JBC Papers in Press. Published on April 2, 2013 as Manuscript M113.460030
The latest version is at http://www.jbc.org/cgi/doi/10.1074/jbc.M113.460030

High-affinity cyclic peptide matriptase inhibitors*

Pedro Quimbar[1], Uru Malik[2], Christian P. Sommerhoff[3], Quentin Kaas[2], Lai Y. Chan[2], Yen-Hua Huang[2], Maresa Grundhuber[3], Kerry Dunse[1], David J. Craik[2], Marilyn A Anderson[1], Norelle L Daly[2,4]

[1]La Trobe Institute for Molecular Science, La Trobe University, Melbourne VIC Australia.
[2]Institute for Molecular Bioscience, The University of Queensland, QLD Australia.
[3]Institute of Laboratory Medicine, University Hospital, Ludwig-Maximilians-University Munich, Germany. [4]Centre for Biodiscovery and Molecular Development of Therapeutics, School of Pharmacy and Molecular Sciences, James Cook University, Cairns QLD Australia

*Running title: *Development of cyclic peptide matriptase inhibitors*

To whom correspondence should be addressed: Norelle L. Daly, James Cook University, Smithfield QLD 4870 Australia Phone: +61 7 4042 1815; Email: norelle.daly@jcu.edu.au

Keywords: NMR spectroscopy; cyclotide; serine proteases

Background: SFTI-1 and MCoTI-II are potent protease inhibitors comprising a cyclic backbone.

Results: Elucidation of structure-activity relationships for SFTI-1 and MCoTI-II was used to design inhibitors with enhanced inhibitory activity.

Conclusion: An analogue of MCoTI-II is one of the most potent inhibitors of matriptase.

Significance: These results provide a solid basis for the design of selective peptide inhibitors of matriptase with therapeutic potential.

SUMMARY

The type II transmembrane serine protease matriptase is a key activator of multiple signaling pathways associated with cell proliferation and modification of the extracellular matrix. Deregulated matriptase activity correlates with a number of diseases, including cancer and hence highly selective matriptase inhibitors may have therapeutic potential. The plant-derived cyclic peptide, sunflower trypsin inhibitor-1 (SFTI-1), is a promising drug scaffold with potent matriptase inhibitory activity. In the current study we have analyzed the structure-activity relationships of SFTI-1 and MCoTI-II, a structurally divergent trypsin inhibitor from *Momordica cochinchinensis* that also contains a cyclic backbone. We show that MCoTI-II is a significantly more potent matriptase inhibitor than SFTI-1 and that all alanine mutants of both peptides, generated using positional scanning mutagenesis, have decreased trypsin affinity, whereas several mutations either maintain or result in enhanced matriptase inhibitory activity. These intriguing results were used to design one of the most potent matriptase inhibitors known to-date with a 290 pM equilibrium dissociation constant, and provide the first indication on how to modulate affinity for matriptase over trypsin in cyclic peptides. This information might be useful for the design of more selective and therapeutically relevant inhibitors of matriptase.

Serine proteases are one of the largest known family of proteases (1), and are involved in a range of cellular processes such as apoptosis, inflammation, blood coagulation and extracellular matrix remodeling (2). Several mechanisms are involved in controlling the activity of serine proteases, including synthesis as inactive zymogens and production of specific protease inhibitors. The deregulation of these endogenous controls has dramatic consequences, and can lead to autoimmune and metabolic diseases and to an increased susceptibility to infections and to cancer (3).

Matriptase is a type-II transmembrane serine protease that is expressed strongly in the human epithelia (4). The expression of this protease, both at the RNA and protein level increases significantly during the progression of prostate cancer (5). The oncogenic potential of matriptase has been demonstrated in transgenic mice where overexpression of matriptase caused spontaneous squamous cell carcinoma. However, over-expression together with a matriptase inhibitor counteracted the oncogenic effects (6), indicating that inhibition of matriptase has significant potential as a therapeutic strategy. Administration of inhibitors in a pancreatic tumor model in mice led to the inhibition of matriptase *in vivo* for at least 24 hours (7), and in a prostate tumor mouse model inhibitors reduced primary tumor growth by 40% as well as reducing the prevalence of metastases (8). Genetic reduction of matriptase in mice resulted in reduced tumor growth, invasiveness and migration *in vitro* (9) and correlated with *in vivo* studies where matriptase-deficient PC3 and DU-145 cells exhibited reduced growth and development compared to control cells when explanted into nude mice (10). Overall, these studies highlight the potential of using matriptase inhibitors as a treatment approach to halt growth and spread of cancer cells.

A major challenge in the design of protease inhibitors is selectivity (11). Broad range inhibitors have caused detrimental effects in clinical trials as a result of interference with signaling pathways and processes that were not foreseen (12). Therefore, when designing inhibitors for pharmaceutical purposes, it is essential that they are selective for the desired target while also having high affinity. Several studies have focused on small molecules for the design of matriptase inhibitors, and have resulted in potent inhibitors with selectivity over other serine proteases such as thrombin (8). However, limited information is available regarding selectivity relative to other closely related serine proteases, particularly the prototypic trypsin.

The active sites of matriptase and trypsin are very similar in structure and both enzymes belong to the S1 family and subfamily A according to the MEROPS database (13). However, there are some important differences, as shown in Figure 1, which could potentially be exploited for the design of novel inhibitors. The active site of all serine proteases is surrounded by eight loops (1) that will be referred to as I to VIII in this manuscript. Loop II of matriptase has ten more residues than that of trypsin, resulting in an additional bulge that constitutes the most obvious topological difference between the two protease active sites (Figure 1). Loops I, III and V also display different sequences between the two proteases, resulting in slightly different conformations. In contrast, loops IV, VI and VIII, which cluster at one end of the active site, have similar conformations in matriptase and trypsin. The active site of matriptase is much more negatively charged than that of trypsin, mostly because of two additional aspartate residues in loop IV and three aspartate residues in the extended loop II. The negative charges borne by matriptase loop II are partly compensated by two positively charged arginine residues located at the tip of the loop.

Naturally occurring peptidic inhibitors are a promising starting point for the design of novel matriptase inhibitors as they can provide a relatively large surface area to enhance the potential of designing selective analogues. The use of naturally occurring amino acids also leaves open the possibility of using plants as "factories" for producing the modified inhibitors via a low cost route (14). SFTI-1 is a 14-residue sunflower peptide that is a potent inhibitor of both trypsin and matriptase, with an inhibition constant against matriptase of 0.92 nM (15). However, SFTI-1 is also active against other serine proteases within the trypsin clan, which limits its therapeutic application. Despite this limitation, the size, versatility and stability of the SFTI-I backbone makes it an appealing scaffold for the design of drugs (16). Swedberg et al (2009), for example, enhanced both the affinity and specificity of the SFTI-1 backbone against kallikrein 4 (KLK4) through modifications of the inhibitor. They used positional scanning of a synthetic combinatorial library to study protease specificity and generated a mutant of SFTI-1 with an inhibition constant of 3.59 nM against KLK4 compared to no inhibition by native SFTI-1 (17). Although this mutant was not tested *in vivo* it retained high inhibitory potential against KLK4 after days of incubation in serum, highlighting the advantages of the SFTI-1 scaffold in the design of drugs.

Several other researchers have also explored the structure-activity relationships of SFTI-1 and shown the importance of particular residues for a

range of proteases. The introduction of naturally and non-naturally occurring residues has been explored, as have the roles of the disulfide bonds in activity (18-22). Enhancements in activity and selectivity have been obtained, highlighting the potential of this cyclic scaffold. However, further study is required to determine if significant selectivity between trypsin and matriptase is achievable.

The stability and versatility of cyclic peptides is not limited to SFTI-1. Other cyclic peptides, such as members of the cyclotide family, have functions as diverse as insecticidal activity (23), HIV inhibition (24), and hemolytic activity (25). Their extraordinary stability to pH, heat and biological fluids also makes them ideal scaffolds for pharmaceutical design (14,17). In the current study we have discovered that a member of the trypsin inhibitory subfamily of cyclotides, *Momordica cochinchinensis* trypsin inhibitor-II (MCoTI-II) (26), a 34-residue cyclic peptide, is also a potent inhibitor of matriptase. Furthermore, we have used alanine scanning and point mutations to design inhibitors, based on SFTI-1 and MCoTI-II, which have enhanced potency and modified selectivity against matriptase. The sequences and structures of SFTI-1 and MCoTI-II are given in Figure 2 and highlight the sequence and structural diversity of these cyclic peptidic trypsin inhibitors.

