inter- Molecular Phase Transitions in Protein Solutions.* Biomolecules, 2019. **9**(12).
84. Sarma, R., et al., *Peptide Solubility Limits: Backbone and Side-Chain Interactions.* J Phys Chem B, 2018. **122**(13): p. 3528-3539.
85. Teufl, M., C.U. Zajc, and M.W. Traxlmayr, *Engineering Strategies to Overcome the Stability-Function Trade-Off in Proteins.* ACS Synth Biol, 2022. **11**(3): p. 1030-1039.
86. *Peptide Tool*. 05.01.2024; 2.2.29.152:[Available from: https://www.protpi.ch/Calculator/PeptideTool.
87. Huang, P.S., S.E. Boyken, and D. Baker, *The coming of age of de novo protein design.* Nature, 2016. **537**(7620): p. 320-7.
88. Bosshard, H.R., D.N. Marti, and I. Jelesarov, *Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings.* J Mol Recognit, 2004. **17**(1): p. 1-16.
89. Behrendt, R., P. White, and J. Offer, *Advances in Fmoc solid-phase peptide synthesis.* J Pept Sci, 2016. **22**(1): p. 4-27.
90. Malencik, D.A. and S.R. Anderson, *Dityrosine as a product of oxidative stress and fluorescent probe.* Amino Acids, 2003. **25**(3-4): p. 233-47.
91. Ghisaidoobe, A.B. and S.J. Chung, *Intrinsic tryptophan fluorescence in the detection and analysis of proteins: a focus on Forster resonance energy transfer techniques.* Int J Mol Sci, 2014. **15**(12): p. 22518-38.
92. Kuzmic, P., *Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase.* Anal Biochem, 1996. **237**(2): p. 260-73.
93. Linse, S., A. Helmersson, and S. Forsén, *Calcium Binding to Calmodulin and Its Globular Domains.* The Journal of Biological Chemistry, 1991. **266**(13): p. 8050-8054.
94. Mirdita, M., et al., *ColabFold: making protein folding accessible to all.* Nat Methods, 2022. **19**(6): p. 679-682.
95. Wu, R., et al., *High-resolution de novo structure prediction from primary sequence.* 2022.
96. Cioci, G., et al., *Beta-propeller crystal structure of Psathyrella velutina lectin: an integrin-like fungal protein interacting with monosaccharides and calcium.* J Mol Biol, 2006. **357**(5): p. 1575-91.
97. Manning, M.C., M. Illangasekare, and R.W. Woody, *Circular dichroism studies of distorted alpha-helices, twisted beta-sheets, and beta turns.* Biophys Chem, 1988. **31**(1-2): p. 77-86.
98. Wolff, M., et al., *Self-Assembly of Exendin-4-Derived Dual Peptide Agonists is Mediated by Acylation and Correlated to the Length of Conjugated Fatty Acyl Chains.* Mol Pharm, 2020. **17**(3): p. 965-978.
99. Wallace, R.W., et al., *Calcium binding domains of calmodulin. Sequence of fill as determined with terbium luminescence.* Journal of Biological Chemistry, 1982. **257**(4): p. 1845-1854.
100. Streicher, W.W. and G.I. Makhatadze, *Calorimetric evidence for a two-state unfolding of the beta-hairpin peptide trpzip4.* J Am Chem Soc, 2006. **128**(1): p. 30-1.
101. Pham, T.L., M. Kovermann, and F. Thomas, *Switchable Zinc(II)-Responsive Globular beta-Sheet Peptide.* ACS Synth Biol, 2022. **11**(1): p. 254-264.
102. Tainer, J.A., et al., *Structure and mechanism of copper, zinc superoxide dismutase.* Nature, 1983. **306**(5940): p. 284-7.

103. Gleason, J.E., et al., *Candida albicans SOD5 represents the prototype of an unprecedented class of Cu-only superoxide dismutases required for pathogen defense.* Proc Natl Acad Sci U S A, 2014. **111**(16): p. 5866-71.
104. Cao, X., et al., *Structures of the G85R variant of SOD1 in familial amyotrophic lateral sclerosis.* J Biol Chem, 2008. **283**(23): p. 16169-77.
105. Kaplan, B.E., et al., *Solid-phase synthesis and characterization of carcinoembryonic antigen (CEA) domains.* J Pept Res, 1998. **52**(4): p. 249-60.
106. Rizzolo, F., et al., *Conventional and microwave-assisted SPPS approach: a comparative synthesis of PTHrP(1-34)NH(2).* J Pept Sci, 2011. **17**(10): p. 708-14.
107. Friligou, I., et al., *Microwave-assisted solid-phase peptide synthesis of the 60-110 domain of human pleiotrophin on 2-chlorotrityl resin.* Amino Acids, 2011. **40**(5): p. 1431-40.
108. Remuzgo, C., et al., *Acanthoscurrin fragment 101-132: total synthesis at 60 degrees C of a novel difficult sequence.* Biopolymers, 2009. **92**(1): p. 65-75.
109. Pedersen, S.L., K.K. Sorensen, and K.J. Jensen, *Semi-automated microwave-assisted SPPS: Optimization of protocols and synthesis of difficult sequences.* Biopolymers, 2010. **94**(2): p. 206-12.
110. Varanda, L.M. and M.T. Miranda, *Solid-phase peptide synthesis at elevated temperatures: a search for and optimized synthesis condition of unsulfated cholecystokinin-12.* J Pept Res, 1997. **50**(2): p. 102-8.
111. Palasek, S.A., Z.J. Cox, and J.M. Collins, *Limiting racemization and aspartimide formation in microwave-enhanced Fmoc solid phase peptide synthesis.* J Pept Sci, 2007. **13**(3): p. 143-8.
112. Liu, J. and J. Song, *Insights into protein aggregation by NMR characterization of insoluble SH3 mutants solubilized in salt-free water.* PLoS One, 2009. **4**(11): p. e7805.
113. Matsumura, Y., et al., *Transient helical structure during PI3K and Fyn SH3 domain folding.* J Phys Chem B, 2013. **117**(17): p. 4836-43.
114. Larson, S.M. and A.R. Davidson, *The identification of conserved interactions within the SH3 domain by alignment of sequences and structures.* Protein Science, 2000. **9**(11): p. 2170-2180.
115. Rulisek, L. and J. Vondrasek, *Coordination geometries of selected transition metal ions (Co2+, Ni2+, Cu2+, Zn2+, Cd2+, and Hg2+) in metalloproteins.* J Inorg Biochem, 1998. **71**(3-4): p. 115-27.
116. Brodin, J.D., et al., *Evolution of metal selectivity in templated protein interfaces.* J Am Chem Soc, 2010. **132**(25): p. 8610-7.
117. Irving, H. and R.J.P. Williams, *Order of Stability of Metal Complexes.* Nature, 1948. **162**(4123): p. 746-747.
118. Smirnova, J., et al., *Evaluation of Zn(2+)- and Cu(2+)-Binding Affinities of Native Cu,Zn-SOD1 and Its G93A Mutant by LC-ICP MS.* Molecules, 2022. **27**(10).
119. Hirose, J., et al., *The pH dependence of apparent binding constants between apo-superoxide dismutase and cupric ions.* Arch Biochem Biophys, 1982. **218**(1): p. 179-86.
120. Pantoliano, M.W., et al., *A pH-dependent superoxide dismutase activity for zinc-free bovine erythrocuprein. Reexamination of the role of zinc in the holoprotein.* J Inorg Biochem, 1982. **17**(4): p. 325-41.
121. Robinett, N.G., et al., *Exploiting the vulnerable active site of a copper-only superoxide dismutase to disrupt fungal pathogenesis.* J Biol Chem, 2019. **294**(8): p. 2700-2713.
122. Pham, T.L., et al., *An Engineered beta-Hairpin Peptide Forming Thermostable Complexes with Zn(II) , Ni(II) , and Cu(II) through a His(3) Site.* Chembiochem, 2023. **24**(3): p. e202200588.
123. Sutherland, M.W. and B.A. Learmonth, *The tetrazolium dyes MTS and XTT provide new quantitative assays for superoxide and superoxide dismutase.* Free Radic Res, 1997. **27**(3): p. 283-9.
124. Vincent, A., et al., *An easy-to-implement combinatorial approach involving an activity-based assay for the discovery of a peptidyl copper complex mimicking superoxide dismutase.* Chem Commun (Camb), 2020. **56**(3): p. 399-402.

125. Klug, D., J. Rabani, and I. Fridovich, *A Direct Demonstration of the Catalytic Action of Superoxide Dismutase through the Use of Pulse Radiolysis.* Journal of Biological Chemistry, 1972. **247**(15): p. 4839-4842.
126. Kotynia, A., et al., *The Analysis of Cu(II)/Zn(II) Cyclopeptide System as Potential Cu,ZnSOD Mimic Center.* Int J Pept Res Ther, 2017. **23**(4): p. 431-439.
127. Strohalm, M., et al., *mMass 3: a cross-platform software environment for precise analysis of mass spectrometric data.* Anal Chem, 2010. **82**(11): p. 4648-51.
128. Pace, C.N., et al., *How to measure and predict the molar absorption coefficient of a protein.* Protein Sci, 1995. **4**(11): p. 2411-23.
129. Golynskiy, M.V., et al., *Metal Binding Studies and EPR Spectroscopy of the Manganese Transport Regulator MntR.* Biochemistry, 2006. **45**(51): p. 15359-15372.
130. Wada, H., T. Ishizuki, and G. Nakagawa, *Synthesis of 2-(2-thiazolylazo)-4-methyl-5-(sulfopropyl-amino) benzoic acid and the application to the flow-injection analysis of copper(II).* Mikrochimica Acta, 1983. **81**(3-4): p. 235-244.

8 Appendix

8.1 Analytical data

8.1.1 Slam series

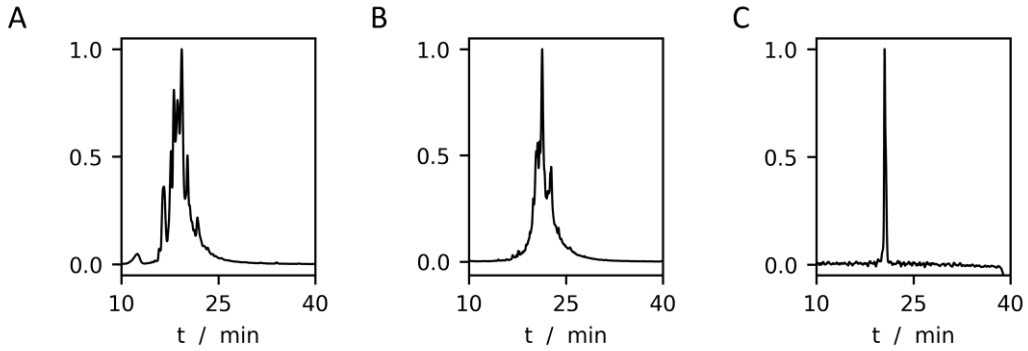


Figure 58. HPLC Chromatograms of Slam1 and Slam2. A) Preparative chromatogram of crude Slam1 in 20-65 % B in A; B) preparative chromatogram of crude Slam2 in 20-50 % B in A; C) analytical chromatogram of purified Slam2 in 20-50 % B in A.

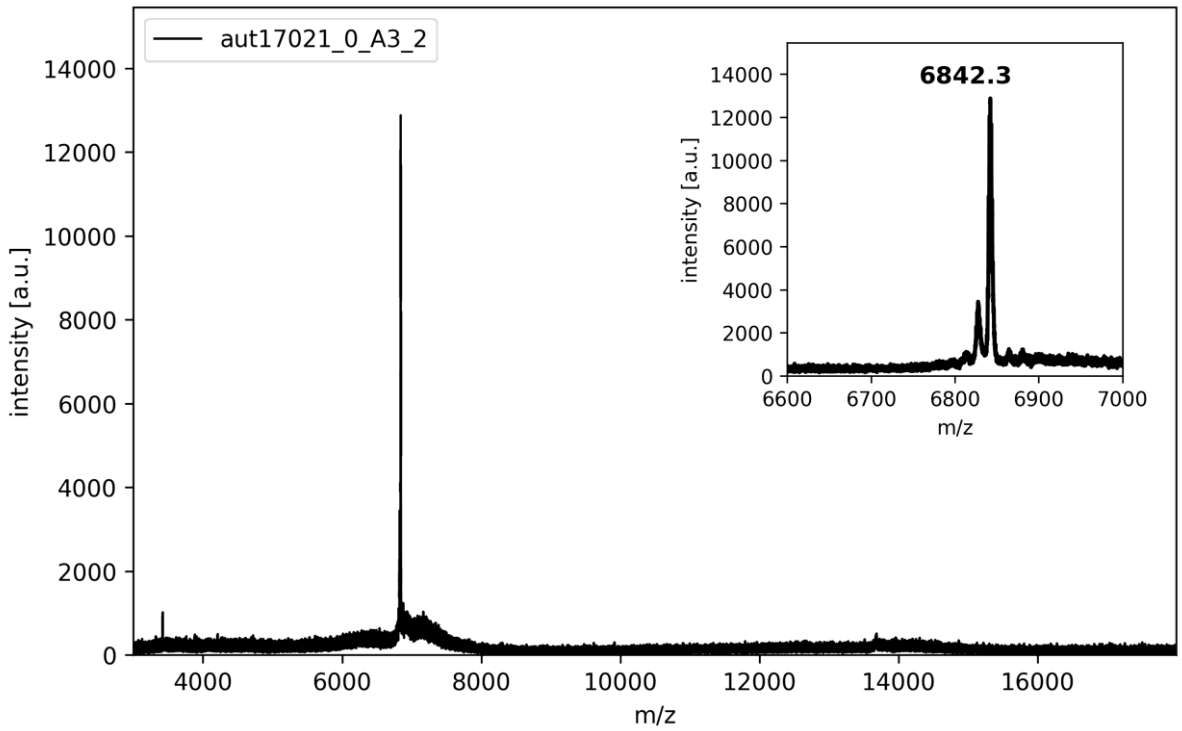


Figure 59. MALDI-TOF Mass spectrum of purified Slam2. Expected mass: 6840.57 m/z, found mass: 6842.3 m/z assigned to $[M+H]^+$.

8.1.2 WWcalm series

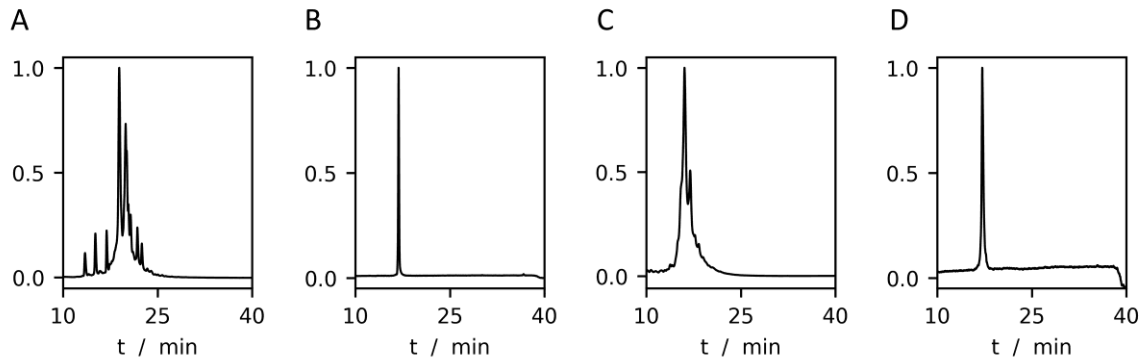


Figure 60. HPLC chromatograms of WWcalm1 and WWcalm2. A) Preparative chromatogram of crude WWcalm1 in 20-80 % B in A; B) analytical chromatogram of purified WWcalm1 in 25-50 % B in A; C) preparative chromatogram of crude WWcalm2 in 20-80 % B in A; C) analytical chromatogram of purified WWcalm2 in 20-50 % B in A.

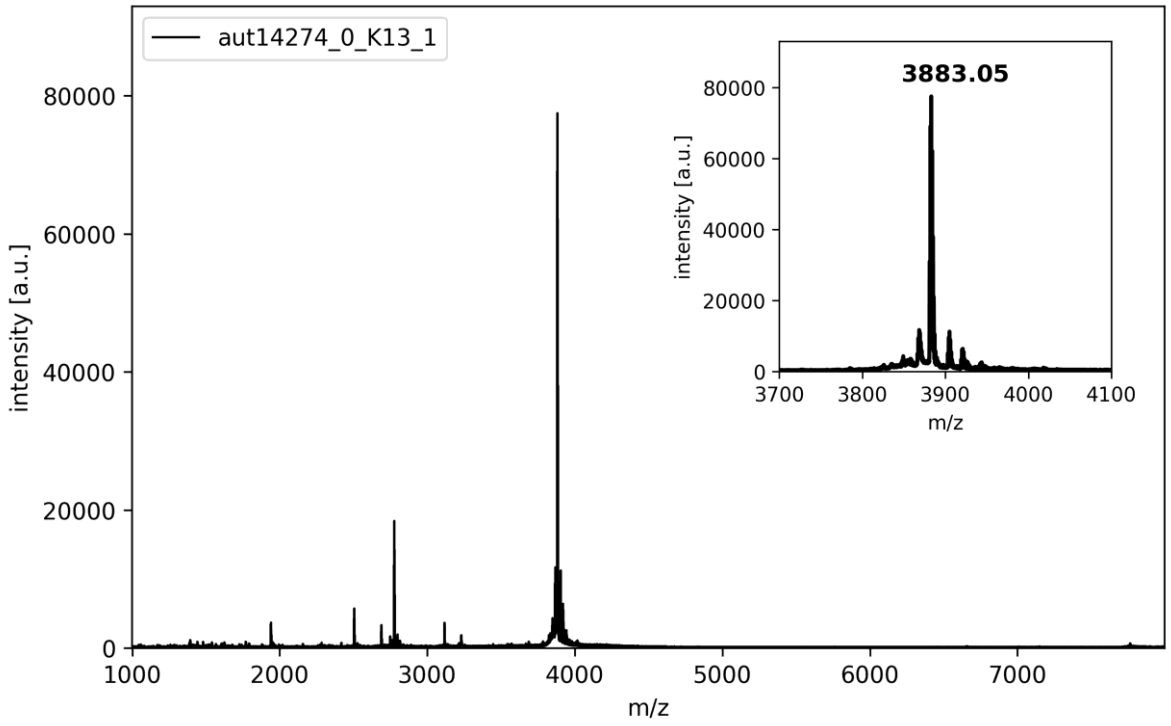


Figure 61. MALDI-TOF Mass spectrum of purified WWcalm1. Expected mass: 3882.38 m/z, found mass: 3883.05 m/z assigned to $[M+H]^+$.

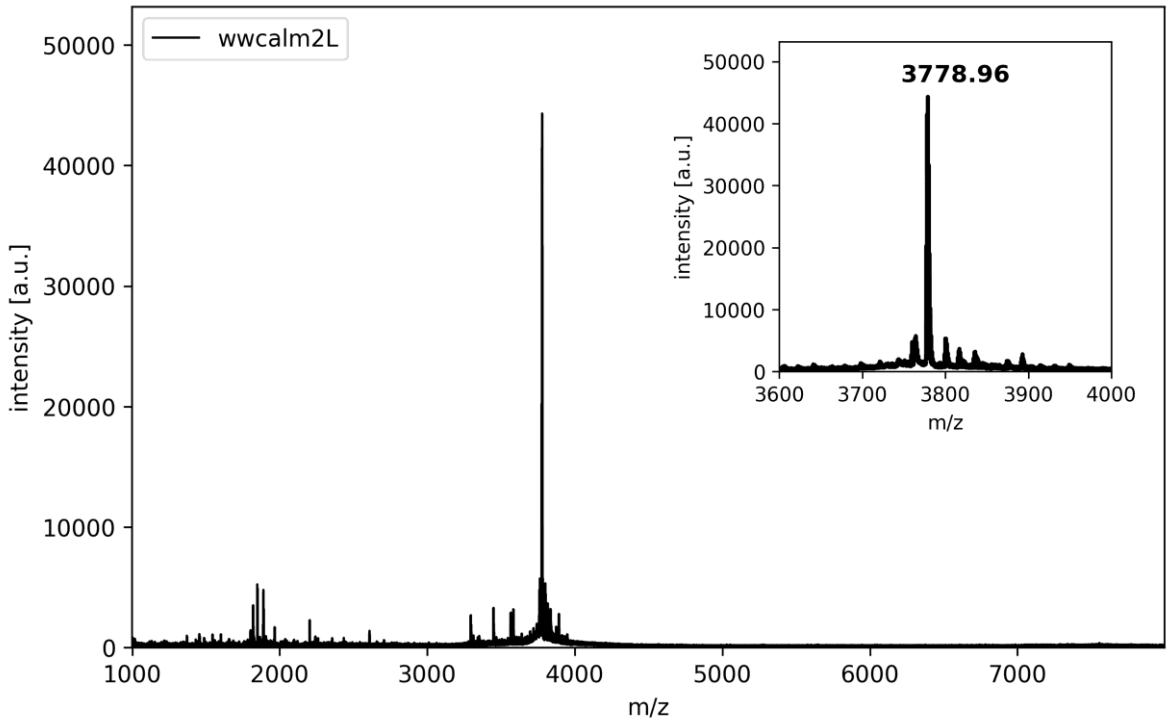


Figure 62. MALDI-TOF Mass spectrum of purified WWcalm2. Expected mass: 3777.03 m/z, found mass: 3778.96 m/z assigned to $[M+H]^+$.