EXPERIMENTAL PROCEDURES

Peptide Synthesis and purification - Peptides were synthesized on a 0.5 mmole scale using manual solid-phase peptide synthesis with Boc-chemistry. A PAM-Gly-Boc resin was used with an S-tritylmercaptopropionic acid linker to facilitate cyclization, as described previously (27,28). Crude peptides were purified using a series of Phenomenex C18 columns on RP-HPLC. Gradients of 1% min^{-1} of 0-80% solvent B (90% Acetonitrile in 0.045% TFA in H_2O) and solvent A (aqueous 0.05% TFA in H_2O) were employed and the eluant was monitored at 215 nm and 280 nm. The purity of the peptides was examined by analytical RP-HPLC on a Phenomenex Jupiter 5μ C18 300 Å 150 x 2.0 mm column and masses were determined by electrospray mass spectrometry. Grafted and native SFTI-1 peptides were folded in solution at 0.1 mg/mL using a range of buffer conditions. Synthetic native SFTI-1 peptide was cyclized and oxidized in a two-step process. Cyclization was achieved in the presence of 0.1 M tris(2-carboxyethyl)phosphine incubated for 24 h and the cyclic product then purified by RP-HPLC. The cyclic reduced peptide was oxidized in 0.1 M ammonium bicarbonate (pH 8.0). By contrast, cyclization and oxidation of MCoTI-II grafted peptides was done in a single step reaction with 0.1 M ammonium bicarbonate (pH 8.5) for 12 h. SFTI-1 grafted peptides were oxidized in 0.1 M ammonium bicarbonate (pH 8.5).

NMR analysis - Peptides were dissolved in 90% H_2O/10% D_2O (v/v) (1 mM). D_2O (99.9%) was obtained from Cambridge Isotope Laboratories, Woburn, MA for ^1H NMR measurements. Spectra were recorded at 290-298 K on Bruker Avance 500 MHz and 600 MHz spectrometers. Two-dimensional spectra included TOCSY, NOESY, and COSY. NOESY mixing times of 200-300 ms were used. Spectra were analyzed using SPARKY (29). The sequential assignment procedure pioneered by Wüthrich (30) was used to sequence-specifically assign the amino acids, using TOCSY and NOESY spectra.

Enzyme kinetics –Titration - The concentrations of active inhibitors were determined by titration against burst-titrated bovine trypsin assuming a one to one interaction between the enzyme and the inhibitor. Reactions were conducted in 96 well plates using buffer (50 mM Tris-HCl, 150 mM NaCl, 0.01% Triton X-100, 0.01% sodium azide, pH 7.6) and serial dilutions of each inhibitor with trypsin (25 nM). After preincubation at 37°C the fluorogenic substrate bz-R-AMC (75 μM) was added. Hydrolysis of the substrate was then measured over 10 min with λ_{em} and λ_{ex} of 360 and 465 nm respectively on a HTS 7000 Bio assay reader (Perkin Elmer, Rodgau-Jügesheim, Germany). The concentrations of inhibitors with a large K_{iapp} could not be determined by titration.

Enzyme kinetics – K_i determination - Inhibition constants of the inhibitors against matriptase and trypsin were calculated by pre-incubating the recombinant human matriptase catalytic domain (2 pM, R&D systems, Inc.) and bovine trypsin (10 pM), respectively, with serial dilutions of the inhibitors in buffer (50 mM Tris-HCl, 150 mM NaCl, 0.01% Triton X-100, 0.01% sodium azide pH 8.0). After pre-incubation of 60 min the substrate Boc-QAR-AMC (5 μM) or tos-GPR-AMC (5 μM), respectively, was added to the reaction mixture and hydrolysis was quantified

as described above. Additional experiments were performed with pre-incubation times of between 60-180 minutes to verify that each inhibitor was in equilibrium with the protease. The inhibition curves generated were fitted to Morrison's equation (31), to obtain the inhibition constant for each of the tested peptides. The results presented are mean ± SEM of ≥ 4 independent experiments.

Three-dimensional structure calculations - Preliminary three-dimensional (3D) structures were calculated using automated NOE assignment within CYANA (32). A final set of 100 structures was calculated and the 20 lowest energy structures were selected for further analysis. Structures were analyzed using MolProbity (http://molprobity.biochem.duke.edu), and MolMol (33) and PyMol (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC) were used to display the structural ensembles and surfaces of the peptides, respectively.

Molecular modeling - The three-dimensional (3D) structures of the complexes between SFTI-1 and trypsin (PDB identifier 1sfi), between SFTI and the protease domain of matriptase (PDB identifier 3p8f), and between MCoTI-II and trypsin (Daly *et al.*, unpublished) were determined experimentally using X-ray crystallography independently by different research groups (16,34). The 3D structure of MCoTI-II in complex with the protease domain of matriptase was built in this study by homology using Modeller9v11 (35). The three previously mentioned X-ray structures were used as structural templates to generate 100 models of the matriptase/MCoTI-II complex. The model with the lowest DOPE score (36) was selected and was then refined using molecular dynamic (MD) simulations. MD simulations were performed using the program PMEMD from the AMBER 12 package with the ff12SB force field (37). The protease domain of matriptase/MCoTI-II complex (274 residues) was solvated in a truncated octahedral periodic box with ~4900 TIP3P water molecules. Three sodium ions were added to neutralize the system. The system was then equilibrated using a similar strategy to Cerutti *et al.* (38). A steepest descent minimization of 2000 steps was first carried out with the solute restrained to its positions and then further 2000 steps of steepest descent minimization was performed with the solvent restrained to its position. MD simulations were then performed with the solute atoms restrained to their position by a harmonic potential whose spring constant was progressively decreased from 16 to 1 kcal/mol Å^2 over 150 ps in the NVT ensemble, before it was progressively decreased from 1 to 0 over 300 ps of standard NPT simulations (300 K, 1 atm). An MD simulation of 300 ps with a 1 fs time step was performed followed by a production run of 20 ns with a 2fs time step. All bonds involving hydrogen atoms were constrained with the SHAKE algorithm (39). Long-range electrostatic interactions were simulated with the particle-mesh Ewald method (40). Similarly prepared 20 ns MD simulations starting from the crystallographic structures of SFTI-1/matriptase (3p8f), SFTI-1/trypsin (1sfi) and MCoTI-II/trypsin were also performed. These simulations gave information on the dynamics of the interactions at the interface. All simulations were performed in triplicates using different random seeds.

The 3D structures of complexes involving mutants of SFTI-1 and MCoTI-II were modeled using the mutation procedure implemented in Modeller9v11. Each homology model was subjected to 4000 steps of minimization using PMEMD. The dynamics of complexes involving a selection of mutants, including [I10R]SFTI-1 [I10D]SFTI-1, [I10K]SFTI-1, [I10R]SFTI-1, [I7A+I10R]SFTI-1, [V3R]MCoTI-II, [I7A]MCoTI-II and [V3R+I7A]MCoTI-II, were studied using 5 ns MD.

RESULTS

Synthesis and characterization of peptides - Variants of SFTI-1 and MCoTI-II were synthesized to identify residues that are essential for inhibition of matriptase or trypsin. The mutants were synthesized using BOC chemistry and thioester-mediated backbone cyclization and included alanine substitutions as well as several other single or double amino acid changes. In general, the SFTI-1 mutants were cyclized and oxidized in separate steps as cyclization occurs more efficiently in the presence of a reducing agent. By contrast, the MCoTI-II analogues could be cyclized and oxidized in a single step. It is possible that the presence of three disulfide-bonds in MCoTI-II compared to one in SFTI-1 facilitates the cyclization process via a thia-zip

mechanism (41). All peptides were purified by RP-HPLC and analyzed by mass spectrometry. The structures of the peptides were analyzed using NMR chemical shift analysis (Figure 3A) to confirm the native fold was present, as shown for [R2A]SFTI-1 and [I10R]-SFTI-1 in Figure 3B. The similarity in αH chemical shifts with the native peptides indicates that the overall fold was maintained despite the mutations.