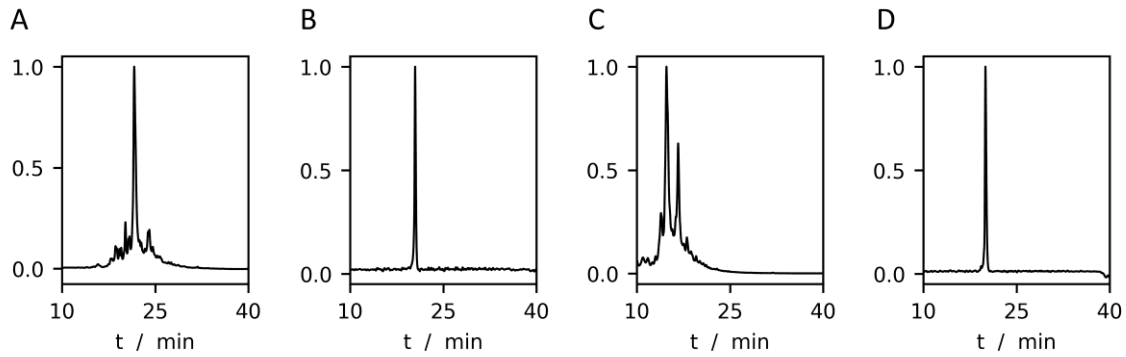


Figure 63. HPLC chromatograms of WWcalm3 and WWcalm4. A) Preparative chromatogram of crude WWcalm3 in 20-50 % B in A; B) analytical chromatogram of purified WWcalm3 in 20-50 % B in A; C) preparative chromatogram of crude WWcalm4 in 25-65 % B in A; C) analytical chromatogram of purified WWcalm4 in 20-50 % B in A.

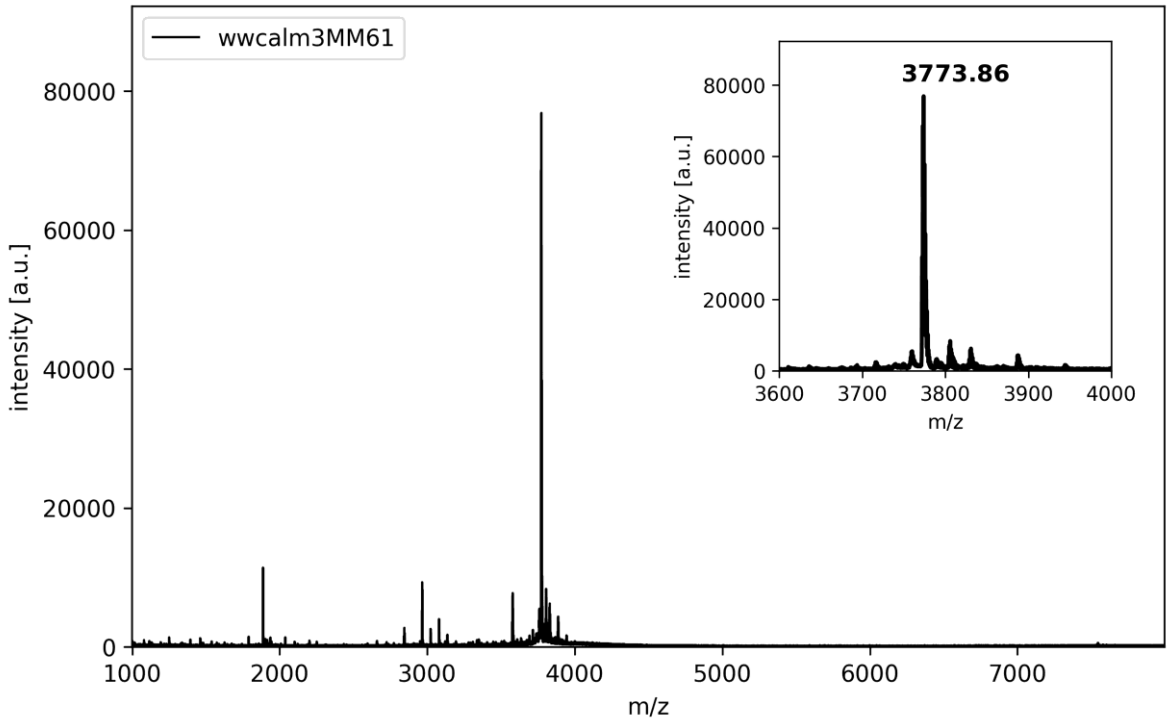


Figure 64. MALDI-TOF Mass spectrum of WWcalm3. Expected mass: 3772.05 m/z, found mass: 3773.86 m/z assigned to $[M+H]^+$.

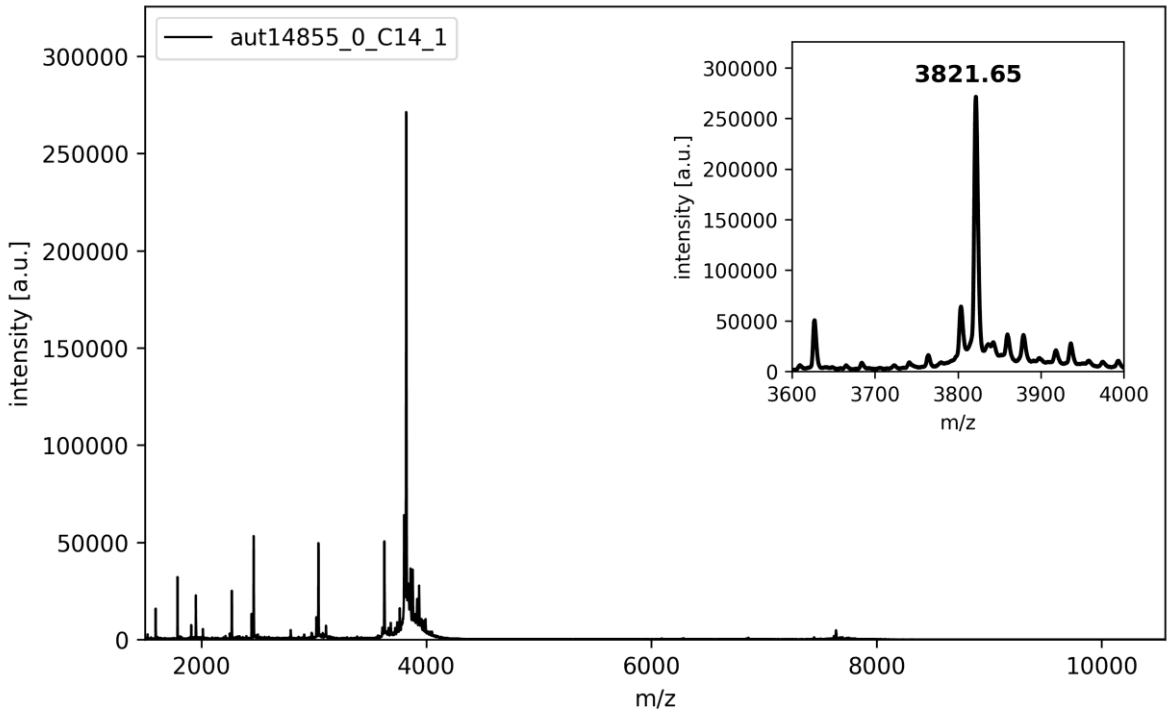


Figure 65. MALDI-TOF Mass spectrum of WWcalm4. Expected mass: 3820.10 m/z, found mass: 3821.65 m/z assigned to $[M+H]^+$.

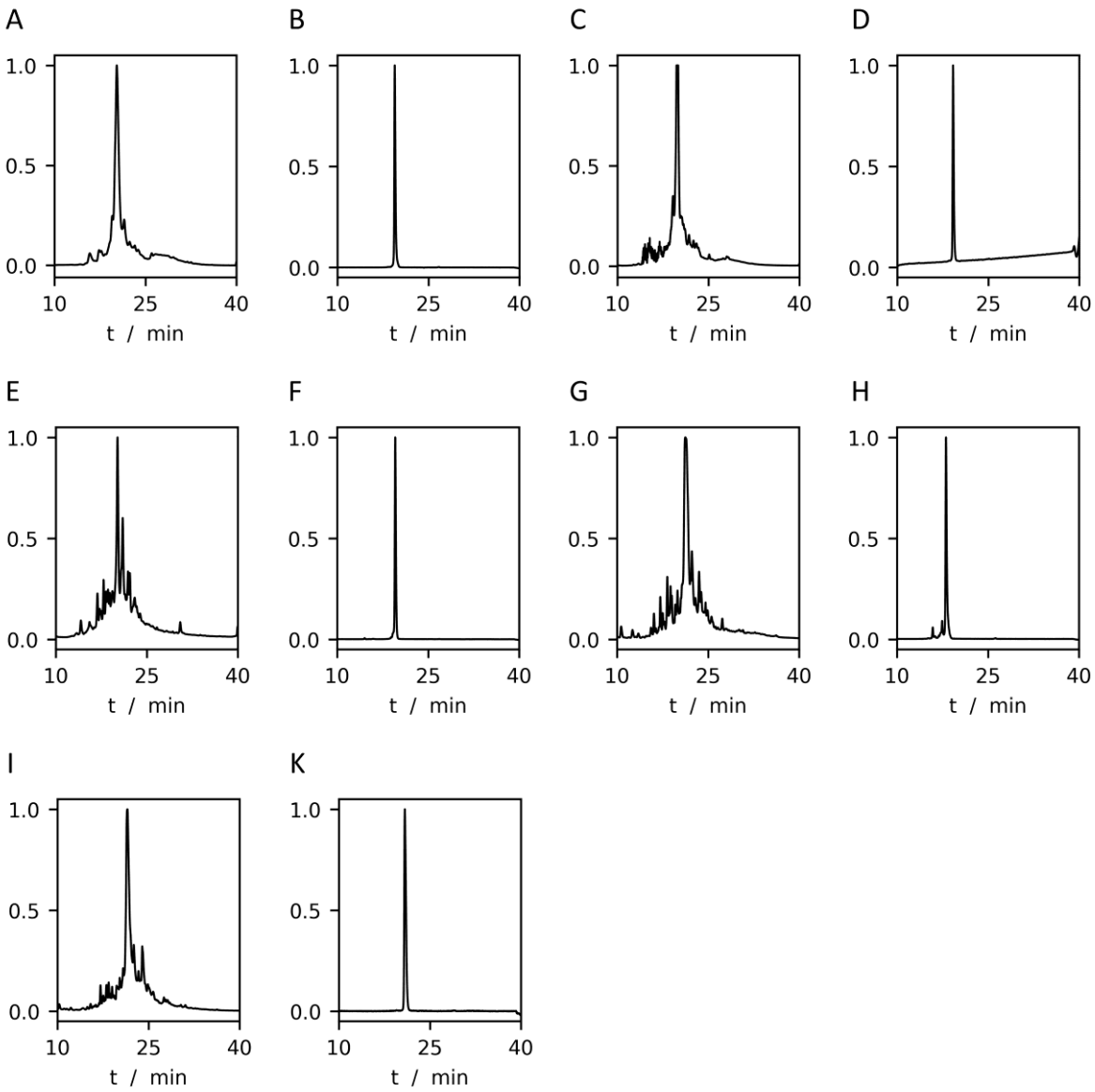


Figure 66. HPLC chromatograms of the WWcalm3.x series and WWcalm4.1. All chromatograms in 20-50 % B in A; A) preparative and (B) analytical chromatogram of WWcalm3.1; C) preparative and (D) analytical chromatogram of WWcalm3.2; E) preparative and (F) analytical chromatogram of WWcalm3.3; G) preparative and (H) analytical chromatogram of WWcalm3.4; I) preparative and (K) analytical chromatogram of WWcalm4.1.

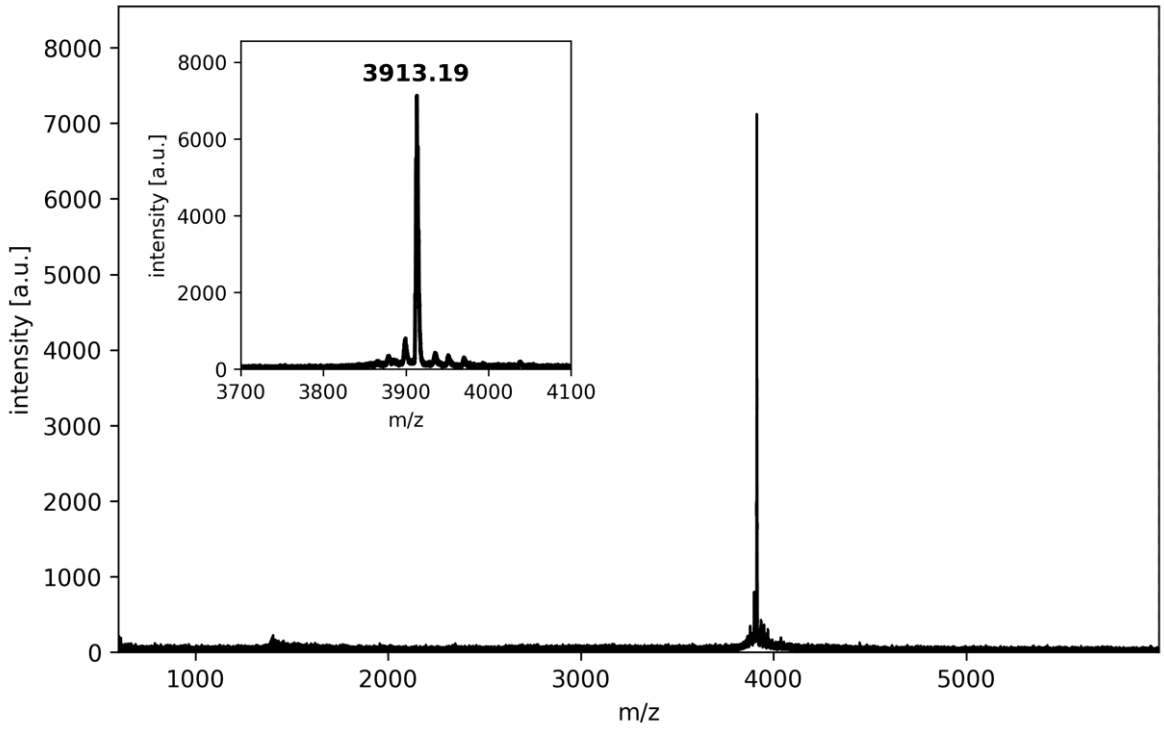


Figure 67. MALDI-TOF Mass spectrum of WWcalm3.1. Expected mass: 3913.22 m/z, found mass: 3913.19 m/z assigned to $[M]^+$.

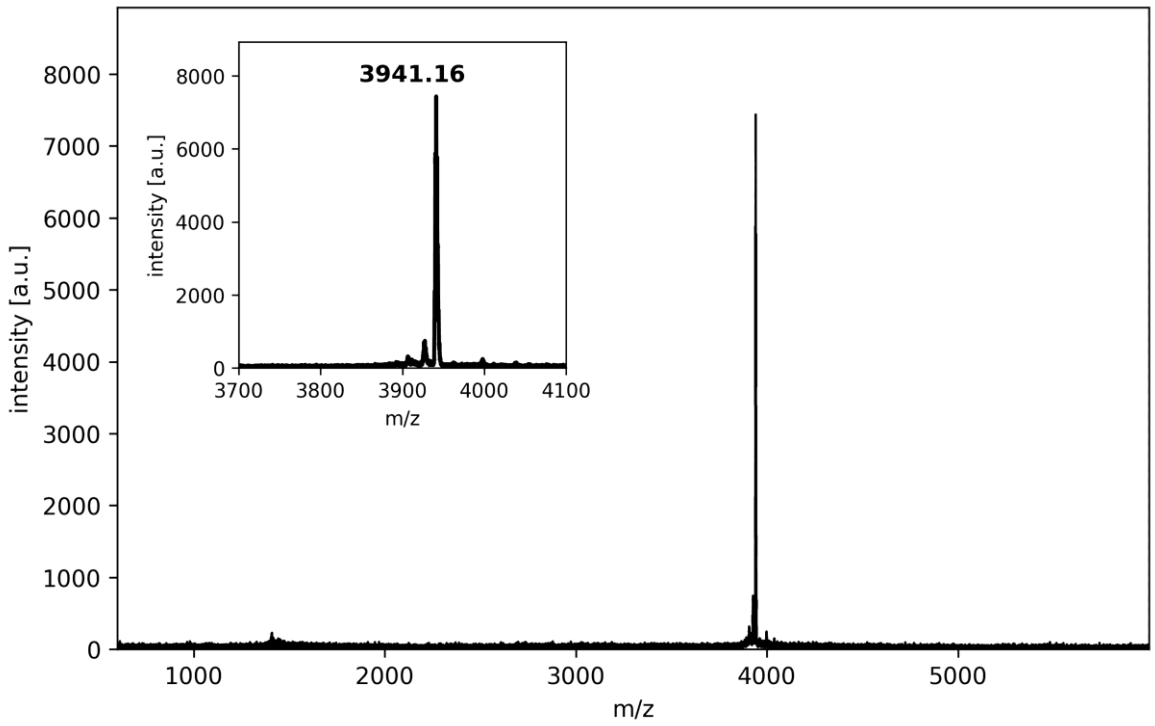


Figure 68. MALDI-TOF Mass spectrum of WWcalm3.2. Expected mass: 3941.28 m/z, found mass: 3941.16 m/z assigned to $[M]^+$.

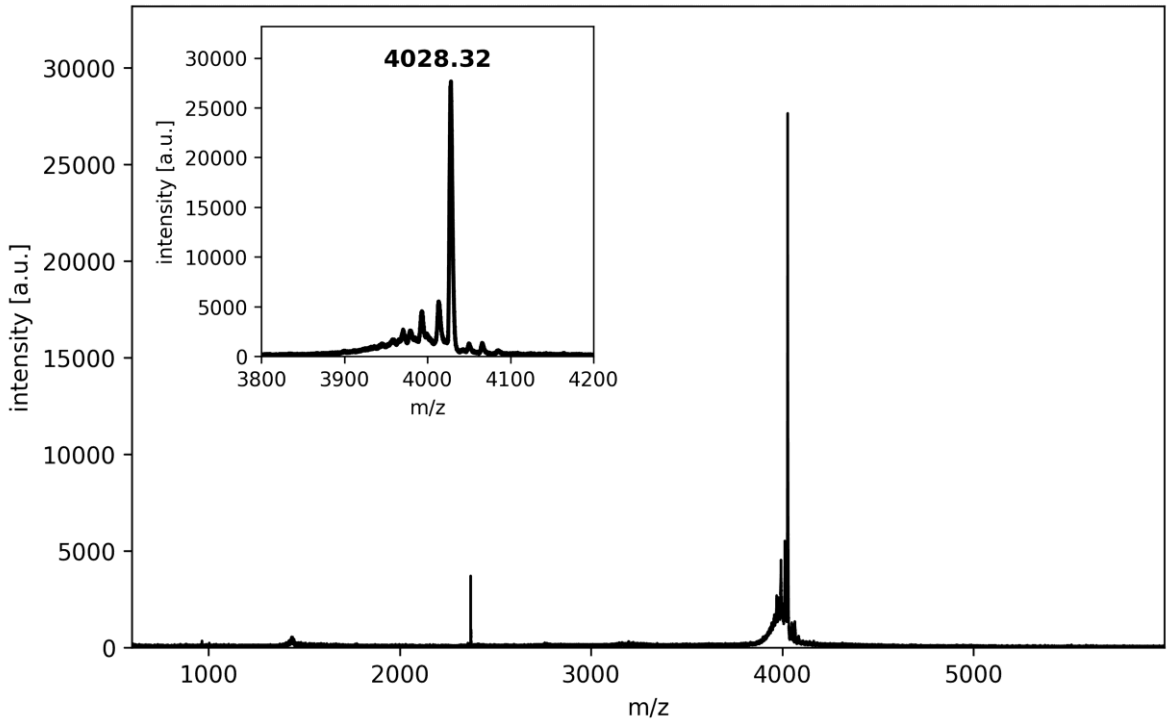


Figure 69. MALDI-TOF Mass spectrum of WWcalm3.3. Expected mass: 4027.41 m/z, found mass: 4028.32 m/z assigned to $[M+H]^+$.