Enzyme kinetics - Inhibition constants for the SFTI-1 and MCoTI-II mutants against matriptase and trypsin are given in Table 1, and graphically presented in Figure 4. Several mutations had selective influences on the inhibitory activity as outlined below for the SFTI-1 and MCoTI-II mutants:

SFTI-1 mutants - The trypsin inhibitory activity for the SFTI-1 alanine mutants was consistent with our previous results (27). Several of the alanine mutants displayed decreased potency compared to the native peptide for both trypsin and matriptase. For example, the alanine substitutions at positions 2, 4, 5, 6 and 14 led to a 7.5 to 50-fold increase in the K_i against matriptase. Although the R2A mutant had significant loss of inhibition against matriptase, it was still a potent inhibitor of trypsin (K_i 160 pM compared to 1.7 pM for the native molecule). Substitution of the active site residue, Lys5, abolished inhibitory activity for both enzymes. Additional substitutions were tested to assess the charge requirements at positions 2 and 5. Substitution of Arg2 with Lys had no effect on trypsin inhibitory activity although the K_i for matriptase increased 6-fold. Substitution of Lys5 with Arg had little effect on trypsin or matriptase inhibition.

[I7A]SFTI-1 exhibited a 700-fold decrease in potency relative to the native peptide against trypsin but was 2-fold more potent against matriptase than the native peptide. Similarly, [P8A]SFTI-1 had a 20-fold decreased trypsin activity but a 7-fold increase in matriptase activity, and [I10A]SFTI-1 a ~100 fold decrease in potency against trypsin but a 2-fold increase in activity against matriptase. The contrasting effect of these mutations against trypsin and matriptase indicated that these positions are potential sites for exploring selectivity. Additional point mutations were made and the selectivity for matriptase was modulated in [I10R]SFTI-1, which had a 30 fold increase in activity against matriptase and a decrease in trypsin inhibitory activity of 2 fold. Mutating position 10 to an aspartic acid or glycine significantly decreased activity against both matriptase and trypsin. A combination of the I7A and I10R mutations resulted in a peptide with a 700-fold decrease in trypsin inhibitory activity and a 4-fold increase in matriptase activity.

MCoTI-II mutants - Analysis of the matriptase inhibitory activity of MCoTI-II revealed it is a more potent inhibitor (K_i = 2.8 nM) than SFTI-1 (K_i = 200 nM) (Table 1). In contrast, both inhibitors had similar K_is against trypsin, consistent with previous studies (27,42). Alanine mutants in the active site loop of MCoTI-II (loop 1) were synthesized to determine the role each residue plays in enzyme inhibitory activity. Additional alanine substitutions were generated for residues in loops 5 and 6 that were predicted to be in close proximity to the enzyme based on modeling studies using the crystal structure of MCoTI-II in complex with trypsin. Sites from loops 5 and 6 that were mutated included residues 2, 3, 26, 28, 30 and 31 (Table 1).

In general, most of the alanine substitutions led to decreased inhibitory activity, but the influence of the substitutions varied for trypsin and matriptase. As expected, substitution of the trypsin active site residue (Lys6) with an alanine abolished inhibitory activity for both trypsin and matriptase. Preference for lysine in this position was confirmed with the K6R substitution, which decreased the inhibitory activity against both trypsin and matriptase relative to the native peptide. Interestingly, the mutation of Val3 in MCoTI-II to an alanine maintained potency against matriptase and decreased potency against trypsin. Based on the selective influence of this mutation at the P4 position of MCoTI-II and the importance of Arg2 in the P4 position in SFTI-1 for potency against matriptase, a MCoTI-II-V3R mutant was synthesized. Analysis of the kinetics of this mutant indicated that it did improve inhibitory activity against matriptase and resulted in one of the most potent matriptase inhibitors known, with a K_i of 290 pM.

The mutation I7A of MCoTI-II significantly decreased the trypsin inhibitory potency but had a lesser impact on the matriptase potency. A combination of the I7A mutation with the V3R mutation resulted in a peptide with a two-fold decrease in matriptase inhibitory activity relative

to the native peptide but a >1000 fold decrease in trypsin inhibitory activity.

Hybrids between SFTI-1 and MCoTI-II - Although MCo-TI-II is a more potent matriptase inhibitor than SFTI-1, SFTI-1 is smaller, which is potentially advantageous for pharmaceutical purposes as it is cheaper to synthesize and easier to modify. A series of grafted peptides was thus generated to investigate the relative importance of the SFTI-1 and MCoTI-II scaffolds and the respective binding loops for the inhibitory activity. Two grafted peptides were generated for the SFTI-1 loop in the MCoTI-II scaffold to investigate the role of the binding loop and loop 5 in interacting with the enzymes. The latter peptide also included the substitution of Lys6 into alanine to prevent interference of MCoTI-II native loop in our studies. These peptides, MCoTI-II-S^1 and MCoTI-II-S^5, displayed poor inhibition of both enzymes. The peptide SFMC, which had the MCoTI-II binding loop grafted onto the SFTI-1 scaffold, was as efficient as native SFTI-1 in inhibition of both trypsin and matriptase, as shown in Table 2. However, the SFMC mutant was not as active as native MCoTI-II against matriptase and the additional peptide/enzyme interactions that occur with the larger MCoTI-II scaffold may be important for enhanced enzyme inhibitory activity.

Structure determination using nuclear magnetic resonance spectroscopy - NMR structures were determined for selected SFTI-1 variants to enhance the understanding of the structure-activity relationships and for use in molecular modeling of inhibitor/enzyme complexes. [I10R]SFTI-1 was chosen because it is the most potent SFTI-1 mutant against matriptase, and R2A was chosen because it was one of the least active analogues against matriptase, aside from the K5A mutant. Structures were determined using torsion angle dynamics in the program CYANA and the 20 lowest energy structures chosen to represent the fold. Energetic and geometric statistics are given in Table 3. The structures were analyzed using PROMOTIF and revealed that the major element of secondary structure in both R2A and I10R is a β-hairpin with the strands involving residues 2-4 and 10-12. Comparison of the structures with the native peptide confirmed the overall fold was maintained and therefore the observed effects of the mutations on the inhibition constants are likely to arise from interactions between the side chains of the substituted amino acids and the proteases (Figure 3B). However, several hydrogen bonds were missing from the mutant structures compared to the native peptide, suggesting that the overall fold had been destabilized, which might also influence the inhibitory activity.

Molecular modeling - The 3D structures of complexes involving matriptase or trypsin were used to propose explanations for the activity of SFTI-1 and MCoTI-II variants. The structures of complexes between SFTI-1 and trypsin, SFTI-1 and matriptase, and MCoTI-II and trypsin have been determined experimentally by X-ray crystallography (16,34) and were used here to model the structure of the matriptase/MCoTI-II complex by homology with refinement using 20 ns MD simulations (Figure 5A). The three other complexes involving wild-type SFTI-1 or MCoTI-II were also simulated for 20 ns by MD (Figure 5A), and these simulations were used to compare the dynamics of the molecular interactions at the interface. All simulations quickly converged, as indicated by the stabilization of the Cα atoms RMSD from the initial homology model (matriptase/MCoTI-II complex) or from the crystal structures (Figure S1). Structural models of the complexes between the peptide variants and the proteases were modeled by comparison based on the wild-type models and refined by 5 ns MD. These simulations allow local conformational change to occur and also provide information on the structural dynamics of the complexes.

Position 10 of SFTI-1 faces loop II of the proteases (Figure 5A), which is ten residues longer in matriptase than trypsin. Several SFTI-1 variants at position 10 were prepared and investigated for their ability to discriminate between the two proteases. The I10A substitution caused small shifts of the positions of SFTI-1 side chains Arg2, Phe12 and Asp14 in both protease complexes, decreasing the distance by ~2 Å between the positively charged guanidinium group from Arg2 of SFTI-1 and the negatively charged carboxylic group of matriptase Asp709 compared to wild-type SFTI-1 (Figure 6 and S2). As a result the electrostatic interactions were improved and at the same time the buried surface area remained globally similar (Table 4). The introduction of a negatively charged residue at position 10 of SFTI-1 (I10D) dramatically decreased the potency for

matriptase, and in the corresponding model the substituted aspartate at position 10 had moved away from loop II and IV of matriptase relative to the position of isoleucine 10 in wild-type SFTI-1 (Figure 6). This probably arose from charge repulsions between Asp10 of SFTI-1 and Asp660, Asp661 and Asp705. Indeed the distance between the C α of matriptase Asp705 and position 10 in SFTI-1 wild-type and the I10D mutant had increased slightly by ~1 Å (Figure 6 and Figure S2). Furthermore the buried surface area in the mutant had decreased on average by ~50 Å^2 (Table 4). By contrast, the substitution of Ile10 by the positively charged residues, arginine or lysine, resulted in a more potent inhibitor of matriptase than wild-type SFTI-1, possibly because of positive electrostatic interactions with loop II (Figure 6). The two variants [I10R]SFTI-1 and [I10K]SFTI-1 exhibited a small decrease in activity for trypsin. The models suggest that [I10R]SFTI-1 is a better inhibitor of matriptase than [I10K]SFTI-1 because an arginine residue at position 10 can establish a salt bridge with matriptase Asp705 whereas the smaller side chain of lysine does not allow for the formation of this interaction (Figure 6 and Figure S2). During most of the 5 ns simulation of the complex [I10R]SFTI-1/matriptase, the guanidinium group of Arg10 established at least one hydrogen bond with the carboxylic group of Asp705 (distance oxygen to nitrogen of ~3 Å), whereas over the 5 ns of the [I10K]SFTI-1/matriptase simulation, Lys10 did not form a hydrogen bond with Asp705 (Figure S2).