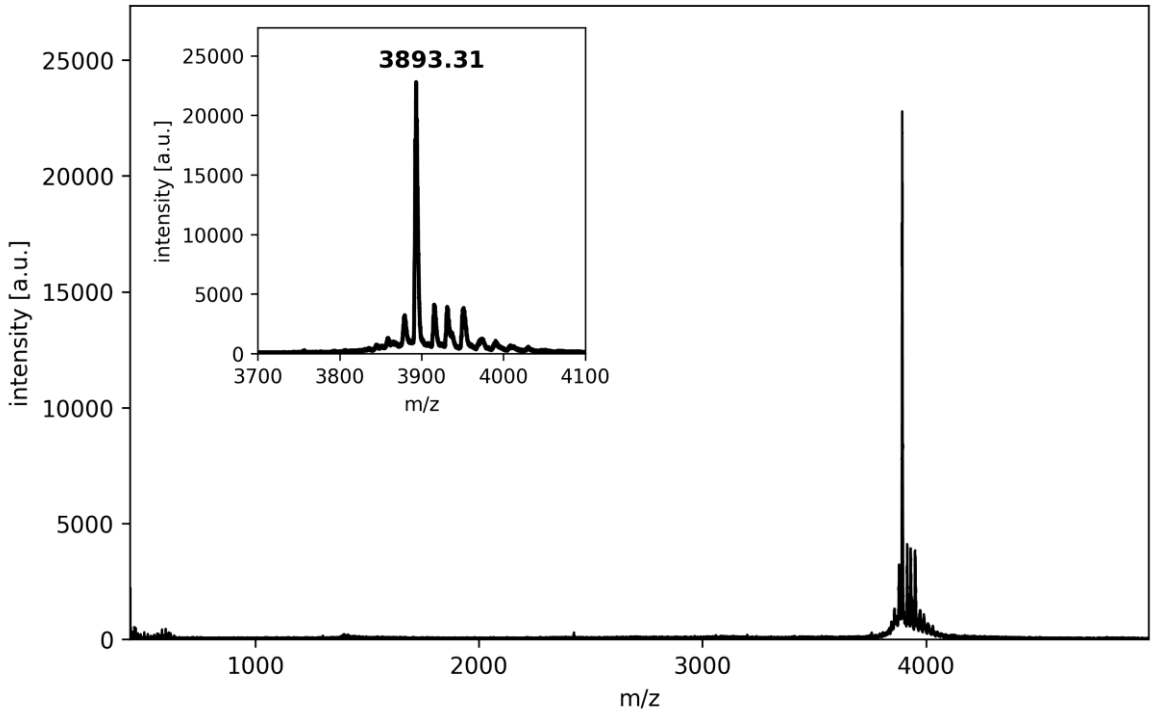


Figure 70. MALDI-TOF Mass spectrum of WWcalm3.4. Expected mass: 3892.21 m/z, found mass: 3893.31 m/z assigned to $[M+H]^+$.

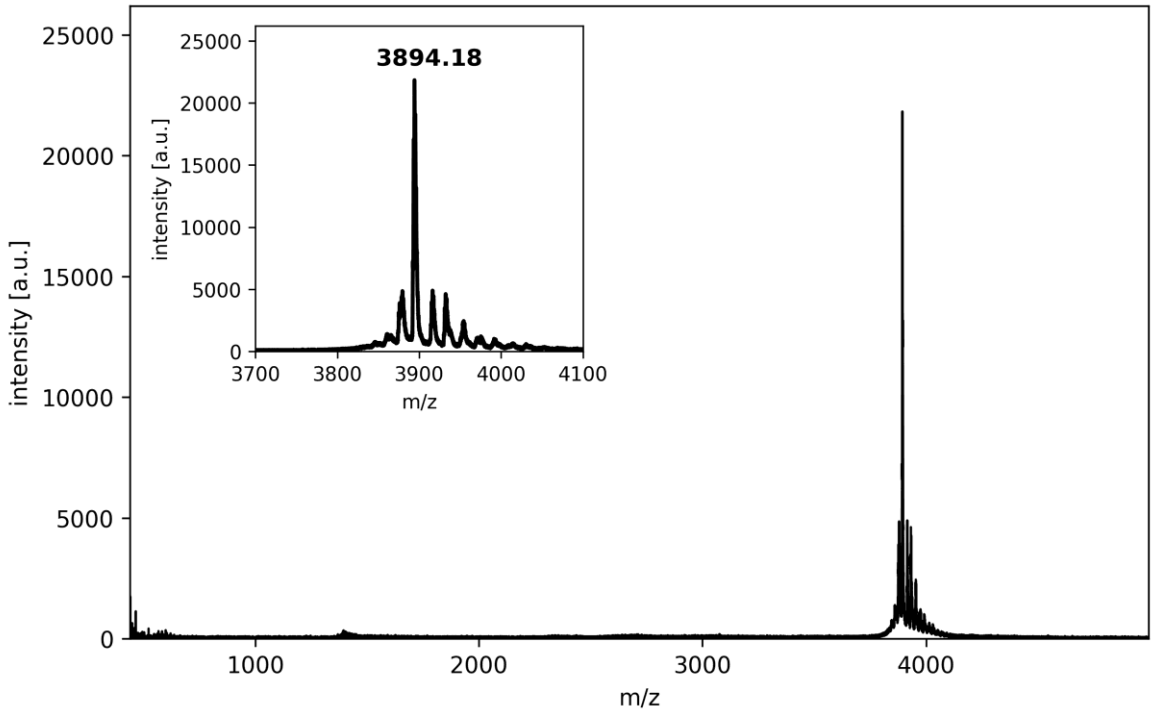


Figure 71. MALDI-TOF Mass spectrum of WWcalm4.1. Expected mass: 3893.19 m/z, found mass: 3894.18 m/z assigned to $[M+H]^+$.

8.1.3 ScaX

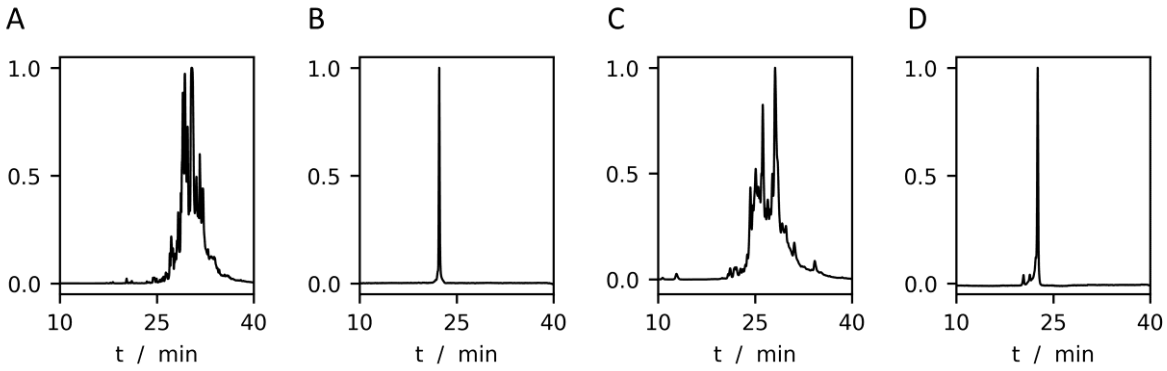


Figure 72. HPLC chromatograms of ScaX and its Y3W mutat. A) Preparative chromatogram in 10-50 % B in A and (B) analytical chromatogram in 20-50 % B in A of ScaX; C) preparative and (D) analytical chromatogram of ScaX Y3W.

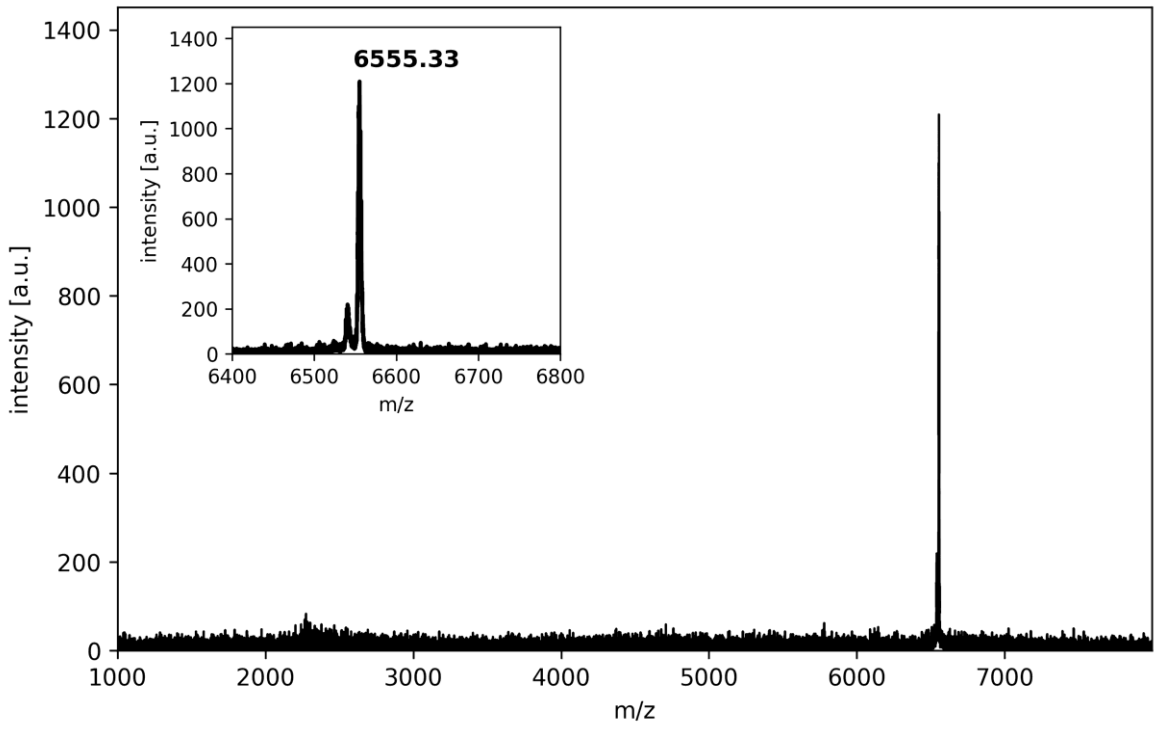


Figure 73. MALDI-TOF Mass spectrum of ScaX. Expected mass: 6555.10 m/z, found mass: 6555.33 m/z assigned to $[M]^+$.

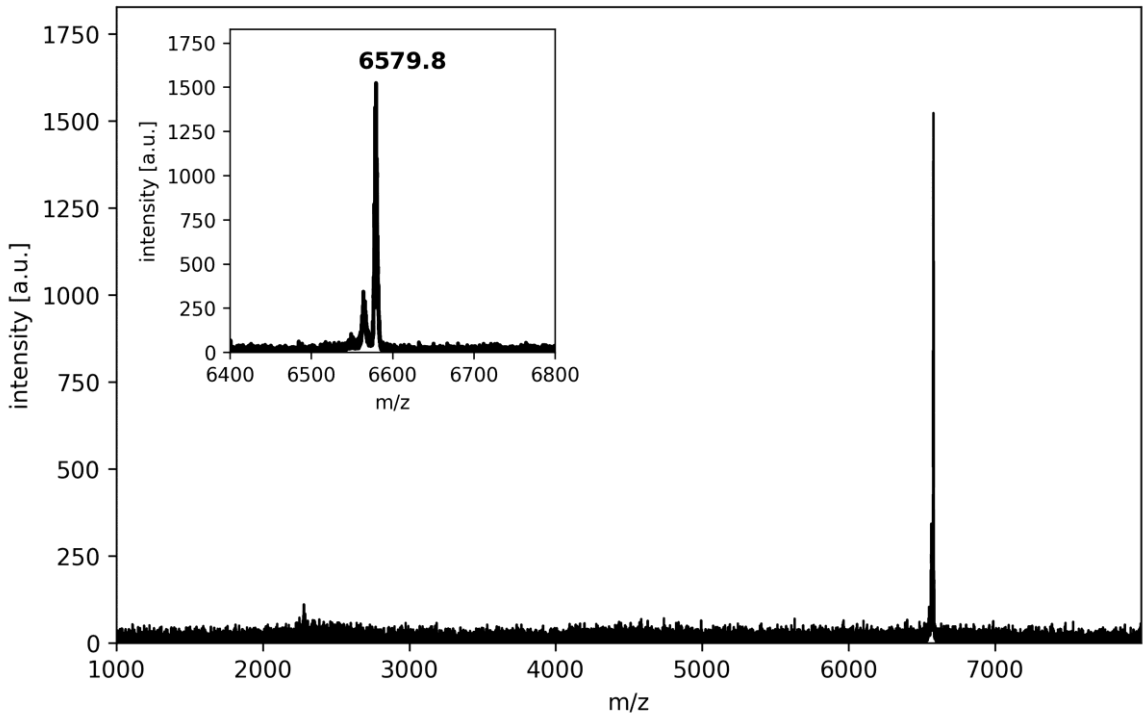


Figure 74. MALDI-TOF Mass spectrum of ScaX Y3W. Expected mass: 6578.14 m/z, found mass: 6579.8 m/z assigned to $[M+H]^+$.

8.1.4 WWturn

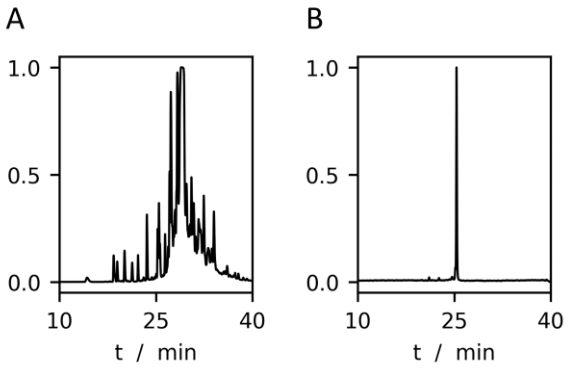


Figure 75. HPLC chromatograms of WWturn. A) Preparative chromatogram and (B) analytical chromatogram in 10-50 % B in A of WWturn.

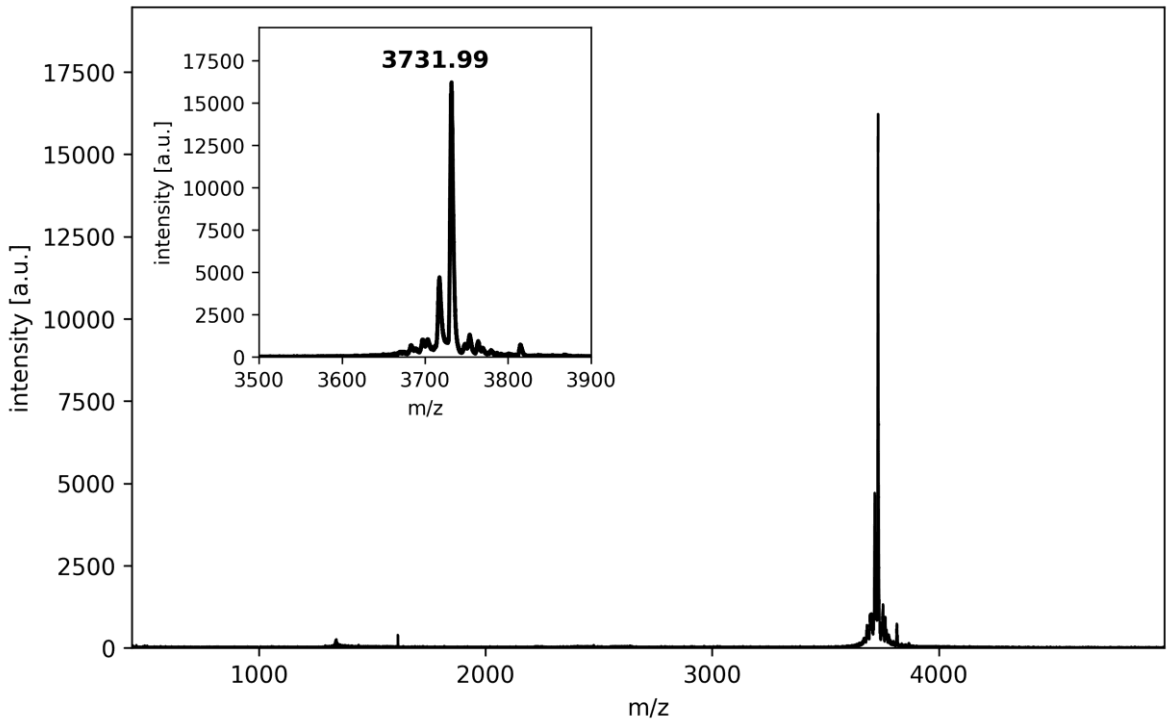


Figure 76. MALDI-TOF Mass spectrum of WWturn. Expected mass: 3730.16 m/z, found mass: 3731.99 m/z assigned to $[M+H]^+$.

8.1.5 WWfan

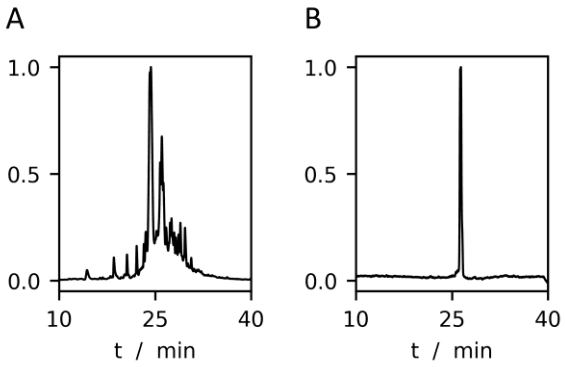


Figure 77. HPLC chromatograms of WWfan. A) Preparative chromatogram and (B) analytical chromatogram in 10-50 % B in A of WWfan.

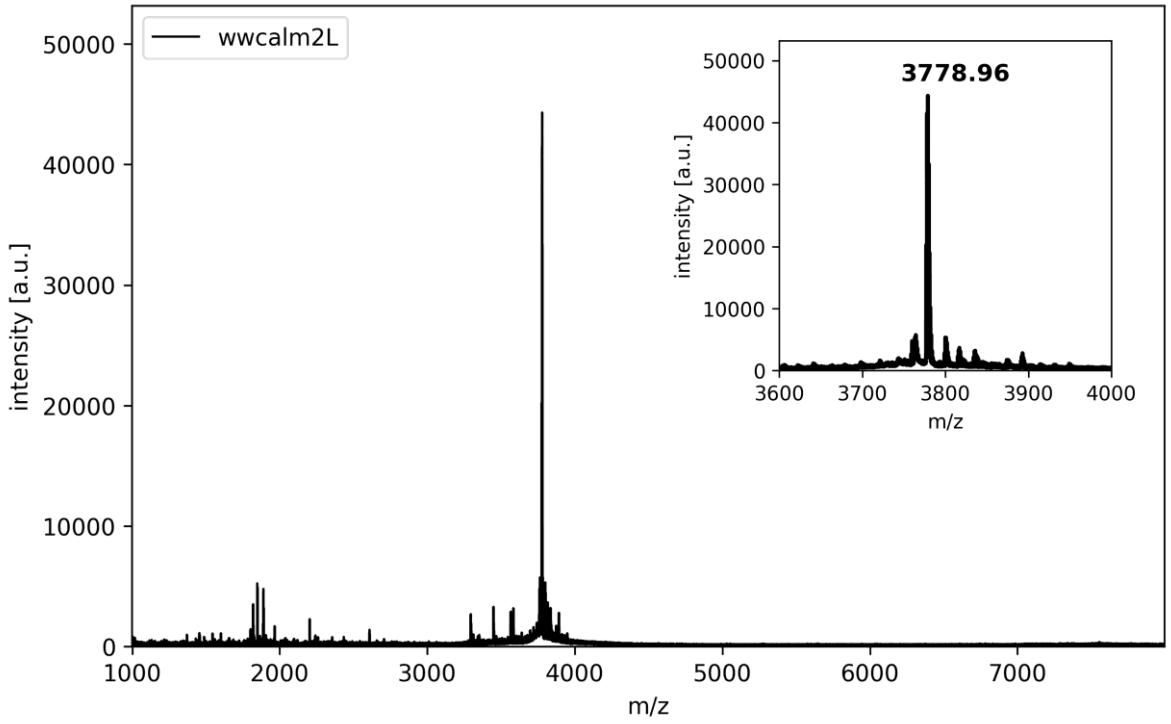


Figure 78. MALDI-TOF Mass spectrum of WWfan. Expected mass: 3777.03 m/z, found mass: 3778.96 m/z assigned to $[M+H]^+$.

8.1.6 Calmcage

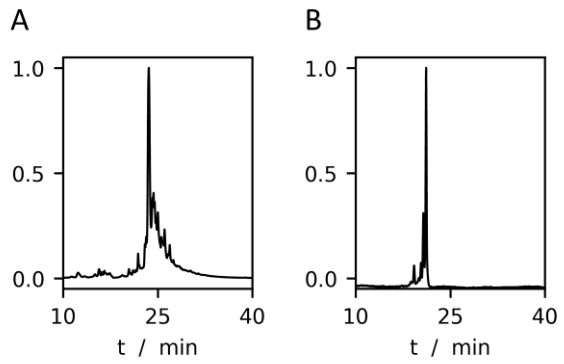


Figure 79. HPLC chromatograms of Calmcage. A) Preparative chromatogram and (B) analytical chromatogram in 20-60 % B in A of Calmcage.