Position 7 of SFTI-1 contacts loop I and V (Figure 5A), which has little sequence conservation between matriptase and trypsin, and this position is therefore potentially important in determining the selectivity for matriptase. According to the modeling, the mutation I7A did not change the orientation of SFTI-1 in the trypsin binding site (Figure 6), but resulted in the largest decrease in buried surface area (~50 Å^2) observed among all the SFTI-1 variants in complex with trypsin (Table 4). The variant [I7A]SFTI-1 was also a more potent inhibitor of matriptase than the wild-type. Unexpectedly the double mutant [I7A+I10R]SFTI-1 was not a more potent inhibitor of matriptase than [I10R]SFTI-1. The corresponding model suggested that this doubly mutated peptide had a different binding mode to [I10R]SFTI-1, and, in this new binding mode, Arg10 cannot form a salt bridge with Asp705 (Figure 6). By contrast, the effect of the double mutation for inhibition of trypsin was cumulative, in line with the models of [I7A]SFTI-1, [I10R]SFTI-1 and [I7A+I10R]SFTI-1 in complex with trypsin that displayed similar binding modes (Figure 6). [I7A+I10R]SFTI-1 was the worst inhibitor for trypsin among all SFTI-1 variants tested in this study, and the most selective for matriptase among SFTI-1 variants.

Substitution of Arg2 of SFTI-1 with an alanine resulted in a loss of potency for both proteases. Arg2 is involved in charge interactions with Asp14 (SFTI-1), and also potentially establishes cation-π interactions with several aromatic side chains, including Phe12 from SFTI-1 and a tryptophan from the proteases (Trp216 for trypsin and Trp826 for matriptase). Matriptase has two phenylalanines, *i.e.* Phe706 and F708 with proximity to Arg2 and these residues might also participate in cation-π interactions with the Arg2 side chain (Figure 7).

MCoTI-II globally displayed more flexibility than SFTI-1 during the MD simulations, and the tip of loop 6 (positions 32, 33, 34 and 1) was the most flexible region of MCoTI-II in the complexes with matriptase and trypsin (Figure S3). By contrast, the inhibitory loop (around Lys5) was the most stable region of the peptide and the conformation of this inhibitory loop was nearly identical between MCoTI-II and SFTI-1 in complex with the two proteases (Figure 5B). As a consequence, the C α at position 3 of MCoTI-II and position 2 of SFTI-1 occupies the same region in the active sites, and because Arg2 of SFTI-1 was shown to be important for the binding affinity, the mutant [V3R]MCoTI-II was predicted to have improved activity. Indeed, the substitution V3R resulted in the best matriptase inhibitor among MCoTI-II variants. For binding to matriptase, the substitution V3R resulted in a large increase of buried surface area (~ 180 Å^2 on average in Table 4) and, similarly to the comments made for the analysis of the mutant [I7A]SFTI-1, Arg3 can potentially establish positive electrostatic interactions with Asp709 in matriptase (Figure 7).

Every modification to the inhibition loops of MCoTI-II, *i.e.* the alanine substitutions in positions 5 to 8, resulted in a drop of activity for both proteases, which could be explained by the

tight fit of the inhibition loop in both active sites (Figure 5B). Interestingly, the I7A substitution caused a decrease of activity that was much more dramatic for trypsin than matriptase, and this substitution resulted in an important loss of ~150 $Å^2$ of buried surface area in the models with trypsin. The double mutant [V3R+I7A]MCoTI-II was as specific for matriptase as for trypsin, with each substitution independently contributing to the loss of activity for trypsin. The models showed that the substitutions at positions 3 and 7 should have an independent impact because these residues are distant from each other and did not cause obvious changes of binding mode. This result contrasts with the change in binding mode predicted for the [I7A+I10R]SFTI-1 double mutant.

DISCUSSION

Naturally occurring peptides with cyclic backbones have significant promise in drug design (14,17,43), and in this study we have highlighted the potential of the frameworks of cyclic trypsin inhibitors from seeds. In particular, we have discovered that MCoTI-II is a potent inhibitor of matriptase and also generated substantial structure-activity data regarding MCoTI-II and SFTI-1 that has provided insights on how to modulate affinity towards matriptase over the prototypic trypsin.

Alanine scanning of SFTI-1 against trypsin and matriptase highlighted enzyme-specific requirements for high affinity inhibition (Table 1, Figure 3). In SFTI-1, Arg2 is indispensable for inhibition of matriptase, whereas for trypsin there is loss of inhibition yet the R2A mutant remains a potent inhibitor with a nanomolar affinity. The importance of this arginine residue was previously highlighted by Long et al. (2001) (15) who suggested that it is involved in a cation-π interaction with Phe706 and Phe708 of matriptase (15). In addition, It has been reported that Arg2 is critical for matriptase inhibitory activity, with mutation to a phenylalanine derivative resulting in a 900 fold decrease in potency (20). We further characterized the role of this residue by substituting it with a lysine to determine whether the charge is the requirement at this position. However, this substitution resulted in a significant loss of inhibitory activity against matriptase suggesting that Arg2 is an absolute requirement. These results are consistent with the substrate specificity of matriptase, as arginine is preferred at the P4 position (44), *i.e.* the position Arg2 occupies when bound to the enzyme (Figure 5).

Mutations in the substrate sites P2, P1 and P1' of SFTI-1 had a major impact on inhibitory activity. As expected, mutation of Lys5 (P1) had the most significant effect, consistent with a previous study (20). Substitution of the adjacent residues (Thr4 and Ser6) also caused a significant loss of affinity against both enzymes. In general, the mutations had more detrimental effects on inhibitory activity against trypsin compared to matriptase, indicating that SFTI-1 has a highly optimized framework in respect to trypsin inhibitory activity, reflecting its biological function in plants as a deterrent of insect and animal predators. Although substitution of Ser6 with an alanine has significant effects on trypsin and matriptase activity it produces an SFTI-1 variant with potent inhibitory activity against chymotrypsin (45).

In contrast to the decreased activity observed for the majority of the SFTI-1 alanine mutants against trypsin, the I7A, P8A and I10A substitutions resulted in enhanced inhibitory activity against matriptase. In addition, F12A had significantly less activity against trypsin but maintained similar activity to wild type SFTI-1 against matriptase. Additional mutations at positions 7 and 10 were explored, and one the variant I10R was identified to have 30-fold enhanced inhibitory activity against matriptase compared to the wild-type peptide. The models generated for the inhibitor [I10R]SFTI-1 in complex with both enzymes illustrated why this substitution was ineffective in the trypsin complex as the side chain of the Arg10 is exposed to the solvent and does not interact with the enzyme. However, in matriptase, Arg10 establishes a salt bridge with Asp705 of matriptase (3.1 Å) when bound to the enzyme. The additional electrostatic interaction between these residues is likely to be responsible for stabilizing the complex, which is reflected in enhanced inhibition.

Combining the I7A and I10R mutants for SFTI-1 resulted in a peptide with significantly decreased trypsin activity (700-fold) and enhanced matriptase activity (4-fold) relative to the wild-type peptide. This selective enhancement of matriptase activity augurs well for the design of

more selective inhibitors using this framework. In particular, further exploration of mutants at positions 8 and 12 might provide useful information for subsequent design studies.

Initial studies using molecular modeling provided significant insight into the binding of SFTI-1 to matriptase and predicted a role for Arg2, Lys5, Ile10 and Phe12 in binding (15). The importance of Arg2 and Lys5 has been confirmed experimentally (20) and is consistent with the current study. The importance of Ile10 and Phe12 has also been explored, and both residues influence activity. For example, replacement of Ile10 with a glutamine enhanced the selectivity for matriptase versus thrombin but decreased the potency against matriptase (20). Ile10 was also highlighted in an analysis of the crystal structure of SFTI-1 bound to matriptase, and it was suggested that this residue would be a useful site for mutational analysis to improve binding (34). However, it was suggested that the replacement of Ile10 with a basic residue, arginine or lysine, would not be beneficial because increasing the flexibility may result in a loss of entropy upon binding. Our results confirm that this site is indeed useful for modulating the inhibitory activity against matriptase and that the increase in potency of the I10R and I10K variants indicates that the introduction of more flexible residues does not impact negatively on activity. Very recently an acyclic SFTI-1 analogue has been synthesized with an I10R substitution in addition to truncation and introduction of a His residue. This peptide also has enhanced matriptase affinity as well as improved selectivity against trypsin (46).

Consistent with the SFTI-1 alanine mutants, several of the alanine mutants of MCoTI-II had more significant losses of activity against trypsin relative to matriptase. Lys9 and Lys10 displayed enzyme-specific dependencies with mutation of Lys9 influencing affinity towards matriptase more than for trypsin. By contrast, mutation of Lys10 influenced inhibition of trypsin more than that of matriptase. Val3 was the only MCoTI-II mutant to maintain activity against matriptase, and mutation of this residue to an arginine resulted in one of the most potent inhibitors of matriptase. The model of [V3R]MCoTI-II shows that the mutated Arg3 residue sits in the substrate S4 pocket (Figure 7) supporting the strong preference for an arginine at this position. Combining the V3R mutant with I7A, one of the least active inhibitors against trypsin, reversed the trend of having higher trypsin potency relative to matriptase, as a result of a 1200-fold decrease in potency against trypsin whilst maintaining potency against matriptase. Based on these results, MCoTI-II also appears to be a valuable framework for the development of more selective inhibitors of matriptase.

Analysis of the hybrid peptides of SFTI-1 and MCoTI-II indicates that the SFTI-1 framework is more amenable to substantial sequence changes than MCoTI-II. Grafting the SFTI-1 active site loop into the MCoTI-II resulted in significant losses in activity against both trypsin and matriptase. By contrast, the loss of activity for the hybrid with the MCoTI-II active site loop grafted into the SFTI-1 framework (SFMC) was minimal, despite having one residue less than the wild-type peptide. A naturally occurring example of a peptide similar to SFTI-1 has been found in sunflower (SFT-L2) that comprises only 12 residues and has significant sequence differences in the bioactive loop (47). Although only containing 12-residues the structure of SFT-L2, braced by a disulfide bond and cyclic backbone, is extremely well defined and has been suggested as a useful framework in drug design. The minimized framework of SFMC might represent a new lead molecule for the design of inhibitors against serine proteases.

In summary, the structure-activity data elucidated in this study are a solid basis for the design of selective peptide inhibitors of matriptase and highlights the versatility of cyclic peptides for drug design. The use of disulfide-rich, cyclic peptides as scaffolds is emerging as a powerful approach for the design of novel drug leads. Using this approach a cyclic cone snail venom peptide has been engineered from the naturally occurring acyclic peptide, that is orally active in a rat model of pain (48). In addition, a modified cyclotide has oral activity in an *in vivo* mouse model of visceral pain (43). The latter study exemplifies the potential of the cyclotide scaffold for conferring oral activity and suggests that selective inhibitors of matriptase, based on cyclic peptide scaffolds, might hold significant promise for the treatment of cancer.

References

1. Page, M.J., and Di Cera, E. (2008) Serine peptidases: classification, structure and function. *Cell. Mol. Life Sci.* **65**, 1220-1236
2. Heutinck, K.M., ten Berge, I.J., Hack, C.E., Hamann, J., and Rowshani, A.T. (2010) Serine proteases of the human immune system in health and disease. *Mol. Immunol.* **47**, 1943-1955
3. Conlan, B.F., and Anderson, M.A. (2011) Circular micro-proteins and mechanisms of cyclization. *Curr. Pharm. Des.* **17**, 4318-4328
4. Oberst, M.D., Singh, B., Ozdemirli, M., Dickson, R.B., Johnson, M.D., and Lin, C.-Y. (2003) Characterization of Matriptase Expression in Normal Human Tissues. *J. Histochem. Cytochem.* **51**, 1017-1025
5. Bergum, C., and List, K. (2010) Loss of the matriptase inhibitor HAI-2 during prostate cancer progression. *Prostate* **70**, 1422-1428
6. List, K., Szabo, R., Molinolo, A., Sriuranpong, V., Redeye, V., Murdock, T., Burke, B., Nielsen, B.S., Gutkind, J.S., and Bugge, T.H. (2005) Deregulated matriptase causes ras-independent multistage carcinogenesis and promotes ras-mediated malignant transformation. *Genes Develop.* **19**, 1934-1950
7. Napp, J., Dullin, C., Muller, F., Uhland, K., Petri, J.B., van de Locht, A., Steinmetzer, T., and Alves, F. (2010) Time-domain in vivo near infrared fluorescence imaging for evaluation of matriptase as a potential target for the development of novel, inhibitor-based tumor therapies. *Int. J. Cancer* **127**, 1958-1974
8. Steinmetzer, T., Schweinitz, A., Sturzebecher, A., Donnecke, D., Uhland, K., Schuster, O., Steinmetzer, P., Muller, F., Friedrich, R., Than, M.E., Bode, W., and Sturzebecher, J. (2006) Secondary amides of sulfonylated 3-amidinophenylalanine. New potent and selective inhibitors of matriptase. *J. Med. Chem.* **49**, 4116-4126
9. Forbs, D., Thiel, S., Stella, M.C., Sturzebecher, A., Schweinitz, A., Steinmetzer, T., Sturzebecher, J., and Uhland, K. (2005) In vitro inhibition of matriptase prevents invasive growth of cell lines of prostate and colon carcinoma. *Int. J. Oncol.* **27**, 1061-1070
10. Sanders, A.J., Parr, C., Davies, G., Martin, T.A., Lane, J., Mason, M.D., and Jiang, W.G. (2006) Genetic reduction of matriptase-1 expression is associated with a reduction in the aggressive phenotype of prostate cancer cells in vitro and in vivo. *J. Exp. Ther. Oncol.* **6**, 39-48
11. Swedberg, J.E., and Harris, J.M. (2012) Natural and engineered plasmin inhibitors: applications and design strategies. *Chembiochem* **13**, 336-348
12. Overall, C.M., and Kleifeld, O. (2006) Tumour microenvironment - opinion: validating matrix metalloproteinases as drug targets and anti-targets for cancer therapy. *Nat. Rev. Cancer* **6**, 227-239
13. Rawlings, N.D., Barrett, A.J., and Bateman, A. (2012) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* **40**, D343-350
14. Craik, D.J., Swedberg, J.E., Mylne, J.S., and Cemazar, M. (2012) Cyclotides as a basis for drug design. *Expert Opin. Drug Discov.* **7**, 179-194
15. Long, Y.Q., Lee, S.L., Lin, C.Y., Enyedy, I.J., Wang, S., Li, P., Dickson, R.B., and Roller, P.P. (2001) Synthesis and evaluation of the sunflower derived trypsin inhibitor as a potent inhibitor of the type II transmembrane serine protease, matriptase. *Bioorg. Med. Chem. Lett.* **11**, 2515-2519
16. Luckett, S., Garcia, R.S., Barker, J.J., Konarev, A.V., Shewry, P.R., Clarke, A.R., and Brady, R.L. (1999) High-resolution structure of a potent, cyclic proteinase inhibitor from sunflower seeds. *J. Mol. Biol.* **290**, 525-533
17. Swedberg, J.E., Nigon, L.V., Reid, J.C., de Veer, S.J., Walpole, C.M., Stephens, C.R., Walsh, T.P., Takayama, T.K., Hooper, J.D., Clements, J.A., Buckle, A.M., and Harris, J.M. (2009) Substrate-guided design of a potent and selective kallikrein-related peptidase inhibitor for kallikrein 4. *Chem. Biol.* **16**, 633-643
18. Legowska, A., Bulak, E., Jaskiewicz, A., Maluch, I., Sieracki, M., Wysocka, M., Lesner, A., and Rolka, K. (2010) Analogues of trypsin inhibitor SFTI-1 with disulfide bridge substituted by various length of carbonyl bridges. *Protein Pept. Lett.* **17**, 1223-1227