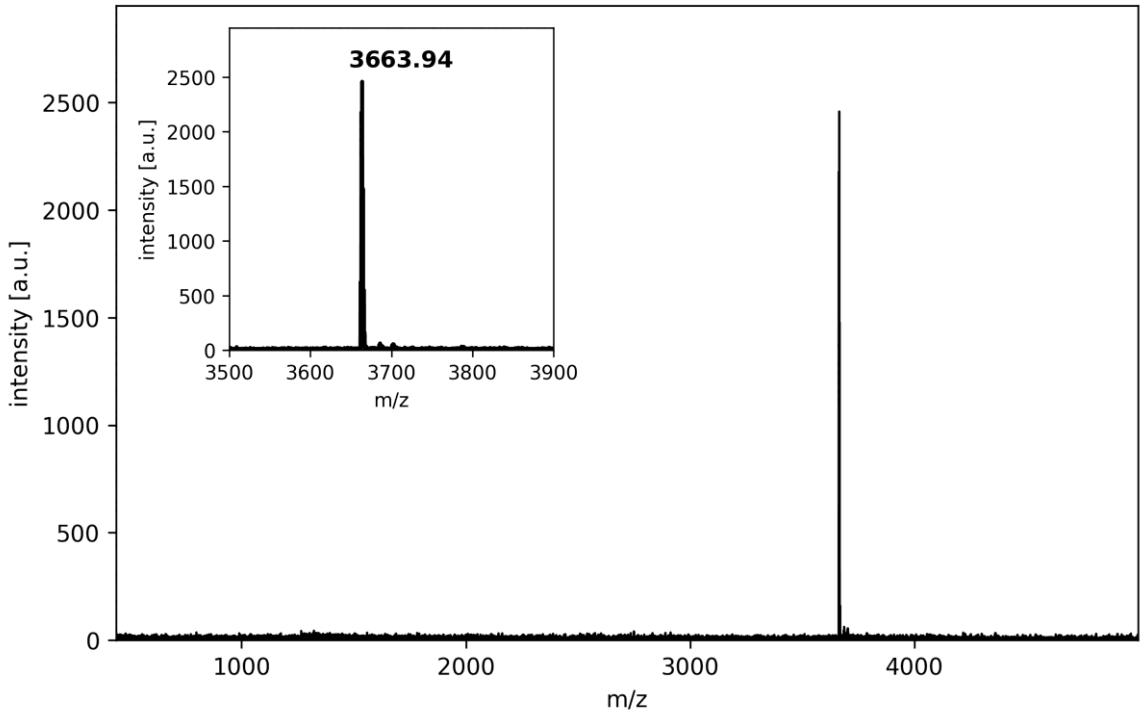


Figure 80. MALDI-TOF Mass spectrum of Calmcage. Expected mass: 3663.91 m/z, found mass: 3663.94 m/z assigned to $[M]^+$.

8.1.7 Calmzip series

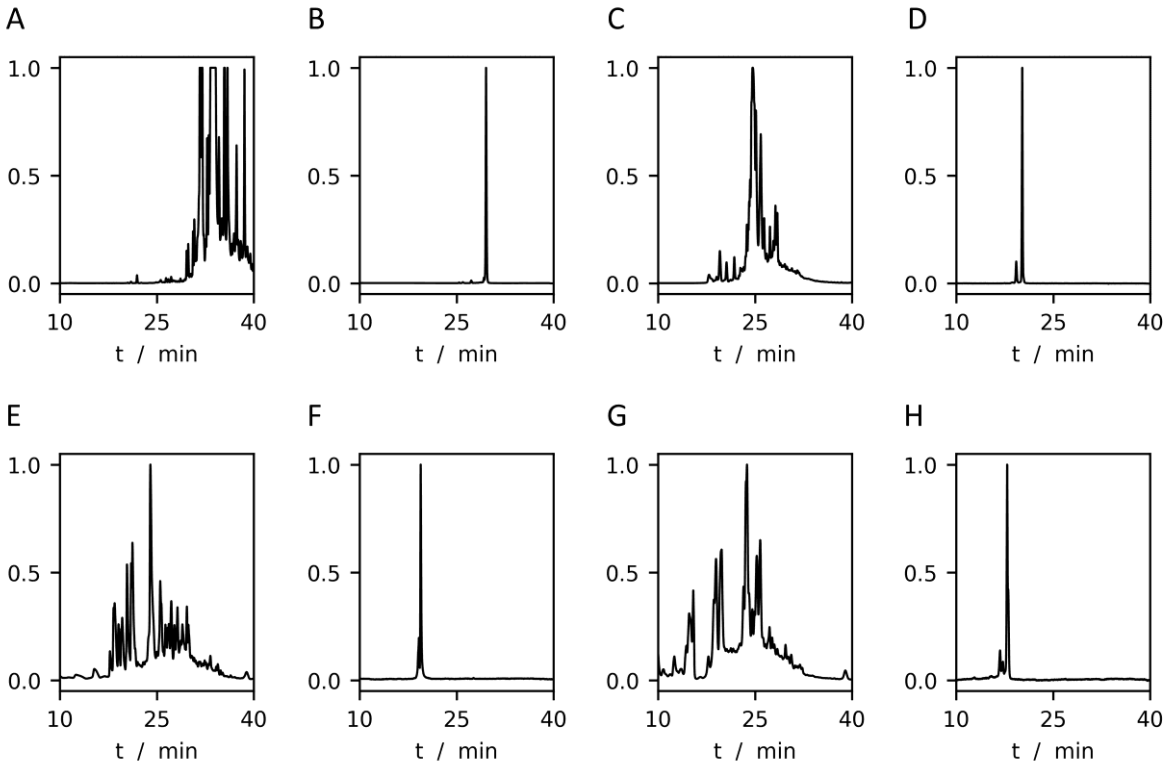


Figure 81. HPLC chromatograms of the Calmzip series. A) Preparative and (B) analytical chromatogram in 5-40 % B in A of Calmzip1; C) preparative chromatogram in 10-50 B in A and (D) analytical chromatogram in 20-60 B in A of Calmzip2; E) preparative and (F) analytical chromatogram in 25-60 % B in A of Calmzip3; G) preparative and (H) analytical chromatogram in 25-60 % B in A of Calmzip3.1.

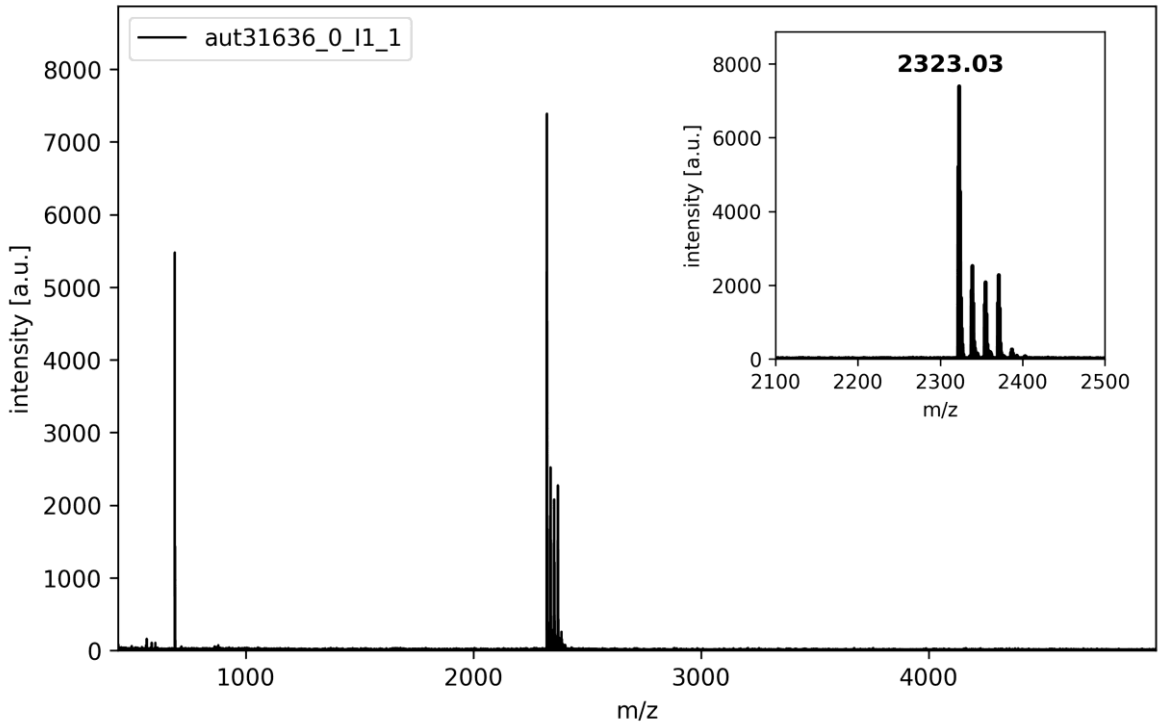


Figure 82. MALDI-TOF Mass spectrum of Calmzip1. Expected mass: 2322.45 m/z, found mass: 2323.03 m/z assigned to $[M+H]^+$.

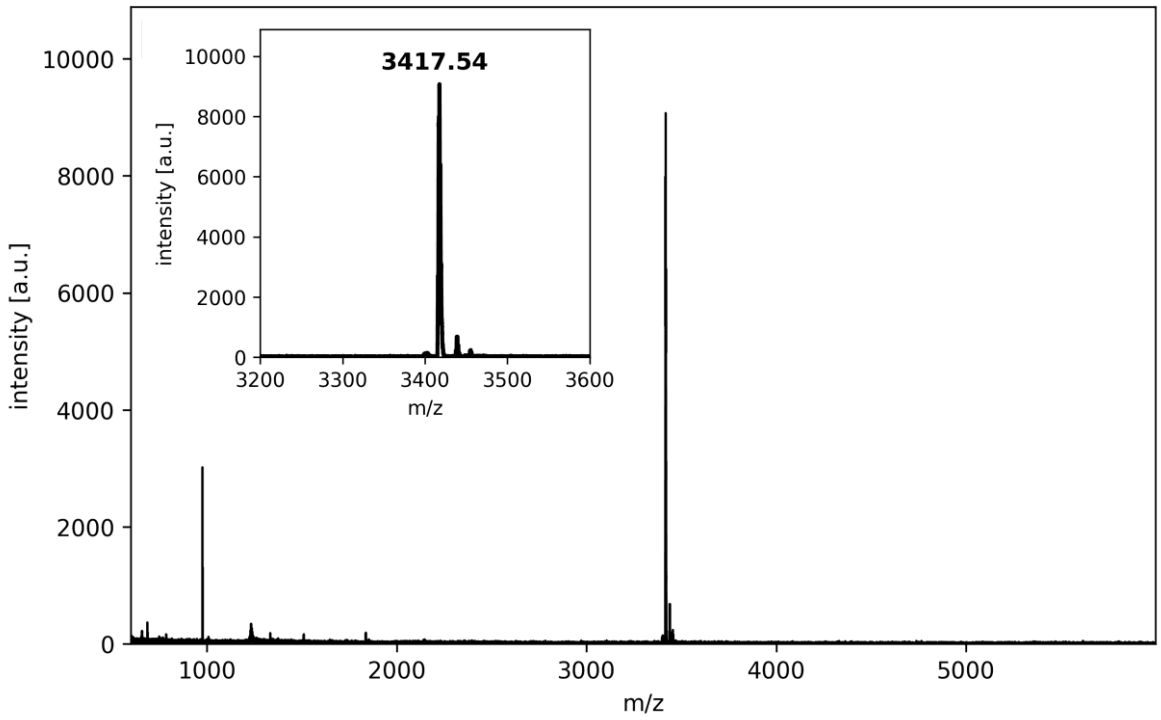


Figure 83. MALDI-TOF Mass spectrum of Calmzip2. Expected mass: 3416.66 m/z, found mass: 3417.54 m/z assigned to $[M+H]^+$.

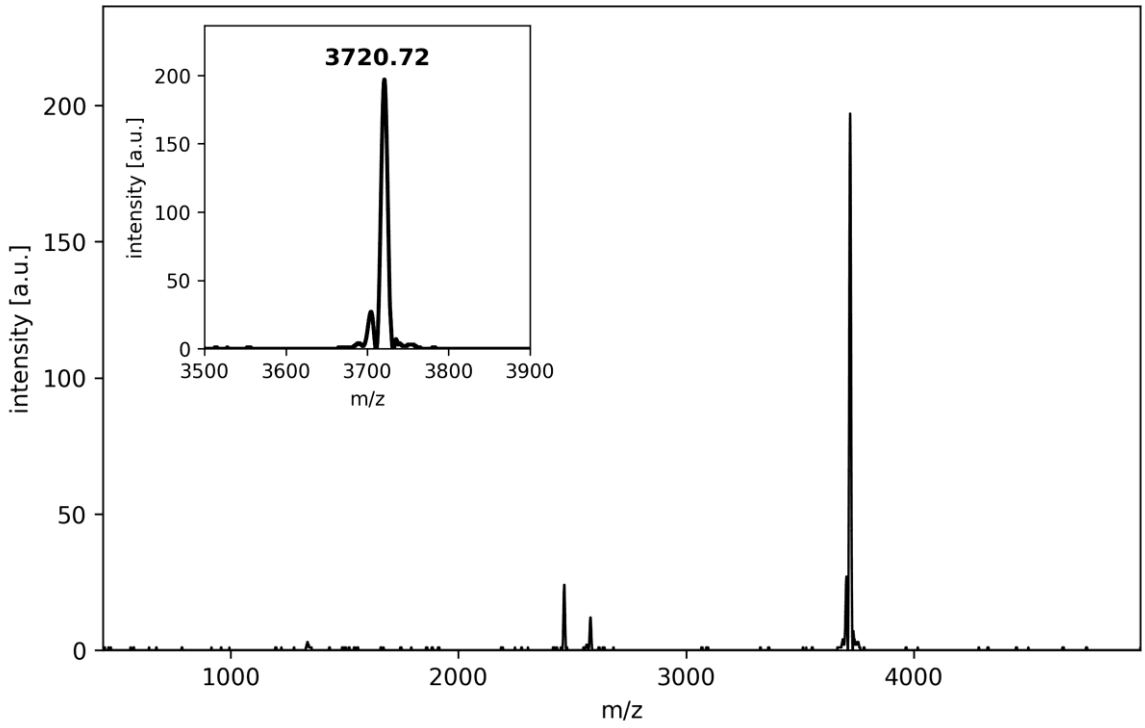


Figure 84. MALDI-TOF Mass spectrum of WWcalm2. Expected mass: 3720.16 m/z, found mass: 3720.72 m/z assigned to $[M]^+$.

8.1.8 SO1

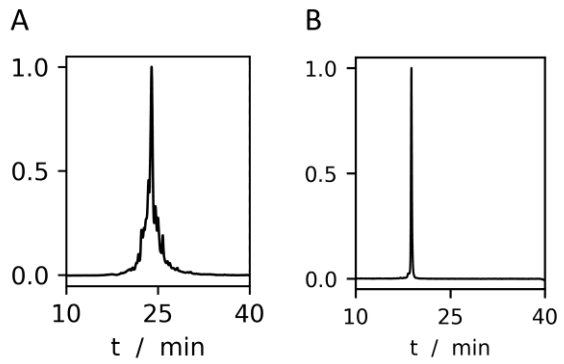


Figure 85. HPLC chromatograms of SO1. A) Preparative chromatogram and (B) analytical chromatogram in 20-50 % B in A of SO1.

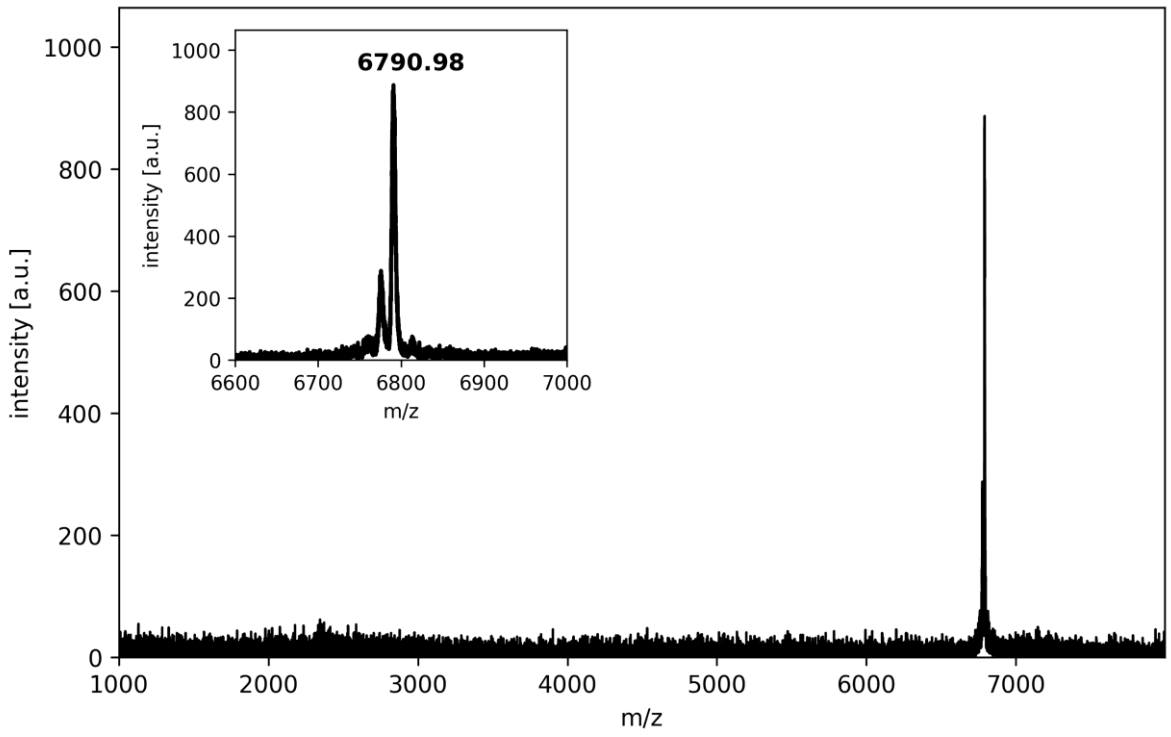


Figure 86. MALDI-TOF Mass spectrum of SO1. Expected mass: 6789.54 m/z, found mass: 6790.98 m/z assigned to $[M+H]^+$.

8.2 Scripts

8.2.1 Design scripts

Script 1. Bash script to call Rosetta3 Match for Slam.

```
counter=1

while [ $counter -le 3 ]

do

/mnt/c/…/rosetta/main/source/bin/match.static.linuxgccrelease @flags_match;

((counter++))

done

echo All done
```

Script 2. Flag file for Rosetta3 Match.

```
-database /mnt/c/…/rosetta/main/database/
-match::scaffold_active_site_residues PATHtoPOS
-match::geometric_constraint_file PATHtoCST
-output_matches_per_group 1
-match:consolidate_matches
-mute all
-packing
-ex1
-ex2
-extrachi_cutoff 0
-use_input_sc
-match:filter_colliding_upstream_residues
-match:filter_upstream_downstream_collisions
-match:updown_collision_tolerance 0.3
-match::bump_tolerance 0.3
-in:ignore_unrecognized_res
-match_grouper SameSequenceAndDSPositionGrouper
-output_format PDB
-enumerate_ligand_rotamers
-only_enumerate_non_match_redundant_ligand_rotamers
-match:euclid_bin_size 0.8
-match:euler_bin_size  10.0
-out::file::output_virtual
-dynamic_grid_refinement
```

Script 3. Positions (POS) file for SH3-based designs.

```
5 7 8 9 10 11 12 13 14 15 16 17 19 21 22 23 24 25 27 29 30 31 32 33 34 35 36 38 40 41
42 43 44 45 46 47 48 50 51 52 53
```

Script 4. Geometric constraint (CST) file for Slam1.

```
CST::BEGIN
TEMPLATE:: ATOM_MAP: 1 atom_name: CA V1 V2
TEMPLATE:: ATOM_MAP: 1 residue3: CA
TEMPLATE:: ATOM_MAP: 2 atom_type: OOC
TEMPLATE:: ATOM_MAP: 2 residue1: D

CONSTRAINT:: distanceAB: 2.47 0.2 40.0 1 1
CONSTRAINT:: angle_A: 136.2  6.0 40.0 360. 1
CONSTRAINT:: angle_B:  1.3  5.0 40.0 360. 1
CONSTRAINT:: torsion_A: 166.2 15.0 40.0 360.0 2
CONSTRAINT:: torsion_AB: 5.8 40.0 40.0 180. 4
CONSTRAINT:: torsion_B: 60.0 10.0 40.0 180. 1
CST::END

CST::BEGIN
TEMPLATE:: ATOM_MAP: 1 atom_name: CA V1 V2
TEMPLATE:: ATOM_MAP: 1 residue3: CA
TEMPLATE:: ATOM_MAP: 2 atom_type: OOC
TEMPLATE:: ATOM_MAP: 2 residue1: D

CONSTRAINT:: distanceAB: 2.52 0.2 40.0 1 1
CONSTRAINT:: angle_A: 148.7  6.0 40.0 360. 1
CONSTRAINT:: angle_B: 89.7  5.0 40.0 360. 1
CONSTRAINT:: torsion_A: -43.2 15.0 40.0 360.0 2
CONSTRAINT:: torsion_AB: 30.0 40.0 40.0 180. 4
CONSTRAINT:: torsion_B: 2.2 10.0 40.0 180. 1
CST::END
```

Script 5. Bash script to call Rosetta3 Design of Slam.

```
/mnt/c/…/rosetta/main/source/bin/enzyme_design.static.linuxgccrelease
@flags_enzdes       -l output_list.txt  -out:file:o scorefile.txt;

/mnt/c/…/rosetta/main/source/bin/enzyme_design.static.linuxgccrelease
@flags_enzdes       -l output_list.txt  -out:file:o scorefile.txt;

/mnt/c/…/rosetta/main/source/bin/enzyme_design.static.linuxgccrelease
@flags_enzdes       -l output_list.txt  -out:file:o scorefile.txt;

/mnt/c/…/rosetta/main/source/bin/enzyme_design.static.linuxgccrelease
@flags_enzdes       -l output_list.txt  -out:file:o scorefile.txt;
```

Script 6. Flags for Rosetta3 Design of Slam.

```
-score::weights /mnt/c/…/
rosetta/main/database/scoring/weights/enzdes.wts
-database /mnt/c/…/rosetta/main/database/
-out:file::o scores.out
-out:overwrite
-out:file::output_virtual
-jd2::enzdes_out
-enzdes
-bb_min_allowed_dev 0.1
-loop_bb_min_allowed_dev 0.3
-enzdes:cstfile
/mnt/c/Users/Florian/Documents/RosettaScripts/input_files/test_cst/CA_LG4_Donly.cst
-extra_res_fa /mnt/c/Users/Florian/Documents/RosettaScripts/input_files/CAX.params
-detect_design_interface
-cut1 6.0
-cut2 8.0
-cut3 10.0
-cut4 12.0
-packing
 -ex1
 -ex2
 -extrachi_cutoff 1
 -use_input_sc
 -flip_HNQ
 -no_optH false
 -optH_MCA false
 -linmem_ig 10
-nblist_autoupdate
-ligand::old_estat
-parser:protocol enzdes_2.xml
-run:preserve_header
-jd2:ntrials 20
-restore_pre_talaris_2013_behavior
-ignore_unrecognized_res
```

Script 7. Parameter (PARAMS) file for octahedral Ca(II).

```
NAME CAX
IO_STRING CAX Z
TYPE LIGAND
AA UNK
ATOM CA Ca2p X 2.00
ATOM V1 VIRT X 0.00
ATOM V2 VIRT X 0.00
ATOM V3 VIRT X 0.00
ATOM V4 VIRT X 0.00
ATOM V5 VIRT X 0.00
BOND CA V1
BOND CA V2
BOND CA V3
BOND CA V4
BOND CA V5
NBR_ATOM CA
NBR_RADIUS 0.01
# chirality is defined so that viewing from V1 to CA, V2-V3-V4-V5 are in clockwise
direction.
ICOOR_INTERNAL CA 0.000000 0.000000 0.000000 CA V1 V2
ICOOR_INTERNAL V1 0.000000 0.000000 2.400000 CA V1 V2
ICOOR_INTERNAL V2 0.000000 90.000000 2.400000 CA V1 V2
ICOOR_INTERNAL V3 90.000000 90.000000 2.400000 CA V1 V2
ICOOR_INTERNAL V4 180.000000 90.000000 2.400000 CA V1 V2
ICOOR_INTERNAL V5 -90.000000 90.000000 2.400000 CA V1 V2
```

Script 8. Geometric contraint (CST) file for Slam2 Match.

```
CST::BEGIN
TEMPLATE:: ATOM_MAP: 1 atom_name: CA V1 V2
TEMPLATE:: ATOM_MAP: 1 residue3: CAX
TEMPLATE:: ATOM_MAP: 2 atom_type: OOC
TEMPLATE:: ATOM_MAP: 2 residue1: D
CONSTRAINT:: distanceAB: 2.30 0.2 40.0 1 1
CONSTRAINT:: angle_A: 114.16 6.0 40.0 360. 2
CONSTRAINT:: angle_B: 130.54 5.0 40.0 360. 2
CONSTRAINT:: torsion_A: 170.72 15.0 40.0 360.0 2
CONSTRAINT:: torsion_AB: 116.86 40.0 40.0 180. 4
CONSTRAINT:: torsion_B: 175.25 10.0 40.0 180. 1
CST::END

CST::BEGIN
TEMPLATE:: ATOM_MAP: 1 atom_name: CA V1 V2
TEMPLATE:: ATOM_MAP: 1 residue3: CAX
TEMPLATE:: ATOM_MAP: 2 atom_type: OOC
TEMPLATE:: ATOM_MAP: 2 residue1: D
CONSTRAINT:: distanceAB: 2.21 0.2 40.0 1 1
CONSTRAINT:: angle_A: 31.78 6.0 40.0 360. 2
CONSTRAINT:: angle_B: 158.32 5.0 40.0 360. 2
CONSTRAINT:: torsion_A: 148.88 15.0 40.0 360.0 2
CONSTRAINT:: torsion_AB: 116.32 40.0 40.0 180. 4
CONSTRAINT:: torsion_B: 27.95 10.0 40.0 180. 1
CST::END

CST::BEGIN
TEMPLATE:: ATOM_MAP: 1 atom_name: CA V1 V2
TEMPLATE:: ATOM_MAP: 1 residue3: CAX
TEMPLATE:: ATOM_MAP: 2 atom_type: OCbb
TEMPLATE:: ATOM_MAP: 2 residue1: AVILGPFYW
CONSTRAINT:: distanceAB: 2.17 0.15 40.0 1 1
CONSTRAINT:: angle_A: 161.11 6.0 40.0 360. 2
CONSTRAINT:: angle_B: 142.97 5.0 40.0 360. 2
CONSTRAINT:: torsion_A: -12.99 15.0 40.0 360 2
CONSTRAINT:: torsion_AB: 177.94 40.0 40.0 180. 4
CONSTRAINT:: torsion_B: -70.48 10.0 40.0 180. 1
CST::END

CST::BEGIN
```

```
TEMPLATE:: ATOM_MAP: 1 atom_name: CA V1 V2
TEMPLATE:: ATOM_MAP: 1 residue3: CAX
TEMPLATE:: ATOM_MAP: 2 atom_type: OCbb
TEMPLATE:: ATOM_MAP: 2 residue1: AVILGPFYW
CONSTRAINT:: distanceAB: 2.23 0.15 40.0 1 1
CONSTRAINT:: angle_A: 92.13 6.0 40.0 360. 2
CONSTRAINT:: angle_B: 155.38 5.0 40.0 360. 2
CONSTRAINT:: torsion_A: 86.76 15.0 40.0 360 2
CONSTRAINT:: torsion_AB: -17.35 40.0 40.0 180. 4
CONSTRAINT:: torsion_B: -54.93 10.0 40.0 180. 1
CST::END
```

Script 9. Coordinate contraint (CST) file for Slam2 Design.

```
CoordinateConstraint C1  58  CA1 58   107.634 141.089   9.374  HARMONIC 0 1
CoordinateConstraint C2  58  CA1 58   106.138 141.375   9.372  HARMONIC 0 1
CoordinateConstraint O1  58  CA1 58   105.874 142.664   9.945  HARMONIC 0 1
CoordinateConstraint C3  58  CA1 58   105.383 140.319  10.164  HARMONIC 0 1
CoordinateConstraint O2  58  CA1 58   103.983 140.475   9.921  HARMONIC 0 1
CoordinateConstraint C4  58  CA1 58   105.793 138.912   9.766  HARMONIC 0 1
CoordinateConstraint O3  58  CA1 58   105.294 138.006  10.757  HARMONIC 0 1
CoordinateConstraint C5  58  CA1 58   107.307 138.765   9.674  HARMONIC 0 1
CoordinateConstraint O4  58  CA1 58   107.871 139.786   8.848  HARMONIC 0 1
CoordinateConstraint C6  58  CA1 58   103.220 140.928  11.022  HARMONIC 0 1
CoordinateConstraint C7  58  CA1 58   101.840 141.412  10.605  HARMONIC 0 1
CoordinateConstraint O5  58  CA1 58   101.968 142.423   9.602  HARMONIC 0 1
CoordinateConstraint C8  58  CA1 58   101.099 141.971  11.809  HARMONIC 0 1
CoordinateConstraint O6  58  CA1 58    99.791 142.417  11.429  HARMONIC 0 1
CoordinateConstraint C9  58  CA1 58   100.991 140.893  12.873  HARMONIC 0 1
CoordinateConstraint O7  58  CA1 58   100.382 141.448  14.048  HARMONIC 0 1
CoordinateConstraint C10 58  CA1 58   102.377 140.356  13.205  HARMONIC 0 1
CoordinateConstraint O8  58  CA1 58   103.090 139.936  12.041  HARMONIC 0 1
CoordinateConstraint C11 58  CA1 58   102.254 139.162  14.099  HARMONIC 0 1
CoordinateConstraint O9  58  CA1 58   102.547 138.058  13.592  HARMONIC 0 1
CoordinateConstraint O10 58  CA1 58   101.989 139.338  15.314  HARMONIC 0 1
```

Script 10. Bash script to call Rosetta3 Design with coordinate constraints for Slam2.

```
/mnt/c/…/rosetta/main/source/bin/enzyme_design.static.linuxgccrelease
@flags_enzdes -l output_file -cstfile DS.cst -nstruct 5
```

Script 11. Flags file for Rosetta3 Design of Slam2.

```
-extra_res_fa GAX.params                #specifies extra residue
-resfile POS.resfile                    #specifies resfile use
-enzdes:detect_design_interface         #automatic detection of designable region
-enzdes:cut1 6.0                        #cutoff value 1
-enzdes:cut2 8.0                        #cv 2
-enzdes:cut3 10.0                       #cv 3
-enzdes:cut4 12.0                       #cv 4
-enzdes:cst_opt                         #invokes catalytic ia optimization
-enzdes:bb_min                          #bb slightly felxible
-enzdes:chi_min                         #dihedrals can move
-enzdes:cst_design                      #invokes sequence design
-enzdes:design_min_cycles 3             #no. of iterations
-enzdes:cst_min                         #invoke minimization after design
-enzdes:bb_min                          #bb slightly felxible
-enzdes:chi_min                         #dihedrals can move
-packing:ex1                            #improved rotamer sampling first dihedral
-packing:ex2                            #improved rotamer sampling second dihedral
-packing:use_input_sc                   #include input rotamer in calc
-packing:linmem_ig 10                   #speeds up the sequence design step
-database …/rosetta/main/database/
```

Script 12. Parameter file for a Ca(II)-bound GlcA-Xyl ligand.

```
NAME XGC
IO_STRING XGC Z
TYPE LIGAND
AA UNK
ATOM  O1  OH    X   -0.59
ATOM  C3  CH1   X   -0.02
ATOM  C2  CH1   X   -0.02
ATOM  C1  CH1   X   -0.02
ATOM  O2  OH    X   -0.59
ATOM  C5  CH2   X   -0.11
ATOM  C4  CH1   X   -0.02
ATOM  O5  OH    X   -0.59
ATOM CA1  Ca2p  X   2.06
ATOM  O10 Oaro  X   -0.59
ATOM  C11 COO   X   0.69
ATOM  C10 CH1   X   -0.02
ATOM  C9  CH1   X   -0.02
ATOM  C8  CH1   X   -0.02
ATOM  C7  CH1   X   -0.02
ATOM  C6  CH1   X   -0.02
ATOM  O7  OH    X   -0.59
ATOM  H9  Hapo  X   0.16
ATOM  O8  OH    X   -0.59
ATOM  H15 Hpol  X   0.50
ATOM  H10 Hapo  X   0.16
ATOM  O6  OH    X   -0.59
ATOM  H12 Hpol  X   0.50
ATOM  H11 Hapo  X   0.16
ATOM  O9  OH    X   -0.59
ATOM  H16 Hpol  X   0.50
ATOM  H13 Hapo  X   0.16
ATOM  H14 Hapo  X   0.16
ATOM  O11 OOC   X   -0.69
ATOM  H4  Hapo  X   0.16
ATOM  H5  Hapo  X   0.16
ATOM  H6  Hapo  X   0.16
ATOM  O3  OH    X   -0.59
ATOM  H7  Hpol  X   0.50
ATOM  H1  Hapo  X   0.16
ATOM  O4  OH    X   -0.59
```

```
ATOM  H8  Hpol  X   0.50
ATOM  H2  Hapo  X   0.16
ATOM  H3  Hapo  X   0.16
BOND_TYPE CA1   O5  1
BOND_TYPE CA1   O10 1
BOND_TYPE CA1   O7  1
BOND_TYPE  C1   C2  1
BOND_TYPE  C1   O2  1
BOND_TYPE  C1   O3  1
BOND_TYPE  C2   C3  1
BOND_TYPE  C2   O4  1
BOND_TYPE  C3   O1  1
BOND_TYPE  C3   C4  1
BOND_TYPE  O1   C6  1
BOND_TYPE  C4   C5  1
BOND_TYPE  C4   O5  1
BOND_TYPE  C5   O2  1
BOND_TYPE  C6   C7  1
BOND_TYPE  C6   O7  1
BOND_TYPE  C7   C8  1
BOND_TYPE  C7   O8  1
BOND_TYPE  C8   O6  1
BOND_TYPE  C8   C9  1
BOND_TYPE  C9   C10 1
BOND_TYPE  C9   O9  1
BOND_TYPE  C10  O7  1
BOND_TYPE  C10  C11 1
BOND_TYPE  C11  O10 1
BOND_TYPE  C11  O11 2
BOND_TYPE  C1   H1  1
BOND_TYPE  C2   H2  1
BOND_TYPE  C3   H3  1
BOND_TYPE  C4   H4  1
BOND_TYPE  C5   H5  1
BOND_TYPE  C5   H6  1
BOND_TYPE  O3   H7  1
BOND_TYPE  O4   H8  1
BOND_TYPE  C6   H9  1
BOND_TYPE  C7   H10 1
BOND_TYPE  C8   H11 1
BOND_TYPE  O6   H12 1
BOND_TYPE  C9   H13 1
```

```
BOND_TYPE  C10  H14 1
BOND_TYPE  O8   H15 1
BOND_TYPE  O9   H16 1
CHI 1  C2   C1   O3   H7
PROTON_CHI 1 SAMPLES 3 60 -60 180 EXTRA 0
CHI 2  C3   C2   O4   H8
PROTON_CHI 2 SAMPLES 3 60 -60 180 EXTRA 0
CHI 3  C8   C7   O8   H15
PROTON_CHI 3 SAMPLES 3 60 -60 180 EXTRA 0
CHI 4  C9   C8   O6   H12
PROTON_CHI 4 SAMPLES 3 60 -60 180 EXTRA 0
CHI 5  C10  C9   O9   H16
PROTON_CHI 5 SAMPLES 3 60 -60 180 EXTRA 0
NBR_ATOM  O1
NBR_RADIUS 6.628899
ICOOR_INTERNAL    O1     0.000000    0.000000    0.000000   O1    C3    C2
ICOOR_INTERNAL    C3     0.000000  180.000000    1.430086   O1    C3    C2
ICOOR_INTERNAL    C2     0.000001   71.253666    1.520374   C3    O1    C2
ICOOR_INTERNAL    C1  -168.793630   69.092779    1.523306   C2    C3    O1
ICOOR_INTERNAL    O2    56.950355   70.456799    1.425265   C1    C2    C3
ICOOR_INTERNAL    C5   -64.286565   67.986981    1.428808   O2    C1    C2
ICOOR_INTERNAL    C4    61.741033   68.968855    1.523774   C5    O2    C1
ICOOR_INTERNAL    O5  -171.154691   70.889809    1.432504   C4    C5    O2
ICOOR_INTERNAL   CA1  -162.302819   40.243359    2.363557   O5    C4    C5
ICOOR_INTERNAL    O10 -108.410981   71.885914    2.510902  CA1    O5    C4
ICOOR_INTERNAL    C11   58.900100   38.155880    1.250984   O10  CA1    O5
ICOOR_INTERNAL    C10   -2.675219   63.653151    1.496225   C11   O10  CA1
ICOOR_INTERNAL    C9   109.819854   70.319406    1.522642   C10   C11   O10
ICOOR_INTERNAL    C8  -171.684835   70.209885    1.518686   C9    C10   C11
ICOOR_INTERNAL    C7    54.845956   70.821915    1.520414   C8    C9    C10
ICOOR_INTERNAL    C6   -58.275834   70.005560    1.521203   C7    C8    C9
ICOOR_INTERNAL    O7    58.059931   70.439030    1.427007   C6    C7    C8
ICOOR_INTERNAL    H9  -118.110438   71.333839    1.069995   C6    C7    O7
ICOOR_INTERNAL    O8  -120.768414   70.193986    1.429296   C7    C8    C6
ICOOR_INTERNAL    H15   60.533209   70.530969    0.969978   O8    C7    C8
ICOOR_INTERNAL    H10 -119.677804   71.039473    1.069994   C7    C8    O8
ICOOR_INTERNAL    O6   121.345774   70.012231    1.433447   C8    C9    C7
ICOOR_INTERNAL    H12  -60.138316   70.529466    0.969949   O6    C8    C9
ICOOR_INTERNAL    H11  119.067845   70.342395    1.070022   C8    C9    O6
ICOOR_INTERNAL    O9  -120.321960   69.878584    1.435017   C9    C10   C8
ICOOR_INTERNAL    H16   60.599024   70.530979    0.970006   O9    C9    C10
ICOOR_INTERNAL    H13 -119.732651   71.255942    1.070007   C9    C10   O9
```

```
ICOOR_INTERNAL    H14  118.139292   68.032884    1.069969   C10   C11   C9
ICOOR_INTERNAL    O11 -173.408496   55.607061    1.255969   C11   O10   C10
ICOOR_INTERNAL    H4  -121.130802   72.353633    1.070027   C4    C5    O5
ICOOR_INTERNAL    H5   120.236462   70.921224    1.070029   C5    O2    C4
ICOOR_INTERNAL    H6   120.501583   71.438843    1.070017   C5    O2    H5
ICOOR_INTERNAL    O3  -116.964376   70.511449    1.480192   C1    C2    O2
ICOOR_INTERNAL    H7    60.007680   70.526549    0.970055   O3    C1    C2
ICOOR_INTERNAL    H1  -121.531746   71.389669    1.069948   C1    C2    O3
ICOOR_INTERNAL    O4  -121.776579   71.045844    1.434410   C2    C3    C1
ICOOR_INTERNAL    H8    60.544804   70.526671    0.970006   O4    C2    C3
ICOOR_INTERNAL    H2  -119.746862   70.809171    1.069994   C2    C3    O4
ICOOR_INTERNAL    H3   118.954265   68.651855    1.069981   C3    O1    C2
```

Script 13. Python script to call Rosetta3 Match for Slam3.

```
import subprocess
import os
cwd = os.getcwd()


def runmatch(cst,suffix):
	subprocess.run(['/mnt/c/…/rosetta/main/source/bin/match.static.linuxgccreleas
e','@/mnt/c/flags_match', '-match::geometric_constraint_file', cst, '-out:suffix',
suffix, '-out:path:all 1_matched', '-mute protocols.idealize','-extra_res_fa
XGC.params'])
	return


path = '/mnt/c/…/'
for l in (1,2,3,4,5,6,7,8,9):
	for i in (1,3):
		cst = path +'lam_XGC_' + str(i) + '.cst'
		suffix = 'XGC_' + str(l) + str(i)
		runmatch(cst,suffix)
```

Script 14. Python script to call Rosetta3 Design for Slam3.

```
import numpy as np
import os
import openbabel
import shutil
import subprocess


def runenzdes(cst,name,out):
	subprocess.run(['/mnt/c/…/rosetta/main/source/bin/enzyme_design.static.linuxgccrelease', '@flags_enzdes', '-out:file:o enzdes.txt', '-nstruct 10', '-out:suffix des', '-s', name, '-cstfile', cst, '-in:auto_setup_metals', '-out:path', out ,'-extra_res_fa /mnt/c/XGC.params' ])
	return


cwd = os.getcwd()


for file in os.listdir('B2_goodmatch/'):
	list=file.split('_')
	cst='/mnt/c/…/lam_XGC_'+list[9]+'.cst'
	print(cst, file)
	runenzdes(cst,'B2_goodmatch/'+file,'B3_enzdes')
```

Script 15. Geometric constraint file for Slam3.

```
CST::BEGIN #D2808
TEMPLATE:: ATOM_MAP: 1 atom_name: CA1 O1 C3
TEMPLATE:: ATOM_MAP: 1 residue3: XGC
TEMPLATE:: ATOM_MAP: 2 atom_type: OOC
TEMPLATE:: ATOM_MAP: 2 residue1: D
CONSTRAINT:: distanceAB: 2.30 0.2 40.0 1 1
CONSTRAINT:: angle_A: 111.9  10.0 40.0 360. 2
CONSTRAINT:: angle_B:  130.5  20.0 40.0 360. 2
CONSTRAINT:: torsion_A: -72.9 10.0 40.0 360.0 2
CONSTRAINT:: torsion_AB: 63.0 20.0 40.0 180. 2
CONSTRAINT:: torsion_B: 175.3 20.0 40.0 180. 2
CST::END

CST::BEGIN #L2825
TEMPLATE:: ATOM_MAP: 1 atom_name: CA1 O1 C3
TEMPLATE:: ATOM_MAP: 1 residue3: XGC
TEMPLATE:: ATOM_MAP: 2 atom_type: OCbb
TEMPLATE:: ATOM_MAP: 2 residue1: VLI
CONSTRAINT:: distanceAB: 2.17 0.2 40.0 1 1
```

```
CONSTRAINT:: angle_A: 124.3  10.0 40.0 360. 2
CONSTRAINT:: angle_B:  143.0  20.0 40.0 360. 2
CONSTRAINT:: torsion_A: 25.9 10.0 40.0 360.0 2
CONSTRAINT:: torsion_AB: -118.5 20.0 40.0 180. 2
CONSTRAINT:: torsion_B: -70.5 20.0 40.0 180. 2
CST::END

CST::BEGIN #I2874
TEMPLATE:: ATOM_MAP: 1 atom_name: CA1 O1 C3
TEMPLATE:: ATOM_MAP: 1 residue3: XGC
TEMPLATE:: ATOM_MAP: 2 atom_type: OCbb
TEMPLATE:: ATOM_MAP: 2 residue1: VLI
CONSTRAINT:: distanceAB: 2.23 0.2 40.0 1 1
CONSTRAINT:: angle_A: 141.3  10.0 40.0 360. 2
CONSTRAINT:: angle_B:  155.4  20.0 40.0 360. 2
CONSTRAINT:: torsion_A: 178.9 10.0 40.0 360.0 2
CONSTRAINT:: torsion_AB: -13.9 20.0 40.0 180. 2
CONSTRAINT:: torsion_B: -54.9 20.0 40.0 180. 2
CST::END

CST::BEGIN #D2876
TEMPLATE:: ATOM_MAP: 1 atom_name: CA1 O1 C3
TEMPLATE:: ATOM_MAP: 1 residue3: XGC
TEMPLATE:: ATOM_MAP: 2 atom_type: OOC
TEMPLATE:: ATOM_MAP: 2 residue1: D
CONSTRAINT:: distanceAB: 2.21 0.2 40.0 1 1
CONSTRAINT:: angle_A: 69.5  10.0 40.0 360. 2
CONSTRAINT:: angle_B:  158.3  20.0 40.0 360. 2
CONSTRAINT:: torsion_A: -147.6 10.0 40.0 360.0 2
CONSTRAINT:: torsion_AB: 72.2 20.0 40.0 180. 2
CONSTRAINT:: torsion_B: 28.0 20.0 40.0 180. 2
CST::END
```