19. Legowska, A., Debowski, D., Lukajtis, R., Wysocka, M., Czaplewski, C., Lesner, A., and Rolka, K. (2010) Implication of the disulfide bridge in trypsin inhibitor SFTI-1 in its interaction with serine proteinases. *Bioorg. Med. Chem.* **18**, 8188-8193
20. Li, P., Jiang, S., Lee, S.-L., Lin, C.Y., Johnson, M.D., Dickson, R.B., Michejda, C.J., and Roller, P.P. (2007) Design and synthesis of novel and potent inhibitors of the type II transmembrane serine protease, matriptase, based upon the sunflower trypsin inhibitor-1. *J. Med. Chem.* **50**, 5976-5983
21. Guo, X., Shi, J., Tang, Z., Cui, D., and Zhang, Y. (2006) Synthesis and biological activity of seleno sunflower trypsin inhibitor analog. *Chem. Biol. Drug Des.* **68**, 341-344
22. Avrutina, O., Fittler, H., Glotzbach, B., Kolmar, H., and Empting, M. (2012) Between two worlds: a comparative study on in vitro and in silico inhibition of trypsin and matriptase by redox-stable SFTI-1 variants at near physiological pH. *Org. Biomol. Chem.* **10**, 7753-7762
23. Barbeta, B.L., Marshall, A.T., Gillon, A.D., Craik, D.J., and Anderson, M.A. (2008) Plant cyclotides disrupt epithelial cells in the midgut of lepidopteran larvae. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1221-1225
24. Gustafson, K.R., McKee, T.C., and Bokesch, H.R. (2004) Anti-HIV cyclotides. *Curr. Protein Pept. Sci.* **5**, 331-340
25. Daly, N.L., Love, S., Alewood, P.F., and Craik, D.J. (1999) Chemical synthesis and folding pathways of large cyclic polypeptides: studies of the cystine knot polypeptide kalata B1. *Biochemistry* **38**, 10606-10614
26. Hernandez, J.F., Gagnon, J., Chiche, L., Nguyen, T.M., Andrieu, J.P., Heitz, A., Trinh Hong, T., Pham, T.T., and Le Nguyen, D. (2000) Squash trypsin inhibitors from *Momordica cochinchinensis* exhibit an atypical macrocyclic structure. *Biochemistry* **39**, 5722-5730
27. Daly, N.L., Chen, Y.-K., Foley, F.M., Bansal, P.S., Bharathi, R., Clark, R.J., Sommerhoff, C.P., and Craik, D.J. (2006) The absolute structural requirement for a proline in the P3'-position of Bowman-Birk protease inhibitors is surmounted in the minimized SFTI-1 scaffold. *J. Biol. Chem.* **281**, 23668-23675
28. Chan, L.Y., Gunasekera, S., Henriques, S.T., Worth, N.F., Le, S.J., Clark, R.J., Campbell, J.H., Craik, D.J., and Daly, N.L. (2011) Engineering pro-angiogenic peptides using stable, disulfide-rich cyclic scaffolds. *Blood* **118**, 6709-6717
29. Goddard, T.D., and Kneller, D.G. SPARKY 3. *University of California, San Francisco*
30. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley-Interscience, New York
31. Morrison, J.F. (1969) Kinetics of the reversible inhibition of enzyme-catalysed reactions by tight-binding inhibitors. *Biochim. Biophys. Acta* **185**, 269-286
32. Guntert, P. (2009) Automated structure determination from NMR spectra. *Eur. Biophys. J.* **38**, 129-143
33. Koradi, R., Billeter, M., and Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 29-32
34. Yuan, C., Chen, L., Meehan, E.J., Daly, N., Craik, D.J., Huang, M., and Ngo, J.C. (2011) Structure of catalytic domain of Matriptase in complex with Sunflower trypsin inhibitor-1. *BMC Struct. Biol.* **11**, 30
35. Sali, A., and Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779-815
36. Shen, M.Y., and Sali, A. (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **15**, 2507-2524
37. Case, D.A., Cheatham, T.E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K.M., Jr., Onufriev, A., Simmerling, C., Wang, B., and Woods, R.J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668-1688
38. Cerutti, D.S., Le Trong, I., Stenkamp, R.E., and Lybrand, T.P. (2009) Dynamics of the streptavidin-biotin complex in solution and in its crystal lattice: distinct behavior revealed by molecular simulations. *J. Phys. Chem. B* **113**, 6971-6985
39. Ryckaert, J.-P., Ciccotti, G., and Berendsen, H.J. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Physics* **23**, 327-341

40. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An N.log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089-10092.
41. Tam, J.P., and Lu, Y.-A. (1998) A biomimetic strategy in the synthesis and fragmentation of cyclic protein. *Protein Sci.* **7**, 1583-1592
42. Avrutina, O., Schmoldt, H.U., Gabrijelcic-Geiger, D., Le Nguyen, D., Sommerhoff, C.P., Diederichsen, U., and Kolmar, H. (2005) Trypsin inhibition by macrocyclic and open-chain variants of the squash inhibitor MCoTI-II. *Biol. Chem.* **386**, 1301-1306
43. Wong, C.T., Rowlands, D.K., Wong, C.H., Lo, T.W., Nguyen, G.K., Li, H.Y., and Tam, J.P. (2012) Orally active peptidic bradykinin B1 receptor antagonists engineered from a cyclotide scaffold for inflammatory pain treatment. *Angew. Chem. Int. Edit.* **51**, 5620-5624
44. Takeuchi, T., Harris, J.L., Huang, W., Yan, K.W., Coughlin, S.R., and Craik, C.S. (2000) Cellular localization of membrane-type serine protease 1 and identification of protease-activated receptor-2 and single-chain urokinase-type plasminogen activator as substrates. *J. Biol. Chem.* **275**, 26333-26342
45. Rafal, L., Anna, L., Magdalena, W., Dawid, D., Adam, L., and Krzysztof, R. (2011) Analogues of trypsin inhibitor SFTI-1 modified in the conserved P(1)' position by synthetic or non-proteinogenic amino acids retain their inhibitory activity. *J. Pept. Sci.* **17**, 281-287
46. Fittler, H., Avrutina, O., Glotzbach, B., Empting, M., and Kolmar, H. (2013) Combinatorial tuning of peptidic drug candidates: high-affinity matriptase inhibitors through incremental structure-guided optimization. *Org. Biomol. Chem.*
47. Mylne, J.S., Colgrave, M.L., Daly, N.L., Chanson, A.H., Elliott, A.G., McCallum, E.J., Jones, A., and Craik, D.J. (2011) Albumins and their processing machinery are hijacked for cyclic peptides in sunflower. *Nat. Chem. Biol.* **7**, 257-259
48. Clark, R.J., Jensen, J., Nevin, S.T., Callaghan, B.P., Adams, D.J., and Craik, D.J. (2010) The engineering of an orally active conotoxin for the treatment of neuropathic pain. *Angew. Chem. Int. Edit.* **49**, 6545-6548
49. Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10037-10041
50. Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S., and Sykes, B.D. (1995) ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J. Biomol. NMR* **5**, 67-81

Acknowledgements- This work was supported by a Queensland Government Smart State Fellowship awarded to NLD, which was co-funded by Hexima Limited, and by project SO 249/1-1, priority program 1394 'Mast cells – promoters of health and modulator of diseases', of the Deutsche Forschungsgemeinschaft to CPS. DJC is a National Health and Medical Research Council Professorial Research Fellow. NLD is an Australian Research Council Future Fellow. PQ was supported by an Australian Postgraduate Award. We thank Phillip Walsh and Philip Sunderland for assistance with peptide synthesis and Sabine Streicher for kinetic measurements.

Abbreviations: 3D, three-dimensional; COSY, correlation spectroscopy; MCoTI-II, Momordica cochinchinensis trypsin inhibitor II; MD, molecular dynamics; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; PAM, phenylacetamidomethyl; RMSD, root mean square deviation; SEM, standard error of the mean; SFTI-1, sunflower trypsin inhibitor-1; RP-HPLC, reverse phase high performance liquid chromatography; TFA, trifluoroacetic acid; TOCSY, total correlation spectroscopy.

FIGURE LEGENDS

Figure 1: Comparison of trypsin and matriptase active sites. The crystallographic structures of SFTI-1 in complex with trypsin (pdb identifier 1sfi) and in complex with matriptase (pdb identifier 3p8f) are represented. The protein backbone of SFTI-1 is in cyan and the functional Lys is displayed in stick representation. The eight loops that surround the active sites are named loops I to VIII. The solvent accessible surfaces of the proteases are colored according to the Poisson-Boltzmann electrostatic potential they generate, as computed by the APBS software (49), with a scale ranging from -5 kT/e (red) to +5 kT/e (blue). The charged positions in each loop are highlighted. The numberings of trypsin and matriptase positions are according to the sequence of bovin trypsin (UniProt P00760) and human matriptase SP1 (Q9Y5Y6), respectively. The figure was generated using PyMol.