Script 16. Python script to rank the Slam3 models and test for key mutations.

```
import numpy as np
import os
import subprocess
cwd = os.getcwd()

def copy(inputFileName,outputFileName):
	with open(inputFileName, "r") as infile:
		with open(outputFileName, "w") as outfile:
			outfile.write(infile.readline())
			for line in infile:
				outfile.write(line)
	return

def dellig(inputFileName,outputFileName):
	with open(inputFileName, "r") as infile:
		with open(outputFileName, "w") as outfile:
			outfile.write(infile.readline())
			for line in infile:
				if line.startswith('ATOM') or line.startswith('TER'):
					outfile.write(line)
				if line.startswith('HETATM') and 'CA1' in line:
					ct = line[:-25]
					ct = ct[29:]
					start = 'HETATM 2714 CA    CA A 101    '
					end = '1.00  0.00          CA\nTER\n'
					newline = start + ct + end
					outfile.write(newline)
	return

def runrelax(infile,x,out):
	subprocess.run(['/mnt/c/…/rosetta/main/source/bin/relax.static.linuxgccrelease', '@/mnt/c/…/RosettaScripts/input_files/FLAGS/flags_relax', '-in:file:s', infile, '-out:suffix', '_'+x, '-nstruct 3', '-out:file:scorefile', x+'_scorefile.txt', '-no_optH    false',    '-in:auto_setup_metals',    '-relax:fast',    '-extra_res_fa /mnt/c/…/RosettaScripts/input_files/PARAMS/XGC/XGC.params',  '-out:path',  out,  '-relax:respect_resfile', '-packing:resfile', 'B4_ranked/'+x+'.resfile'])



with open('B3_enzdes.txt', 'r') as score:
	s = []
	e = []
	i = 0
```

```
	for line in score:
		if i > 0:
			list = line.split()
			full = float(list[0])
			sr4 = float(list[42])
			name = list[50]
			if sr4 < -1.0:
				if full < -145.:
					print(name,'CA:',sr4,'Score:',full)

	copy('B3_enzdes/'+name+'.pdb','B4_ranked/enzdes_rank/'+name+'.pdb')
		i = i+1


path = 'B4_ranked/enzdes_rank/'
out = 'B4_ranked/nolig_rank/'
for file in os.listdir(path):
	if file.endswith(".pdb"):
		nowfile = path + file
		outfile = out + file
		dellig(nowfile,outfile)
		x = 'relaxed'
		runrelax(outfile,x,out)


for i in ('P32S','R27TFY','salt','YF'): #'W36R','full'
	print('-------------------------------------------------------------------------
-------------------------------------------------------------------'+i+'------
-------------------------------------------------------------------------------
-----------------------------------------------------')
	path = 'B4_ranked/nolig_rank/'
	out = 'B4_ranked/'+i+'_rank/'
	for file in os.listdir(path):
		if file.endswith(".pdb"):
			nowfile = path + file
			outfile = out + file
			copy(nowfile,outfile)
			runrelax(outfile,i,out)


with open('B4_ranked/nolig_rank/out/scorefile.txt', 'r') as score:
	base = []
	names = []
	k = 0
	oldname = ''
	REU = 0.
```

```
	for line in score:
		if k > 1:
			list = line.split()
			name = list[22]
			name = name[:-2]
			full = float(list[1])
			if name == oldname:
				if full <= REU:
					REU = full
			if name != oldname:
				if REU != 0.:
					base.append(REU)
					REU = full
					names.append(name)
			oldname = name
		k = k+1
base.append(REU)
names.append(name)
print(base)
matrix = []
for i in ('P32S','R27TFY','salt','YF','W36R'):
	s = []
	k = 0
	oldname = ''
	REU = 0.
	with open('B4_ranked/'+i+'_rank/out/'+i+'_scorefile.txt', 'r') as score:
		for line in score:
			if k > 1:
				list = line.split()
				name = list[22]
				name = name[:-2]
				full = float(list[1])
				if name == oldname:
					if full <= REU:
						REU = full
				if name != oldname:
					if REU != 0.:
						s.append(REU)
						REU = full
				oldname = name
			k = k+1
	s.append(REU)
```

```
	diff =[]
	for k in np.arange(len(s)):
		d = s[k] - base[k]
		if d < 0:
			d = round(d,2)
			diff.append(d)
		else:
			diff.append(0.)
	matrix.append(diff)
matrix[4].append(0.)
data=np.column_stack((base, matrix[0], matrix[1], matrix[2], matrix[3], matrix[4]))
np.savetxt('B4_ranked/all_scorefile.txt',data, fmt='%.6f')
np.savetxt('B4_ranked/all_names.txt',names, fmt = '%s')
```

Script 17. Python script to relax the highest-ranked models.

```
from math import sqrt
import math
import subprocess
import numpy as np
import os
cwd = os.getcwd()


def runrelax(infile,x,out):
	subprocess.run(['/mnt/c/Users/Florian/Documents/rosetta/main/source/bin/relax
.static.linuxgccrelease',
'@/mnt/c/Users/Florian/Documents/RosettaScripts/input_files/FLAGS/flags_relax',   '-
in:file:s',  infile,  '-out:suffix',  '_'+x,  '-nstruct  5',  '-out:file:scorefile',
'5_scorefile.txt', '-no_optH false', '-in:auto_setup_metals', '-out:path', out])


path = 'B4_ranked/nolig_rank/best/'
out = 'B5_relaxed/'
for file in os.listdir(path):
	if file.endswith(".pdb"):
		nowfile = path + file
		x = 'relaxed'
		runrelax(nowfile,x,out)
```

Script 18. Bash script to call Rosetta3 Match for WWcalm.

```
/mnt/c/…/rosetta/main/source/bin/match.static.linuxgccrelease
@flags_match -match::geometric_constraint_file D20D22.cst;
/mnt/c/…/rosetta/main/source/bin/match.static.linuxgccrelease
@flags_match -match::geometric_constraint_file D20D24.cst;
/mnt/c/…/rosetta/main/source/bin/match.static.linuxgccrelease
@flags_match -match::geometric_constraint_file D20E31.cst;
/mnt/c/…/rosetta/main/source/bin/match.static.linuxgccrelease
@flags_match -match::geometric_constraint_file D22D24.cst;
/mnt/c/…/rosetta/main/source/bin/match.static.linuxgccrelease
@flags_match -match::geometric_constraint_file D22E31.cst;
/mnt/c/…/rosetta/main/source/bin/match.static.linuxgccrelease
@flags_match -match::geometric_constraint_file D24E31.cst;
echo All done
```

Script 19. Constraint blocks (CST) for calmodulin EF-hand 1 active site.

```
CST::BEGIN #Asp20
TEMPLATE:: ATOM_MAP: 1 atom_name: CA V1 V2
TEMPLATE:: ATOM_MAP: 1 residue3: CA
TEMPLATE:: ATOM_MAP: 2 atom_type: OOC
TEMPLATE:: ATOM_MAP: 2 residue1: D

CONSTRAINT:: distanceAB: 2.4 0.2 40.0 1 1
CONSTRAINT:: angle_A: 147.1  6.0 40.0 360. 1
CONSTRAINT:: angle_B:  0.0  5.0 40.0 360. 2
CONSTRAINT:: torsion_A: 93.4 15.0 40.0 360.0 2
CONSTRAINT:: torsion_AB: 0.2 40.0 40.0 180. 4
CONSTRAINT:: torsion_B: 28.0 20.0 40.0 180. 4
CST::END

CST::BEGIN #Asp22
TEMPLATE:: ATOM_MAP: 1 atom_name: CA V1 V2
TEMPLATE:: ATOM_MAP: 1 residue3: CA
TEMPLATE:: ATOM_MAP: 2 atom_type: OOC
TEMPLATE:: ATOM_MAP: 2 residue1: D

CONSTRAINT:: distanceAB: 2.4 0.2 40.0 1 1
CONSTRAINT:: angle_A: 145.7  6.0 40.0 360. 1
CONSTRAINT:: angle_B:  90.0  5.0 40.0 360. 2
CONSTRAINT:: torsion_A: -113.8 15.0 40.0 360.0 2
```

```
CONSTRAINT:: torsion_AB: 141.2 40.0 40.0 180. 4
CONSTRAINT:: torsion_B: -1.5 20.0 40.0 180. 4
CST::END

CST::BEGIN #Asp24
TEMPLATE:: ATOM_MAP: 1 atom_name: CA V1 V2
TEMPLATE:: ATOM_MAP: 1 residue3: CA
TEMPLATE:: ATOM_MAP: 2 atom_type: OOC
TEMPLATE:: ATOM_MAP: 2 residue1: D

CONSTRAINT:: distanceAB: 2.4 0.2 40.0 1 1
CONSTRAINT:: angle_A: 135.3  6.0 40.0 360. 1
CONSTRAINT:: angle_B:  90.0  5.0 40.0 360. 2
CONSTRAINT:: torsion_A: -107.3 15.0 40.0 360.0 2
CONSTRAINT:: torsion_AB: 149.1 40.0 40.0 180. 4
CONSTRAINT:: torsion_B: -78.2 20.0 40.0 180. 4
CST::END

CST::BEGIN #Glu31
TEMPLATE:: ATOM_MAP: 1 atom_name: CA V1 V2
TEMPLATE:: ATOM_MAP: 1 residue3: CA
TEMPLATE:: ATOM_MAP: 2 atom_type: OOC
TEMPLATE:: ATOM_MAP: 2 residue1: E

CONSTRAINT:: distanceAB: 2.4 0.2 40.0 1 1
CONSTRAINT:: angle_A: 90.7  6.0 40.0 360. 1
CONSTRAINT:: angle_B: 90.0  10.0 40.0 360. 2
CONSTRAINT:: torsion_A: -173.3 15.0 40.0 360.0 2
CONSTRAINT:: torsion_AB: -94.8 40.0 40.0 180. 4
CONSTRAINT:: torsion_B: 127.4 20.0 40.0 180. 4
CST::END
```

Script 20. Bash script for Rosetta3 Design of WWcalm.

```
/mnt/c/…/rosetta/main/source/bin/enzyme_design.static.linuxgccrelease
@flags_Donly_enzdes_ww     -l list_of_D20D22_designs.txt     -cstfile D20D22.cst;

/mnt/c/…/rosetta/main/source/bin/enzyme_design.static.linuxgccrelease
@flags_Donly_enzdes_ww     -l list_of_D20D24_designs.txt     -cstfile D20D24.cst;

/mnt/c/…/rosetta/main/source/bin/enzyme_design.static.linuxgccrelease
@flags_Donly_enzdes_ww     -l list_of_D20E31_designs.txt     -cstfile D20E31.cst;

/mnt/c/…/rosetta/main/source/bin/enzyme_design.static.linuxgccrelease
@flags_Donly_enzdes_ww     -l list_of_D22D24_designs.txt     -cstfile D22D24.cst;
```

Script 21. Bash script to call Rosetta3 Match for SO1.

```
# Call the match algorithm of Rosetta.
/PATH/Rosetta3/main/source/bin/match.static.linuxgccrelease
-lig_name CU
-geometric_constraint_file 3cqp_D.cst
-in:file:s 1ckb.pdb
-scaffold_active_site_residues sh3.pos
-extrachi_cutoff 0
-ex1
-ex2
-use_input_sc
-no_opt_H false
-enumerate_ligand_rotamers false
-dynamic_grid_refinement
-match_group SameSequenceAndDSPositionGrouper
-consolidate_matches
-output_matches_per_group 1
-out:file:output_virtual
```

Script 22. Geometric constraint file for the SOD1 (pdb: 3CQP) active site.

```
# CST constraint file for CuHHH of SOD1 (3CQP//D)
# Block 1: His-46, Nhis/ND1 binding to Cu
CST::BEGIN
TEMPLATE:: ATOM_MAP: 1 atom_name: CU V3 V4
TEMPLATE:: ATOM_MAP: 1 residue_3: CU
TEMPLATE:: ATOM_MAP: 2 atom_type: Nhis
TEMPLATE:: ATOM_MAP: 1 residue_3: HIS
CONSTRAINT:: distanceAB: 2.28 0.3 40.0 1 1
CONSTRAINT:: angle_A: 130.1 20.0 40.0 360.0 1
CONSTRAINT:: angle_B: 123.6 15.0 40.0 360.0 1
CONSTRAINT:: torsion_A: -123.9 30.0 40.0 360.0 1
CONSTRAINT:: torsion_AB: -16.5 30.0 40.0 360.0 1
CONSTRAINT:: torsion_B: 23.7 30.0 40.0 360.0 1
CST::END
# Block 2: His-48, Ntrp/NE2 binding to Cu
CST::BEGIN
TEMPLATE:: ATOM_MAP: 1 atom_name: CU V4 V2
TEMPLATE:: ATOM_MAP: 1 residue_3: CU
TEMPLATE:: ATOM_MAP: 2 atom_type: Ntrp
TEMPLATE:: ATOM_MAP: 1 residue_3: HIS
CONSTRAINT:: distanceAB: 2.00 0.3 40.0 1 1
CONSTRAINT:: angle_A: 115.4 20.0 40.0 360.0 1
CONSTRAINT:: angle_B: 108.4 15.0 40.0 360.0 1
CONSTRAINT:: torsion_A: -140.3 30.0 40.0 360.0 1
CONSTRAINT:: torsion_AB: -87.9 30.0 40.0 360.0 1
CONSTRAINT:: torsion_B: -166.0 30.0 40.0 360.0 1
CST::END
# Block 3: His-120, Ntrp/NE2 binding to Cu
CST::BEGIN
TEMPLATE:: ATOM_MAP: 1 atom_name: CU V2 V3
TEMPLATE:: ATOM_MAP: 1 residue_3: CU
TEMPLATE:: ATOM_MAP: 2 atom_type: Ntrp
TEMPLATE:: ATOM_MAP: 1 residue_3: HIS
CONSTRAINT:: distanceAB: 2.03 0.3 40.0 1 1
CONSTRAINT:: angle_A: 96.2 20.0 40.0 360.0 1
CONSTRAINT:: angle_B: 119.9 15.0 40.0 360.0 1
CONSTRAINT:: torsion_A: -131.1 30.0 40.0 360.0 1
CONSTRAINT:: torsion_AB: -121.2 30.0 40.0 360.0 1
CONSTRAINT:: torsion_B: 148.1 30.0 40.0 360.0 1
CST::END
```

Script 23. Bash script to call Rosetta3 Design for SO1.

```
# Call the design algorithm of Rosetta.
/PATH/Rosetta3/main/source/bin/enzyme_design.static.linuxgccrelease
-cst_file 3cqp_D.cst
-s $matchoutput.pdb
-enzdes:detect_design_interface
-enzdes:cut1 6.0
-enzdes:cut2 8.0
-enzdes:cut3 10.0
-enzdes:cut4 12.0
-enzdes:cst_opt
-enzdes:bb_min
-enzdes:chi_min
-enzdes:cst_design
-enzdes:design_min_cycles 3
-enzdes:lig_packer_weight 1
-enzdes:cst_min
-packing:ex1
-packing:ex2
-packing:use_input_sc
-packing:soft_rep_design
-packing:linmen_ig 10
-packing:extrachi_cutoff 0
-packing:no_opt_H false
-out:file:o scorefile_$matchnumber.txt
```

Script 24. Bash script to call Rosetta3 Relax for SO1.

```
# Call the relax algorithm of Rosetta.
/PATH/Rosetta3/main/source/bin/relax.static.linuxgccrelease
-database /PATH/Rosetta3/main/database
-in:file:s PEPTIDE.pdb
-relax:respect_resfile
-packing:resfile file.resfile
-nstruct 10
-relax:constrain_relax_to_start_coords
-relax:ramp_constraints false
-ex1
-ex2
-use_input_sc
-flip_HNQ
-no_optH false
-extrachi_cutoff 0
```

Script 25. Residue file for Rosetta3 Relax to control for redundant mutations.

```
# Resfile to control for redundant mutations.
NATAA
AUTO
EX 1 EX 2
USE_INPUT_SC
start
11     A      PIKAA  TN
16     A      PIKAA  VQ
36     A      PIKAA  LW
52     A      PIKAA  VP
53     A      PIKAA  LY
```

Script 26. Residue file for Rosetta3 Relax to mutate Asp in aspartimide-promoting sequences.

```
# Resfile to mutate aspartates.
NATAA
AUTO
EX 1 EX 2
USE_INPUT_SC
start
9      A      PIKAA  DN
14     A      PIKAA  DN
24     A      PIKAA  DNE
30     A      PIKAA  DN
41     A      PIKAA  N
```

Script 27. Bash script to call Rosetta3 Score for SO1.

```
# Call the score algorithm of Rosetta.
/PATH/Rosetta3/main/source/bin/score_jd2.static.linuxgccrelease
-in:file:l PDBlist.txt
-score::weights ref2015_ap
-ex1
-ex2
-use_input_sc
-flip
```

Script 28. Weights file for a Rosetta3 Scoring function.

```
# beta_nov15
#    beta energy function following parameter refitting (Frank DiMaio and Hahnbeom Park), November 2015
#
#   Two sets of reference weight are provided.
#       The first is for use in "minimization context" (e.g., RTmin, min_pack, or sidechain relax).
#       The second, and default set, is for use in "packing context" (e.g. Rotamer trials or packing)
#
#METHOD_WEIGHTS ref 1.82468 3.75479 -2.14574 -2.72453 1.21829 0.79816 -0.30065 2.30374 -0.71458 1.66147 2.15735 -1.34026 -1.94321 -1.45095 -0.59474 -0.28969 1.15175 2.64269 2.26099 0.58223
METHOD_WEIGHTS ref 1.32468 3.25479 -2.14574 -2.72453 1.21829 0.79816 -0.30065 2.30374 -0.71458 1.66147 1.65735 -1.34026 -1.64321 -1.45095 -0.09474 -0.28969 1.15175 2.64269 2.26099 0.58223
fa_atr 1
```

```
fa_rep 0.55
fa_sol 1.0
fa_intra_sol_xover4 1.0
lk_ball_wtd 1.0
fa_intra_rep 0.005
fa_elec 1.0
pro_close 1.25
hbond_sr_bb 1.0
hbond_lr_bb 1.0
hbond_bb_sc 1.0
hbond_sc 1.0
dslf_fa13 1.25
rama_prepro 0.45
omega 0.4
p_aa_pp 0.6
fa_dun 0.7
yhh_planarity 0.625
ref 1
aspartimide_penalty 1.0
INCLUDE_INTRA_RES_PROTEIN
NO_HB_ENV_DEP
```

8.2.2 Analysis scripts

HPLC analysis

Script 29. Python script to plot HPLC chromatograms from *Jasco* HPLC raw data.

```
#----------------------------------import---------------------------------------
-----------------------

import os

import matplotlib.pyplot as plt

import numpy as np

import pandas as pd


def plotme(x,y,lbl,name):

    plt.figure(figsize=(1.3,1.6), dpi=600)

    p=8

    plt.plot(x, y, color='black' , linewidth=0.8, label= lbl)

    #plt.plot(df2['t'], y280, color='slategrey' , markersize=2, label= pep2)


    #x axis label

    plt.xlabel('t  /  min', size = p)

    plt.xlim(10,40)

    plt.yticks(size = p)

    plt.xticks([10,25,40],size = p)

    #y axis label

    #plt.ylabel('intensity / a.u.', size=p)

    #create legend

    #plt.legend(fontsize=p,loc = 'upper left',handlelength = .8,labelspacing = 0.3)

    #save as png file

    filename = name + '_semiprep.png'

    plt.savefig(filename,bbox_inches='tight')

    plt.close()


#----------------------------------input----------------------------------------
-----------------------

cwd = os.getcwd()

for file in os.listdir(cwd):

    if file.endswith('.csv'):

        df2=pd.read_csv(file,   sep=';',   skiprows=1,   nrows=5857,   decimal=',',
names=['t', 'I'])

        plotme(df2['t'],df2['I']/np.amax(df2['I']),file.split('_')[0],file[:-4])
```

Script 30. Python script to plot HPLC chromatograms from *Shimadzu* HPLC raw data.

```
#----------------------------------import-----------------------------------------
------------------------
import os
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd



#----------------------------------figure-----------------------------------------
------------------------


def plotme(x,y,lbl,name):
    plt.figure(figsize=(1.3,1.6), dpi=600)
    p=8
    plt.plot(x, y, color='black' , linewidth=0.8, label= lbl)
    #x axis label
    plt.xlabel('t  /  min', size = p)
    plt.xlim(10,40)
    plt.yticks(size = p)
    plt.xticks([10,25,40],size = p)
    if np.amax(x) < 40.0:
        plt.xlim(0,30)
        plt.xticks([0,15,30],size = p)
    #save as png file
    filename = name + '_semiprep.png'
    plt.savefig(filename,bbox_inches='tight')
    plt.close()

#----------------------------------input-----------------------------------------
------------------------
cwd = os.getcwd()
for file in os.listdir(cwd):
    if file.endswith('.txt'):
        if '40min' in file:
            length = 5325
        elif 'SEC' in file:
            length = 3282
        else:
            length = 3820
        print(file)
        with open(file, 'rb') as myfile:
```

```
            for num, line in enumerate(myfile, 1):
                lookup1 = b'Wavelength(nm)\t280'
                lookup2 = b'Wavelength(nm)\t220'
                lookup3 = b'Wavelength(nm)\t254'
                if lookup1 in line:
                    print(file, '280')
                    df2=pd.read_csv(file,  sep='\t',  skiprows=num+2,  nrows=length,
decimal='.', names=['t', 'I'])
                    #print(df2)

plotme(df2['t'],df2['I']/np.amax(df2['I']),file.split('_')[2],file[:-4]+'_280')
                if lookup2 in line:
                    print(file, '220')
                    df1=pd.read_csv(file,  sep='\t',  skiprows=num+2,  nrows=length,
decimal='.', names=['t', 'I'])
                    #print(df1)

plotme(df1['t'],df1['I']/np.amax(df1['I']),file.split('_')[2],file[:-4]+'_220')
                if lookup3 in line:
                    print(file, '254')
                    df0=pd.read_csv(file,  sep='\t',  skiprows=num+2,  nrows=length,
decimal='.', names=['t', 'I'])
                    #print(df0)

plotme(df0['t'],df0['I']/np.amax(df0['I']),file.split('_')[2],file[:-4]+'_254')
```

Script 31. Python script to plot analytical HPLC chromatograms.

```
#----------------------------------import-----------------------------------------
------------------------

import os

import matplotlib.pyplot as plt

import numpy as np

import pandas as pd



#----------------------------------figure-----------------------------------------
------------------------


def plotme(x,y,lbl,name):

    plt.figure(figsize=(1.3,1.6), dpi=600)

    p=8

    plt.plot(x, y, color='black' , linewidth=0.8, label= lbl)

    #x axis label

    plt.xlabel('t  /  min', size = p)

    plt.ylim(-0.05,1.05)

    plt.xlim(10,40)

    plt.yticks([0,0.5,1],size = p)

    plt.xticks([10,25,40],size = p)

    if np.amax(x) < 40.0:

        plt.xlim(0,30)

        plt.xticks([0,15,30],size = p)

    #save as png file

    filename = name + '_analytical.png'

    plt.savefig(filename,bbox_inches='tight')

    plt.close()


#----------------------------------input-----------------------------------------
------------------------

limiter = '.'

cwd = os.getcwd()

for file in os.listdir(cwd):

    if file.endswith('.ctx'):

        if '280' in file:

            if 'blank' in file:

                dfb = pd.read_csv(file, sep=';', skiprows=38, nrows=6750, decimal=
limiter, names=['t', 'I','nan'], encoding = 'latin1')

            else:

                df1 = pd.read_csv(file, sep=';', skiprows=38, nrows=6750, decimal=
limiter, names=['t', 'I','nan'], encoding = 'latin1')
```

```
                plotme(df1['t'],(df1['I']-dfb['I'])/np.amax(df1['I']-
dfb['I']),file.split('_')[1],file[:-4])
        if '220' in file:
            if 'blank' in file:
                dfb = pd.read_csv(file, sep=';', skiprows=38, nrows=6750, decimal=
limiter, names=['t', 'I','nan'], encoding = 'latin1')
            else:
                df1 = pd.read_csv(file, sep=';', skiprows=38, nrows=6750, decimal=
limiter, names=['t', 'I','nan'], encoding = 'latin1')
                plotme(df1['t'],(df1['I']-dfb['I'])/np.amax(df1['I']-
dfb['I']),file.split('_')[1],file[:-4])
        else:
            if 'blank' in file:
                dfb = pd.read_csv(file, sep=';', skiprows=38, nrows=6750, decimal=
limiter, names=['t', 'I','nan'], encoding = 'latin1')
            else:
                df1 = pd.read_csv(file, sep=';', skiprows=38, nrows=6750, decimal=
limiter, names=['t', 'I','nan'], encoding = 'latin1')
                plotme(df1['t'],(df1['I']-dfb['I'])/np.amax(df1['I']-
dfb['I']),file.split('_')[1],file[:-4])
```

Script 32. Python script to plot MALDI-TOF mass spectra.

```
#----------------------------------import-------------------------------------------
------------------------

import os

import sys

import numpy as np

import matplotlib as mpl

import matplotlib.pylab as plt

import pandas as pd

from scipy.signal import find_peaks

from numpy import sin, cos, tan, arcsin, arcsin, arctan, arctan2, sinh, cosh, tanh,
exp, log, log10, sqrt, nan, pi, array, arange, linspace, logspace

from matplotlib.pylab import plot, show


#-----Define Functions-----------------------------------------------


def plotme(num,xdata,ydata):
       plt.figure(figsize=(8,5), dpi=100)
       plt.plot(xdata, ydata, color='black', linewidth=1 , label= num)
       ymax = np.amax(ydata)
#      peaks, _ = find_peaks(ydata, threshold = 150)
#      print(xdata[peaks], ydata[peaks])
#      plt.plot(xdata[peaks], ydata[peaks], 'xr')
       #
       plt.xticks(fontsize=10)
       plt.yticks(fontsize=10)
       #x axis label
       plt.xlabel('m/z', size=10)

       #y axis label
       plt.ylabel('intensity [a.u.]', size=10)
       plt.xlim([np.min(df1['mass']),np.max(df1['mass'])])
       plt.ylim([0.,ymax+ymax/5])
       #create legend
       plt.legend(fontsize = 10, loc = 'upper left')
       #save as png file
       figname2 = 'plot_all/'+num+'_maldi.png'
       plt.savefig( figname2, dpi=300,bbox_inches='tight')
       plt.close()

def plotzoom(num,xdata,ydata,xmax,plusminus):
       limup=100*round(xmax/100, 0)+plusminus
```

```
	limdown=limup-2*plusminus
	plt.figure(figsize=(3,3), dpi=100)
	plt.plot(xdata, ydata, color='black', linewidth=2 , label=r"$\bf{" + str(np.round(xmax,2)) + "}$")
	ymax = np.amax(ydata)
#	peaks, _ = find_peaks(ydata, threshold = 150)
#	print(xdata[peaks], ydata[peaks])
#	plt.plot(xdata[peaks], ydata[peaks], 'xr')
	#
	plt.xticks(fontsize=10)
	plt.yticks(fontsize=10)
	#x axis label
	plt.xlabel('m/z', size=10)
	plt.legend(fontsize = 12, loc = 'upper center', frameon = False, handlelength = 0)
	#y axis label
	plt.ylabel('intensity [a.u.]', size=10)
	plt.xlim([limdown,limup])#[np.min(xdata),np.max(xdata)])
	plt.ylim([0.,ymax+ymax/5])
	#save as png file
	figname2 = 'plot_all/'+num+'_zoom_maldi.png'
	plt.savefig( figname2, dpi=300,bbox_inches='tight')
	plt.close()



#-----Input-----------------------------------------------------
#num = 'S19' #name of the sample
cwd = os.getcwd() #defines the working directory path

datx = [] #create x data dictionary
daty = [] #create y data dictionary
for file in os.listdir(cwd): #all files in current working directory
	if file.endswith('.txt'): #only txt files
		#if '29621' in file: #only autXXXX number
		df1 = pd.read_csv(file, sep=' ', decimal='.', names=['mass','int']) #read data with pandas
		datx = df1['mass'] #define x data
		daty = df1['int'] #define y data
		xmax=df1['mass'][df1['int'].idxmax()] #find mass of highest intensity
		plotme(file[:-4],datx,daty) #plot full mass spectrum of sample, x data and y data
		plotzoom(file[:-4],datx,daty,xmax,200) #plot zoom spectrum of sample, x data, y data, x of maximum y, range around maximum y
```

Binding analysis

Script 33. Python script to analyze Tb(III) FRET titration data.

```
import os
import sys
import numpy as np
import matplotlib as mpl
import matplotlib.pylab as plt
import pandas as pd

from scipy.optimize import curve_fit

from numpy import sin, cos, tan, arcsin, arcsin, arctan, arctan2, sinh, cosh, tanh,
exp, log, log10, sqrt, nan, pi, array, arange, linspace, logspace

from matplotlib.pylab import plot, show

#-----Define Functions-----------------------------------------------
def R_square(x,y,f,a,b): # x,y: Daten ;  f: Funktion ; a,b: Parameter der Funktion

	res_square=[]
	for i in arange(len(y)):
		y1=(y[i]-f(x[i],a,b))**2
		res_square.append(y1)

	SS_res=sum(res_square)



	y_ymean_square=[]
	y_mean=np.mean(y)
	for i in arange(len(y)):
		y2=(y[i]-y_mean)**2
		y_ymean_square.append(y2)

	SS_tot=sum(y_ymean_square)



	R_2 = 1 - (SS_res/SS_tot)
	return R_2

def R_square3(x,y,f,a,b,c): # x,y: Daten ;  f: Funktion ; a,b: Parameter der Funktion

	res_square=[]
```

```
	for i in arange(len(y)):
		y1=(y[i]-f(x[i],a,b,c))**2
		res_square.append(y1)

	SS_res=sum(res_square)



	y_ymean_square=[]
	y_mean=np.mean(y)
	for i in arange(len(y)):
		y2=(y[i]-y_mean)**2
		y_ymean_square.append(y2)

	SS_tot=sum(y_ymean_square)
	R_2 = 1 - (SS_res/SS_tot)
	return R_2

def Hypnew(x,Kd,Bmax):
	return Bmax*x/(Kd+x)

def Hyp(A,B,Kd,S):
	a=Kd+A+B
	return S*((a-sqrt((a**2)-(4*A*B)))/(2*B))
#-----get Y-----------------------------------------------------
cwd = os.getcwd()

daty = []

for file in os.listdir(cwd):
	if file.endswith((".txt")):
		if "blank" in file:
			df1 = pd.read_csv(file, sep='\t', skiprows=509, nrows=5,
decimal=',', usecols=[1], names=['Y'],encoding='latin-1')

	by=(df1['Y'][0]+df1['Y'][1]+df1['Y'][2]+df1['Y'][3]+df1['Y'][4])/5
		else:
			df2 = pd.read_csv(file, sep='\t', skiprows=509, nrows=5,
decimal=',', usecols=[1], names=['Y'],encoding='latin-1')

	sy=(df2['Y'][0]+df2['Y'][1]+df2['Y'][2]+df2['Y'][3]+df2['Y'][4])/5-by
			daty.append(sy)
print(daty)
daty = 1-daty/daty[0]
```

```
#-----get X-----------------------------------------------------
datx = [0.0,31.25,62.5,125.,250.,500.,1000.,2000.]

#-----Fit curve-------------------------------------------------------
A_opt,A_cov=curve_fit(Hypnew,datx,daty, maxfev=1000000)

K_A=A_opt[0]
S1_A=A_opt[1]
print(K_A)
Err_K_A,Err_S1_A=sqrt(np.diag(A_cov))
#-----Set up plot-----------------------------------------------------
x_fit=arange(0.0,2005.,5.)

y_fit=[]
for i in arange(len(x_fit)):
       A=x_fit[i]
       S=Hypnew(A,K_A,S1_A)
       y_fit.append(S)
#-----Plot curve-----------------------------------------------------
p = 8
plt.figure()
plt.figure(figsize=(2.2,1.6), dpi=600)
plt.plot(datx,(daty-np.amin(daty))/(np.amax(daty)-np.amin(daty)),               'o',
color='slategrey', markersize=1, label='calmcage + $ \mathrm{Tb^{3\!+}} $' )
plt.plot(x_fit,(y_fit-np.amin(daty))/(np.amax(daty)-np.amin(daty)),color='black',
linewidth=1,   label   =   'Kd   =   '+str(int(round(K_A,0)))+   '   $   \pm   $   '
+str(int(round(Err_K_A,0)))+ ' $ \mathrm{\mu} $M ')
#x axis label
plt.xlabel('$ \mathrm{c_L} $  /  $ \mathrm{\mu} $M ', size = p)
plt.ylabel('fraction bound', size=p)
plt.xticks(fontsize = p)
#y axis label
plt.yticks(fontsize = p)
#create legend
plt.legend(fontsize=p,  loc='lower  right',  borderpad  =  0.3,  handlelength  =
0.2,labelspacing = 0.2, handletextpad = 0.2)#save as png file
plt.savefig('titration.png', bbox_inches='tight')
```

Script 34. DynaFit script to analyze competitive dye binding data.

```
Competitive fluorescence displacement assay to determine the Kd for both labeled and
unlabeled ligand.

;_____________________________________________________________________________
[task]
   data = equilibria
   task = fit
[mechanism]
   ;
   ; P = protein
   ; L = ligand
   ; M = metal
   ;
   P + M <==> PM      :    Kd*   dissoc
   L + M <==> ML      :    Kd1   dissoc

[constants]
   Kd* = 20. ?
   Kd1 = 1.7

[responses]

   intensive

[data]
   variable  M, L, P
   set Abs.01 | resp L = 0.0002 ?, ML = 0.041 ?

[output]
   directory ./out
   rate-file outfile
[set:Abs.01]

M,microM     L,microM     P,microM     Abs
0.0000000000 10.0000000000 10.0000000000 0.2660000000
2.4968789014 9.9875156055 9.9875156055 0.2310000000
4.9875311721 9.9750623441 9.9750623441 0.2040000000
7.4719800747 9.9626400996 9.9626400996 0.1830000000
9.9502487562 9.9502487562 9.9502487562 0.1660000000
12.4223602484 9.9378881988 9.9378881988 0.1550000000
14.8883374690 9.9255583127 9.9255583127 0.1480000000
17.3482032218 9.9132589839 9.9132589839 0.1430000000
```

```
19.8019801980 9.9009900990 9.9009900990 0.1390000000
22.2496909765 9.8887515451 9.8887515451 0.1360000000
24.6913580247 9.8765432099 9.8765432099 0.1360000000
29.5566502463 9.8522167488 9.8522167488 0.1310000000

[end]
```

Script 35. Python script to simulate a competitive dye binding experiment.

```
"""Fitting example, determining Kd from 1:1:1 competition data"""
import matplotlib.pylab as plt
import numpy as np
import pybindingcurve as pbc
import sys
import pandas as pd
import os
KIT=(0.11,0.59,0.51)
BLU='xkcd:azure'
HU=(0.78,0.09,0.15)
#viridis colors
Ni = '#fde725' #yellow
Fe = '#7ad151' #leaf
Cu = '#22a884' #green
Co = '#2a788e' #teal
Zn = '#414487' #violet
Mn = '#440154' #purple
#_______________Define
Functions_____________________________________________________________________#
def plotme(num,xdata,ydata,c_n, A_plot,K_A,Err_K_A,R_A):
      plt.figure(figsize=(3,3), dpi=300)
      plt.plot(c_n,A_plot, color='red')
      plt.plot(xdata,   ydata,'.',   color='black'   ,   markersize=5,   label   =
'$K_{\mathrm{d}}=%i         \pm         %i$         $\mu$M,         $R^2         =
%1.3f$'%(round(K_A,0),round(Err_K_A,0),round(R_A,3)))
#      plt.errorbar(xdata,ydata,yerr=erry, fmt=' ', ecolor= 'slategrey', capsize = 5)
      #
      plt.xticks(fontsize=10)
      plt.yticks(fontsize=10)
      #x axis label
      plt.xlabel('$c[\mathrm{L}]$ / $\mu$M', size=10)

      #y axis label
      plt.ylabel('Absorbance [a.u.]', size=10)

      #create legend
      plt.legend(fontsize = 8, loc = 'lower right')
      plt.xlim([np.min(c_n),np.max(c_n)])
      #save as png file
      figname1 = 'out/'+num+'_competitive_assay.png'
      figname2 = 'out/'+num+'_400_competitive_assay.png'
```

```
	plt.savefig( figname1, dpi=300,bbox_inches='tight')
	plt.xlim([0,400])
	plt.savefig( figname2, dpi=300,bbox_inches='tight')


def R_square(A,Acalc):
	Amax = np.max(A)
	Amin = np.min(A)
	Amean = np.mean(A)
	rsd = 0.
	for i in np.arange(len(A)):
		rsd = rsd + (A[i]-Acalc[i])**2
	msd = 0.
	for i in np.arange(len(A)):
		msd = msd + (A[i]-Amean)**2
	R_2 = 1-rsd/msd
	return R_2


def fit_pbc(xcoords,ycoords,cqel,cpep,Kq):
	# Construct the PyBindingCurve object, operating on a 1:1:1 (compeittion) system and add experimental data to the plot
	my_system = pbc.BindingCurve("1:1:1")
	my_system.add_scatter(xcoords, ycoords)
	# 'p' for particle to bind
	# 'l' for labeled molecule
	# 'i' for molecule of interest

	# Known system parameters, kdpi will be added to this by fitting
	system_parameters = {
		"p": xcoords,
		"l": cqel,
		"i": cpep,
		"kdpl": Kq,
		"ymin": 0,
	}

	# Now we call fit, passing the known parameters, followed by a dict of parameters to be fitted along
	# with an initial guess, pass the ycoords, and what the readout (ycoords) is
	fitted_system, fit_accuracy = my_system.fit(
		system_parameters, {"kdpi": 20, "ymax": np.min(ycoords)}, ycoords
	)
	# Print out the fitted parameters
```

```
for k, v in fit_accuracy.items():
	print(f"Fit: {k}={fitted_system[k]} +/- {v}")
print('Fit successful')
# Assign more points to 'p' to make a smooth plot
fitted_system["p"] = np.linspace(0, np.max(xcoords))

# Add a new curve, simulated using fitted parameters to our BindingCurve object
my_system.add_curve(fitted_system)

# Get Fit:
c_n=np.arange(0.,2000.,1)

my_parameters = {
	"p": c_n,
	"l": cqel,
	"i": cpep,
	"kdpl": Kq,
	"kdpi": fitted_system['kdpi'],
	"ymin": fitted_system['ymin'],#np.min(ycoords),
	"ymax": fitted_system['ymax']
}

test_parameters = {
	"p": c_n,
	"l": cqel,
	"i": cpep,
	"kdpl": Kq,
	"kdpi": 2,
	"ymin": fitted_system['ymin'],#np.min(ycoords),
	"ymax": fitted_system['ymax']
}
B_plot = (my_system.query(test_parameters))

test_parameters = {
	"p": c_n,
	"l": cqel,
	"i": cpep,
	"kdpl": Kq,
	"kdpi": 200,
	"ymin": fitted_system['ymin'],#np.min(ycoords),
	"ymax": fitted_system['ymax']
}
```

```
	C_plot = (my_system.query(test_parameters))


	R_parameters = {
		"p": xcoords,
		"l": cqel,
		"i": cpep,
		"kdpl": Kq,
		"kdpi": fitted_system['kdpi'],
		"ymin": np.min(ycoords),
		"ymax": fitted_system['ymax']
	}
	R_plot = (my_system.query(R_parameters))
	R_A = R_square(ycoords,R_plot)
	A_plot = (my_system.query(my_parameters))


	return xcoords, ycoords, c_n, A_plot, fitted_system['kdpi'],
fit_accuracy['kdpi'], R_A, B_plot, C_plot


def get_pbc(c_n,cqel,cpep,Kq,Kp):
	# Construct the PyBindingCurve object, operating on a 1:1:1 (compeittion)
system and add experimental data to the plot
	my_system = pbc.BindingCurve("1:1:1")
	# 'p' for particle to bind
	# 'l' for labeled molecule
	# 'i' for molecule of interest
	# Get Fit:


	test_parameters = {
		"p": c_n,
		"l": cqel,
		"i": cpep,
		"kdpl": Kq,
		"kdpi": Kp,
		"ymin": 0,
		"ymax": 1
	}
	B_plot = (my_system.query(test_parameters))
	return B_plot


def readone(name):
	df1=pd.read_csv(name+'_data.txt', sep=' ', decimal='.',encoding = 'latin1',
names=['met', 'lig', 'pep','dati'])
	return df1['met'],df1['dati']#,df1['yr']
```

```
#_________________Get
File_____________________________________________________________________________#
# We can choose to work in a common unit, typically nM, or uM, as long as all
# numbers are in the same unit, the result is valid.  We assume uM for all
# concentrations bellow.

cwd = os.getcwd()

#for file in os.listdir(cwd):
       #if file.endswith('_titration.txt'):
              #fnum = file[:-14]

#name = fnum + '_titration.txt'



##_________________Get
Data_____________________________________________________________________________#

cqel = 28.52
cpep = 28.52
Kq = 1.4#uM

c_n = [0,4.5,9,13.5,18,22.5,27.,36.,45,54,63,72,80,89,107,124,142,159,176]
c_n4 = np.arange(0.,180.,1.)
y1 = get_pbc(c_n,cqel,cpep,Kq,0.1)
y2 = get_pbc(c_n,cqel,cpep,Kq,10)
y3 = get_pbc(c_n,cqel,cpep,Kq,100)
y4 = get_pbc(c_n4,cqel,cpep,Kq,100000)

p = 8
w = 1
m = 0.5
plt.figure(figsize=(1.5,1.6), dpi=600)
plt.plot(c_n4, y4, '-', color='slategrey', linewidth=w, label = '$\infty$')
plt.plot(c_n, y1, 'o', color=Fe, markersize=m, label = '100 nM')
plt.plot(c_n, y2, 'o', color=Co, markersize=m, label = '10 $\mathrm{\mu}$M')
plt.plot(c_n, y3, 'o', color=Zn, markersize=m, label = '100 $\mathrm{\mu}$M')

#x axis label
plt.xlabel('c / $\mathrm{\mu}$M', size = p)
plt.xlim([0,180])
```

```
plt.ylim([0,1.1])

#create legend

plt.legend(fontsize=p, loc = 'lower right', borderpad = 0.3, handlelength =
0.5,labelspacing = 0.2, handletextpad = 0.5)

plt.yticks([0,0.5,1.0],fontsize=p)

plt.xticks([0,60,120,180],fontsize=p)

#save as png file

plt.savefig('out/b5nsimKdatP20.png', bbox_inches='tight')

#y axis label

plt.ylabel('fraction bound', size=p)

#save as png file

plt.savefig('out/b5nsimKdatP20-lbl.png', bbox_inches='tight')

plt.close()


x1,y1=readone('FH30-32_B2B_ca')

x2,y2=readone('FH30-32_B2B_s08')


p = 8

w = 1

m = 0.5


plt.figure(figsize=(1.5,1.6), dpi=600)

#plt.errorbar(x1,y1/np.amax(y1),yerr=r1/np.amax(y1), fmt=' ', color=Cu, linewidth =
m, capsize = m)

plt.plot(c_n4, y4, '-', color='slategrey', linewidth=w, label = '$K_{\mathrm{d}}$ =
$\infty$')

plt.plot(x1,   1-(y1-np.amin(y1))/(np.amax(y1)-np.amin(y1)),   'o',   color='black',
markersize=m, label = 'B2B')

#plt.plot(x2,    1-(y2-np.amin(y2))/(np.amax(y2)-np.amin(y2)),    'o',    color=Cu,
markersize=m, label = 'B2B vs. Slam2')

#x axis label

plt.xlabel('c / $\mathrm{\mu}$M', size = p)

plt.xlim([0,180])

plt.ylim([0,1.1])

#create legend

plt.legend(fontsize=p, loc = 'lower right', borderpad = 0.3, handlelength =
0.5,labelspacing = 0.2, handletextpad = 0.5)

plt.yticks([0,0.5,1.0],fontsize=p)

plt.xticks([0,60,120,180],fontsize=p)

#save as png file

plt.savefig('out/none_vs_b2b.png', bbox_inches='tight')

#y axis label

plt.ylabel('fraction bound', size=p)

#save as png file
```

```
plt.savefig('out/none_vs_b2b-lbl.png', bbox_inches='tight')

plt.close()


plt.figure(figsize=(1.5,1.6), dpi=600)

#plt.errorbar(x1,y1/np.amax(y1),yerr=r1/np.amax(y1), fmt=' ', color=Cu, linewidth =
m, capsize = m)

plt.plot(c_n4, y4, '-', color='slategrey', linewidth=w, label = '$K_{\mathrm{d}}$ =
$\infty$')

plt.plot(x2,     1-(y2-np.amin(y2))/(np.amax(y2)-np.amin(y2)),     'o',     color=Cu,
markersize=m, label = 'Slam2')

#x axis label

plt.xlabel('c / $\mathrm{\mu}$M', size = p)

plt.xlim([0,180])

plt.ylim([0,1.1])

#create legend

plt.legend(fontsize=p,  loc  =  'lower  right',  borderpad  =  0.3,  handlelength  =
0.5,labelspacing = 0.2, handletextpad = 0.5)

plt.yticks([0,0.5,1.0],fontsize=p)

plt.xticks([0,60,120,180],fontsize=p)

#save as png file

plt.savefig('out/slam2_vs_b2b.png', bbox_inches='tight')

#y axis label

plt.ylabel('fraction bound', size=p)

#save as png file

plt.savefig('out/slam2_vs_b2b-lbl.png', bbox_inches='tight')

plt.close()
```