Figure 2: Comparison of the sequences and structures of SFTI-1 and MCoTI-II. The structures were determined based on NMR spectroscopy (pdb identifier 1jbl and 1ib9, respectively) and drawn using MolMol.

Figure 3: A) Comparison of the secondary shifts of selected SFTI-1 and MCoTI-II mutants. The shifts were generated by subtracting the αH shifts from random coil shifts (50). B) Alignment of NMR structures of SFTI-1 backbone with the SFTI-1 variants I10R (left) and R2A (right). The parent peptide SFTI-1 is displayed in pale cyan. The variants I10R and R02A are displayed in purple and green respectively.

Figure 4: Comparison of the differences in inhibitory activity of the SFTI-1 (A) and MCoTI-II (B) mutants relative to the native peptides. The side-chains are highlighted on the structures shown on the right of the diagram. The disulfide-bonds are shown as yellow sticks. The surface diagrams were generated using PyMol.

Figure 5: A) Binding modes of SFTI-1 (cyan) and MCoTI-II (green) in the active sites of trypsin (left) and matriptase (right). The three major differences between the active sites of trypsin and matriptase are the longest and more charged loop II of matriptase, the different conformation of loop III and the more negatively charged loop IV of matriptase. The structures in this figure are the final conformations of 20 ns molecular dynamics simulations carried out for each complex. SFTI-1 and MCoTI-II adopted similar binding modes in the two protease active sites, but MCoTI-II displayed more conformational variability than SFTI-1. The solvent accessible surfaces of the proteases are colored according to the Poisson-Boltzmann electrostatic potential they generate, as computed by the APBS software (49), with a scale ranging from -5 kT/e (red) to +5 kT/e (blue). B) Comparison of the binding modes of the inhibitory loops of SFTI-1 and MCoTI-II when they are in complex with matriptase and trypsin. The backbones of the proteases are shown in white using cartoon representations, and the inhibitory loops are in cartoon and stick representations. The proteases were

superimposed using PyMol. The remaining parts of the SFTI-1 and MCoTI-II beside the inhibitory loops are not shown.

Figure 6: Comparison of the binding modes of six SFTI-1 variants modeled in the matriptase and trypsin active sites. The models of the variants were generated by comparison using the structure of the complexes with wild-type SFTI-1 and were simulated by molecular dynamics for at least 5 ns. The proteases in the mutated complex are shown in green using cartoon and stick representations, and the SFTI-1 variants are shown in cyan using cartoon and stick representations. The structures of the corresponding complexes but without mutations are shown in transparency with the backbone of SFTI-1 in violet and the proteases in white. The mutated position is circled, and residues and loops discussed in the text are highlighted.

Figure 7: Comparison of complexes involving matriptase by focusing around position 2 of SFTI-1 (A), position 3 of MCoTI-II (B), and position 3 of MCoTI-II V3R (C). These three positions occupy equivalent coordinates in the matriptase active site. SFTI-1, MCoTI-II and [V3R]MCoTI-II are shown in cyan, and matriptase is in green. Positions discussed in the text are highlighted. The represented structures are the final conformations from molecular dynamics simulations.

Table 1. Equilibrium dissociation constant K_i for the inhibition of trypsin and matriptase by SFTI-1, MCoTI-II, and mutants.

Inhibitor name	K_i Trypsin (nM)	K_i Matriptase (nM)
SFTI-1 native	0.0017 ± 0.00026	200 ± 22
SFTI-1 alanine mutants		
G1A	0.0048 ± 0.00032	190 ± 26
R2A	0.16 ± 0.016	~10,000
T4A	0.16 ± 0.016	1500 ± 170
K5A	>1000	>10,000
S6A	0.15 ± 0.018	1600 ± 390
I7A	0.89 ± 0.21	84 ± 15
P8A	0.035 ± 0.0044	27 ± 1.4
P9A	0.0017 ± 0.0001	370 ± 58
I10A	0.086 ± 0.016	73 ± 11
F12A	0.21 ± 0.031	320 ± 35
P13A	0.0037 ± 0.0003	240 ± 37
D14A	0.01 ± 0.0014	1500 ± 150
SFTI-1 additional mutants		
R2K	0.002 ± 0.00003	1,200 ± 160
K5R	0.0027 ± 0.00078	310 ± 52
I7R	0.01 ± 0.00095	4500 ± 500
I10D	0.051 ± 0.006	~10,000
I10G	0.081 ± 0.011	3,700 ± 490
I10R	0.0038 ± 0.00062	6.4 ± 1.3
I10K	0.0057 ± 0.0015	40 ± 5.8
I7A+I10R	1.2 ± 0.25	51 ± 4.9
MCoTI-II native	0.0023 ± 0.0007	2.8 ± 0.51
MCoTI-II alanine mutants		
G02A	0.38 ± 0.043	180 ± 17
V03A	0.15 ± 0.028	2.3 ± 0.12
P05A	0.056 ± 0.0076	39 ± 5.4
K06A	>1,000	>10,000
I07A	1.2 ± 0.22	9.8 ± 1.9
L08A	0.13 ± 0.011	12 ± 1.9
K09A	0.0034 ± 0.0048	76 ± 14
K10A	0.051 ± 0.0095	4.1 ± 1
N26A	0.008 ± 0.0027	12 ± 0.41
Y28A	0.15 ± 0.021	11 ± 1.3
G30A	0.0032 ± 0.00036	3.7 ± 0.033
S31A	0.069 ± 0.0079	98 ± 10
MCoTI-II additional mutants		
V03R	0.01 ± 0.0025	0.29 ± 0.054
K6R	0.13 ± 0.0055	39 ± 4.4
I7R	5.2 ± 0.62	110 ± 8.5
V3R+I7A	2.8 ± 0.23	3.5 ± 0.32

mean ± SEM, n≥4

Table 2. Equilibrium dissociation constant K_i for the inhibition of trypsin and matriptase by SFTI-1 and MCoTI-II hybrids.

Peptide	**Sequence**	**Inhibition constant**	
		Trypsin (nM)	**Matriptase (nM)**
SFTI-1	GRCTKSIPPICFPD	0.0017 ± 00026	200 ± 22
MCoTI-II	GGVCPKILKKCRRDSDCPGACICRGNGY--CGSGSD	0.0023 ± 0.0007	2.8 ± 0.51
SFMC	GRC**PKILKK**CFPD	0.0023 ± 0.00018	290 ± 34
MCoTI-II-S[1]	GGVC**TKSIPP**CRRDSDCPGACICRGNGY--CGSGSD	450 ± 27	3,500 ± 660
MCoTI-II-S[5]	GGVCP**A**ILKKCRRDSDCPGACIC**TKSIPPI**CGSGSD	68 ± 4.5	>10,000

Table 3: Structural statistics for SFTI-1 mutant structures

	I10R	R2A
Experimental restraints		
Interproton distances	95	88
Intraresidue	28	28
Sequential	33	36
Medium range, $i - j < 5$	9	3
Long range, $i - j \geq 5$	25	21
Disulfide bond restraints	2	2
Dihedral bond (ϕ, ψ, χ_1) restraints	6	8
RMSD from mean coordinate structure (Å)		
Backbone atoms (residues 1–14)	0.19 ± 0.11	0.20 ± 0.09
All heavy atoms (residues 1–14)	1.05 ± 0.19	0.70 ± 0.09
Stereochemical quality		
Residues in most favoured Ramachandran region (%)	87.5	85.8
Ramachandran outliers (%)	12.5	14.2
Overall MolProbity score[1]	1.76 ± 0.32	2.15 ± 0.47

[1]Determined using MolProbity

Table 4: Average buried surface area in $Å^2$ of SFTI-1 wild-type and variants in their complex with matriptase and trypsin.

	Complex with trypsin	**Complex with matriptase**
SFTI-1 wild-type	*1340 +/- 49*	*1483 +/- 55*
[I10A]SFTI-1	1465 +/- 52	1476 +/- 54
[I10D]SFTI-1	1413 +/- 60	1431 +/- 62
[I10K]SFTI-1	1384 +/- 59	1417 +/- 55
[I10R]SFTI-1	1407 +/- 64	1538 +/- 81
[I7A]SFTI-1	1295 +/- 44	1428 +/- 52
[I7A+I10R]SFTI-1	1316 +/- 58	1414 +/- 66
[R2A]SFTI-1	1274 +/- 62	1275 +/- 73
MCoTI-II	*1550 +/- 64*	*1971 +/- 124*
[V3R]MCoTI-II	1637 +/- 97	2152 +/- 62
[I7A]MCoTI-II	1396 +/- 56	2021 +/- 73
[V3R+I7A]MCoTI-II	1632 +/- 70	2081 +/- 57

The buried surface areas were computed over the last 15 ns of the simulations involving wild-type SFTI-1 and over the last 4 ns of the simulations involving SFTI-1 variants using the LCPO approximation implemented in the ccptraj software from the Amber12 package. Standard-deviations are provided for each measure.