Script 36. DynaFit script to analyze a competitive ligand binding experiment.

```
Competitive fluorescence displacement assay to determine the Kd for both labeled and
unlabeled ligand.

;_____________________________________________________________________________
[task]
   data = equilibria
   task = fit
[mechanism]
   ;
   ; P = protein
   ; L = ligand
   ; M = metal
   ;
   P + M <==> PM      :    Kd*   dissoc
   P + L <==> PL      :    Kd1   dissoc

[constants]
   Kd* = 1000 ?
   Kd1 = 0.203

[responses]

   intensive

[data]
   variable  M, L, P
   set Abs.01 | resp L = 0.0002 ?, PL = 0.041 ?

[output]
   directory ./out
   rate-file outfile
[set:Abs.01]

M,microM     L,microM     P,microM     Abs
0.0000000000 0.2000000000 0.0100000000 1.0000000000
37.5000000000 0.2000000000 0.0100000000 0.9108922063
75.0000000000 0.2000000000 0.0100000000 0.7917085363
150.0000000000 0.2000000000 0.0100000000 0.7403664373
300.0000000000 0.2000000000 0.0100000000 0.6495426985

[end]
```

Script 37. Python script to analyze a Superoxide Dismutase activity assay by endpoint determination.

```
#----------------------------------import-----------------------------------------
------------------------

import os

import matplotlib.pyplot as plt

import numpy as np

import pandas as pd

from scipy.optimize import curve_fit

KIT=(0.11,0.59,0.51)

BLU='xkcd:azure'

HU=(0.78,0.09,0.15)

#Mn(II): clr = 'magenta'

#Fe(II): clr = 'coral'

#Co(II): clr = 'steelblue'

#Ni(II): clr = 'goldenrod'

#Cu(II): clr = HU

#Zn(II): clr = KIT


#viridis colors

Ni = '#fde725' #yellow

Fe = '#7ad151' #leaf

Cu = '#22a884' #green

Co = '#2a788e' #teal

Zn = '#414487' #violet

Mn = '#440154' #purple


def func(x,a,b):

       return a*x+b


def getdata(path):

    df1 = pd.read_excel(path, sheet_name = 'Summary', names = ('time','abs') ,
skiprows = 52, decimal = '.')

    #print(df1)

    A = df1['abs'].iloc[80:90]

    #print(path,A)

    #print(df1.iloc[80:90])

    meany=A.mean(axis=0)

    erry=A.std(axis=0)

    return meany,erry


def getpoint(n,nr,path):
```

```
    s,r = getdata(path)
    L = n/s
    #rL = ((-n/(s**2))*nr**2+r**2)**0.5
    I = 100*(n-s)/n
    #rI = (1/n*(nr)**2+r**2)**0.5
    return L, I#rL, I, rI

def getmeans(a,b,c):
    L, I = np.column_stack((a,b,c))
    mL=L.mean()
    rL=L.std()
    mI=I.mean()
    rI=I.std()
    return mL,rL, mI,rI

#-------------------------------------input-----------------------------------------
-------------------------
#path = '/mnt/c/Users/Florian/Downloads/DynaFit4/out/txt/'
#filenames

#FH302 find IC50
n1, rn1 = getdata('FH302_None.xlsx')
s1 = getpoint(n1,rn1,'FH302_SO1-2.5nM_CU-2.4nM.xlsx')
s2 = getpoint(n1,rn1,'FH302_SO1-10nM_CU-9.6nM.xlsx')
s3 = getpoint(n1,rn1,'FH302_SO1-20nM_CU-19.6nM.xlsx')
s4 = getpoint(n1,rn1,'FH302_SO1-50nM_CU-48nM.xlsx')
s5 = getpoint(n1,rn1,'FH302_SO1-100nM_CU-96nM.xlsx')
n2, rn2 = getdata('FH303_None.xlsx')
n3, rn3 = getdata('FH304_None.xlsx')
#print(s1,s2,s3,s4,s5)
datL, datI = np.column_stack((s1,s2,s3,s4,s5))
#print(datI)
datp = [2.5,10,20,50,100]#nM
#FH303 IC50
scu1 = getpoint(n2,rn2,'FH303_Cu-SO1-1.xlsx')
s1 = getpoint(n2,rn2,'FH303_SO1_noMetal.xlsx')
cutot1 = getpoint(n2,rn2,'FH303_Cu-30.7.xlsx')
cuf1 = getpoint(n2,rn2,'FH303_freeCu-1.xlsx')
#FH304 repeat
scu2 = getpoint(n3,rn3,'FH304_Cu-SO1-2.xlsx')
scu3 = getpoint(n3,rn3,'FH304_Cu-SO1-3.xlsx')
sod1 = getpoint(n3,rn3,'FH304_SOD-1.xlsx')
```

```
sod2 = getpoint(n3,rn3,'FH304_SOD-2.xlsx')

sod3 = getpoint(n3,rn3,'FH304_SOD-3.xlsx')

#FH305 repeat free Copper

n4, rn4 = getdata('FH305_None.xlsx')

cutot2 = getpoint(n2,rn2,'FH305_Cutot-2.xlsx')

cutot3 = getpoint(n2,rn2,'FH305_Cutot-3.xlsx')

cuf2 = getpoint(n2,rn2,'FH305_freeCu-2.xlsx')

cuf3 = getpoint(n2,rn2,'FH305_freeCu-3.xlsx')

#get mean data

scu = getmeans(scu1,scu2,scu3)

sod = getmeans(sod1,sod2,sod3)

print(scu)

#s = getmeans(s1,s1,s1)

cutot = getmeans(cutot1,cutot2,cutot3)

cuf = getmeans(cuf1,cuf2,cuf3)

data=np.column_stack((['linear
rate','error','inhibition','error'],sod,scu,cutot,cuf))

np.savetxt('out/activity_data.txt',data, delimiter=" ", fmt="%s")

#--------------------------------------fit-----------------------------------------
-------------------------

popt1, pcov1 = curve_fit(func, datp, datL, maxfev=1000000, p0 = (.0312,1.0), bounds
= ((0.0310,0),(0.0314,2)))

A = popt1[0]

B = popt1[1]

Ar,Br=np.sqrt(np.diag(pcov1))


print(str(A)+' +/- '+str(Ar), str(B)+' +/- '+str(Br))

print(1/A)

dati = []

num = np.arange(-5.,110,1.)

for i in np.arange(len(num)):

    n = func(num[i],A,B)

    dati.append(n)

##---------------------------------figure---------------------------------------
-------------------------

p = 10

cap=3

plt.figure()

plt.figure(figsize=(1.2,1.8), dpi=600)#5,4

plt.plot(num,dati, color='slategrey', linewidth=2)

plt.plot(datp,datL,    'o',    color='black',    markersize=cap)#,    label='SO1+$
\mathrm{Cu^{2\!+}} $' )

plt.errorbar(32,scu[0],yerr=scu[1],  fmt='o',  color=Cu,  capsize  =  cap,  label='$
\mathrm{IC_{50}} $') #, label='$ \mathrm{IC_{50}} $ = 32 nM' )
```

```
#x axis label
plt.xlabel('$ \mathrm{c_{SO1}}$  /  nM', fontsize=p)
plt.xticks([0,32,100],fontsize=p)
#y axis label
plt.ylabel('$ \mathrm{A^{0}}$  /  $ \mathrm{A^{i}}$', fontsize=p)
plt.yticks([1,2,3,4,5],fontsize=p)
#create legend
plt.legend(fontsize=p, handlelength=0.4,handletextpad = 0.5)
##save as png file
plt.xlim([-5,105])
plt.ylim([.8,5])
filename = 'out/lineplot_activity.png'
plt.savefig(filename, bbox_inches='tight')#, transparent=True)


#bar diagram
plt.figure()
plt.figure(figsize=(1.4,1.8), dpi=600)
plt.bar('Cu/Zn-  \nSOD1',  sod[2],  yerr=sod[3],color  ='slategrey',edgecolor  =
'black',width = 0.4, capsize = cap)
plt.bar('Cu- \nSO1', scu[2], yerr=scu[3],color =Cu,edgecolor = 'black',width = 0.4,
capsize = cap)
#plt.bar('30.7  nM\n$  \mathrm{Cu^{2\!+}}  $',  cutot[2],  yerr=cutot[3],  color  =
Fe,edgecolor = 'black',width = 0.4, capsize = 3)
plt.bar('free \n$ \mathrm{Cu^{2\!+}} $', cuf[2], yerr=cuf[3],color = Fe, edgecolor =
'black',width = 0.4, capsize = 3)
plt.yticks([0,25,50],fontsize=p)
plt.xticks(fontsize=p)
plt.ylabel('Inhibition / %', size = p)
plt.ylim([0,60])
plt.savefig('out/so1vsbarplot.png', dpi=600,bbox_inches='tight')
```

Script 38. Python script to analyze a Superoxide Dismutase activity assay by rate determination.

```
#----------------------------------import-----------------------------------------
------------------------
import os
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
from scipy.optimize import curve_fit
from scipy.stats import linregress
KIT=(0.11,0.59,0.51)
BLU='xkcd:azure'
HU=(0.78,0.09,0.15)
#Mn(II): clr = 'magenta'
#Fe(II): clr = 'coral'
#Co(II): clr = 'steelblue'
#Ni(II): clr = 'goldenrod'
#Cu(II): clr = HU
#Zn(II): clr = KIT

#viridis colors
Ni = '#fde725' #yellow
Fe = '#7ad151' #leaf
Cu = '#22a884' #green
Co = '#2a788e' #teal
Zn = '#414487' #violet
Mn = '#440154' #purple

def func(x,a,b):
       return a*x+b

def getrate(a,b):
    res = linregress(a,b)
    return res.slope

def getdata(table,n):
    if n==3:
        df1 = pd.read_excel('activities.xlsx', sheet_name = table, names =
('time','abs1','abs2','abs3') , decimal = '.')
        print(df1)
        #r1 = getrate(df1['time'],df1['abs1']-df1['abs1'][0])
        r2 = getrate(df1['time'],df1['abs2']-df1['abs2'][0])
        r3 = getrate(df1['time'],df1['abs3']-df1['abs3'][0])
```

```
        dati = np.array([df1['abs2']-df1['abs2'][0],df1['abs3']-df1['abs3'][0]])
        A = np.array([r2,r3])
    elif n==2:
        df1 = pd.read_excel('activities.xlsx', sheet_name = table, names =
('time','abs1','abs2') , decimal = '.')
        r1 = getrate(df1['time'],df1['abs1']-df1['abs1'][0])
        r2 = getrate(df1['time'],df1['abs2']-df1['abs2'][0])
        dati = np.array([df1['abs1']-df1['abs1'][0],df1['abs2']-df1['abs2'][0]])
        A = np.array([r1,r2])
    else:
        print('Something went wrong for table:  ',table)
    meany=dati.mean(axis=0)
    erry=dati.std(axis=0)
    r = A.mean(axis=0)
    rr = A.std(axis=0)
    return r,rr,meany,erry,df1['time']


def getpoint(n,nr,s,sr):
    I = 100*(n-s)/n
    rI = 100*((-sr/n)**2+((s*nr)/n**2)**2)**0.5
    return I, rI


#-------------------------------------input-----------------------------------------
--------------------------
o,ro,dato,rdato,x = getdata('Blank',3)
s,rs,dats,rdats,x = getdata('CU-SO1',3)
a,ra,data,rdata,x = getdata('apo-SO1',2)
print(s,rs,o,ro)
sI,rsI = getpoint(o,ro,s,rs)
aI,raI = getpoint(o,ro,a,ra)
kxtt = 8.59*10**4 #cxtt = 100 uM
k = (100*kxtt)/0.14 #kP*IC50 = kQ*cQ
rk = ((100/0.14)*0.81*10**4)
print('Cu-SO1: '+ str(int(round(sI,0)))+' +/- '+str(int(round(rsI,0))),'apo-SO1: '+
str(int(round(aI,0)))+' +/- '+str(int(round(raI,0))))
print('{:.2e}'.format(k),'{:.2e}'.format(rk))
dati=np.column_stack((['none',o,ro],['sample',s,rs],['apo',a,ra]))
np.savetxt('out/activity_rates.txt',dati, delimiter=" ", fmt="%s")
##-------------------------------------figure-----------------------------------------
-------------------------
p = 10
m = 4
cap = 3.0
```

```
plt.figure()
plt.figure(figsize=(3.34,1.9), dpi=600)#5,4


plt.plot(x,o*x*0.1, color='slategrey', linewidth=2)
plt.errorbar(x,dato*0.1,yerr=rdato*0.1, fmt='o', color='slategrey', markersize = m, capsize = cap, label='No inhibitor')
plt.plot(x,a*x*0.1, color=Fe, linewidth=2)
plt.errorbar(x,data*0.1,yerr=rdata*0.1, fmt='o', color=Fe, markersize = m, capsize = cap, label='SO1')
plt.plot(x,s*x*0.1, color=Cu, linewidth=2)
plt.errorbar(x,dats*0.1,yerr=rdats*0.1, fmt='o', color=Cu, markersize = m,capsize = cap, label='Cu-SO1')


#x axis label
plt.xlabel('time  /  s', fontsize=p)
plt.xticks([0,30,60,90],fontsize=p)
#y axis label
plt.ylabel('$ \mathrm{\Delta A}$  /  a.u.', fontsize=p)
plt.yticks([0,2,4],fontsize=p)
#create legend
plt.legend(fontsize=p, handlelength=0.5)
##save as png file
plt.xlim([-3,93])
plt.ylim([-0.2,4.2])
filename = 'out/lineplot_activity.png'
plt.savefig(filename, bbox_inches='tight')
```

Danksagung

An erster Stelle möchte ich *Jun.-Prof. Dr. Franziska Thomas* für die Betreueung und Finanzierung der Labormittel während meiner Doktorarbeit danken. Viele späte Stunden sind in diese Arbeit geflossen, aber Du warst bei Problemen immer in Reichweite und hast auch in kurzfristigen Dringlichkeiten eine helfende Hand gereicht. Auch für die finanzielle Unterstützung will ich mich bedanken: einerseits mit einer HiWi-Stelle für eine ausgeglichene Bezahlung unter Angestellten und Stipendiaten und andererseits für die Beschaffung von Versuchschemikalien, egal wie teuer Farbstoffe und Zucker wurden.

Ein großes Dankeschön geht natürlich auch an die *Carl-Zeiss-Stiftung*, die meine Arbeit mit einem Doktorandenstipendium unterstützte und diese Forschung ermöglichte.

Vielen Dank auch an das *3DMM2O Cluster of Excellence* der HeiKa Initiative für abwechslungsreiche Workshops, die Chance neben der Promotion einen MBA zu absolvieren, zahlreiche Treffen mit anderen Doktoranden und jährliche Konferenzen. Der offene Austausch mit dem Cluster-Büro und offene Diskussionsrunden vermittelten Offenheit und das Gefühl, man könne als Doktorand auch neben der Wissenschaft etwas zum Cluster beitragen.

Ich bedanke mich auch bei *Prof. Mastalerz* und *Prof. Klein* für die Nutzung ihrer CD-Spektrometer. Außerdem bedanke ich mich bei *Prof. Milan Kivala* für die bereitwillige Annahme der Zweitprüferschaft meiner Arbeit.

Vielen Dank natürlich auch an meine Kollegen aus dem *AK Thomas*. Danke an *Christina*, *Truc Lam* und *Vanessa,* dass ihr mit mir den manchmal holprigen Start nach dem Corona-Lockdown durchgestanden habt. Wir sind gemeinsam gewachsen und haben die Zielgerade erreicht. Ich denke gerne an unsere Zeit im kleinen Kreis zurück und auch an die Zeit in Barcelona beim EPS, von Diskussionen über Vorträge, Poster oder Essen bis hin zu den nächtlichen Ausgängen. Danke auch an meine Kollegen aus dem CAM-Büro *Christina* – ich muss Dich wohl zwei Mal nennen, *Niklas, Roman* und später *Marcos* und *Agi* für lustige Stunden, kleine Ablenkungen bei anstrengenden Arbeiten und Gespräche beim Kaffee. Ihr seid eine tolle Truppe und habt mir den Alltag versüßt oder manchmal auch einfach erträglich gemacht. Auch an den übrigen AK Danke für spannende Diskussionen am Mittagstisch, leckere Desserts an Geburtstagen und Spaß im Labor, beim Hüte basteln, Karaoke singen, Kickern oder Mario Kart spielen.

Auch vielen Dank an meine Praktikanten über die 4 Jahre. Insbesondere *Stefan Germer* für die wunderbare Arbeit an SO1, *Luise Karle* für die geduldigen Stunden beim Verifizieren des Calcium-Bindungsassays und *Matilde Macrí* für die Hilfe bei der Charakterisierung der Metallbindungen von SO1. Ihr wart eine echte Unterstützung im Labor und tolle Gesellschaft außerhalb.

Ich danke meinen Eltern *Susanne und Jürgen Häge* für die finanzielle Unterstützung, wenn es mal eng wurde, und den Rückhalt und Zuspruch in schwierigen Phasen.

An letzter Stelle danke ich meiner Frau *Franziska Häge* für ein offenes Ohr und Halt in schweren Zeiten, sowie Begeisterung und Eifer in guten Zeiten. Danke, dass Du den Weg mit mir gegangen bist, auch wenn er nicht immer leicht war. Du hast mir immer wieder Mut gemacht an meinen Ambitionen festzuhalten und mir den Rücken gestärkt, um den letzten Meter zu schaffen oder zu sehen, wann eine Pause sinnvoller ist.