Figure 1

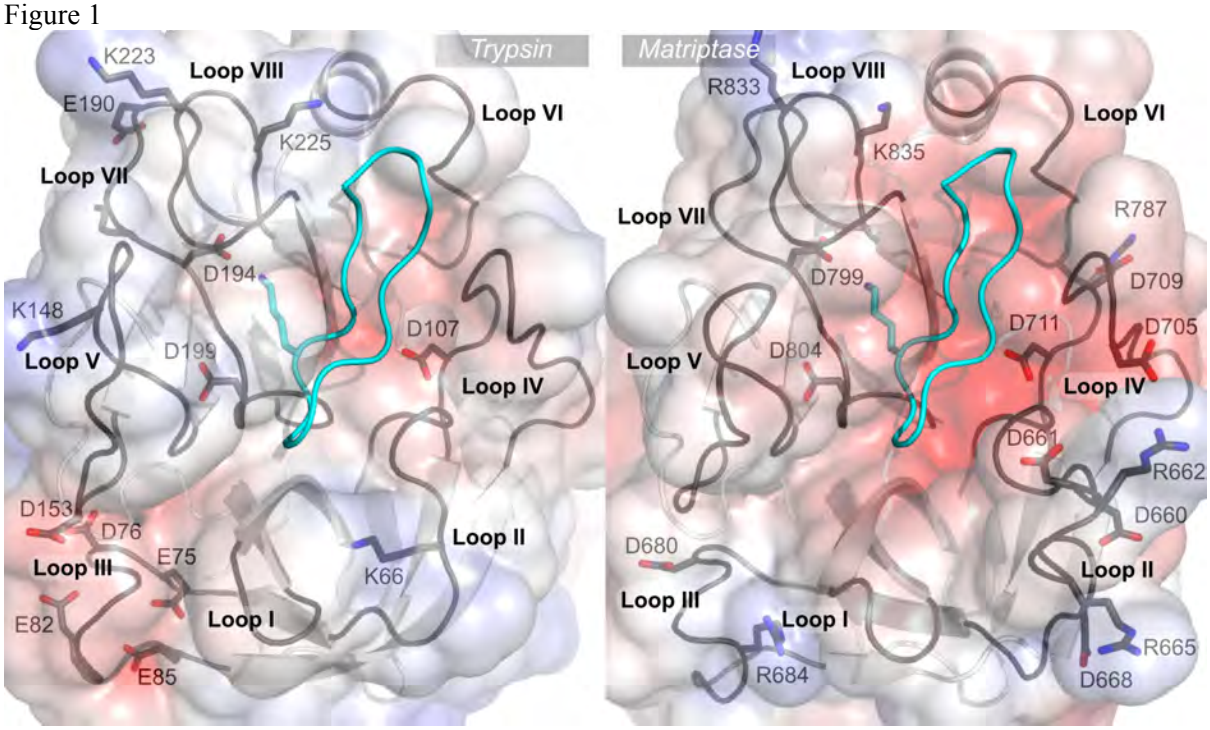

Figure 2

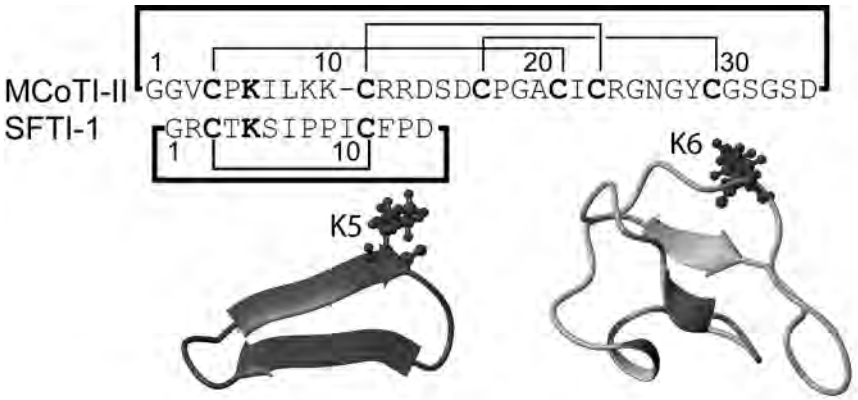

Figure 3

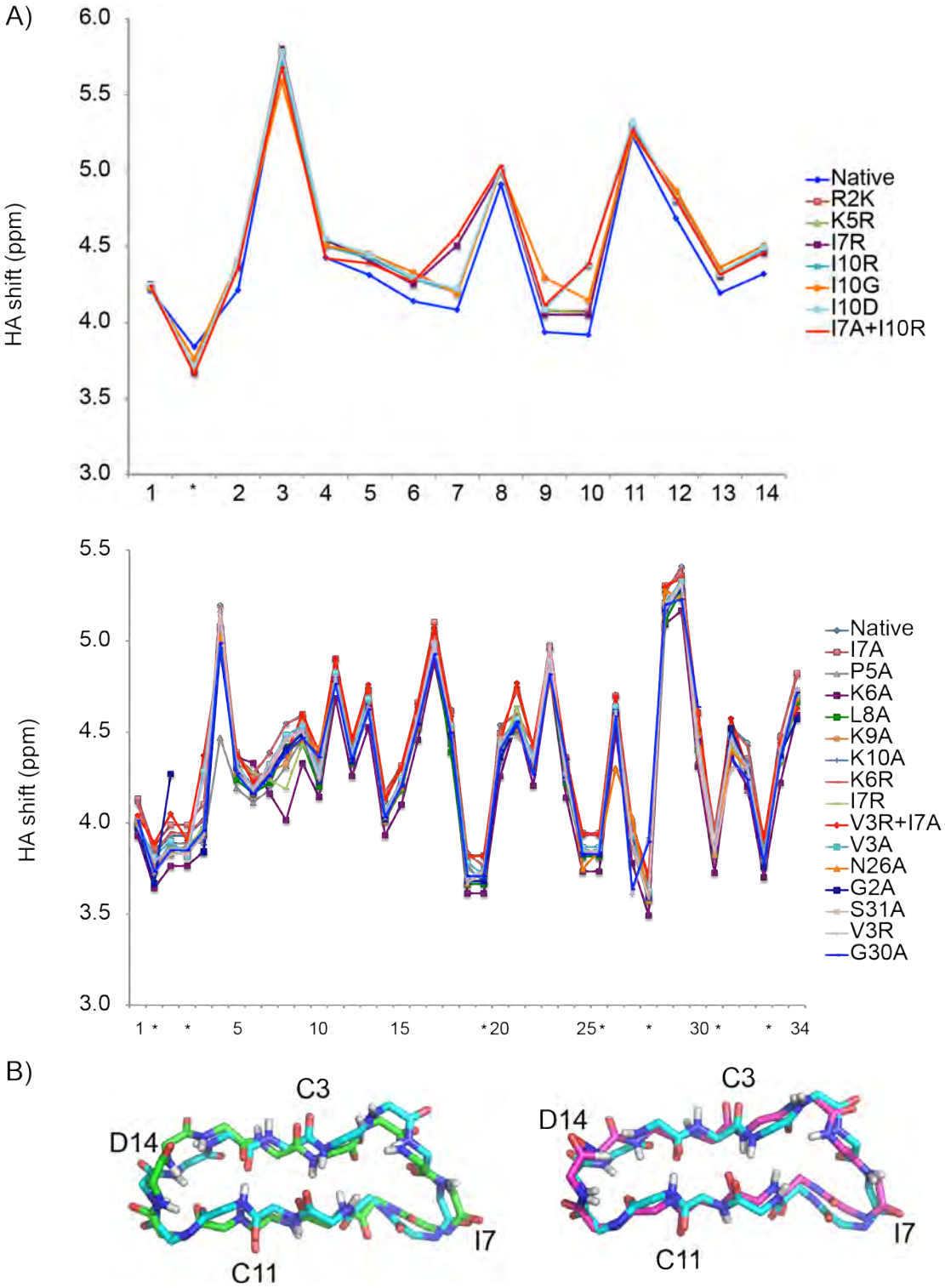

Figure 4

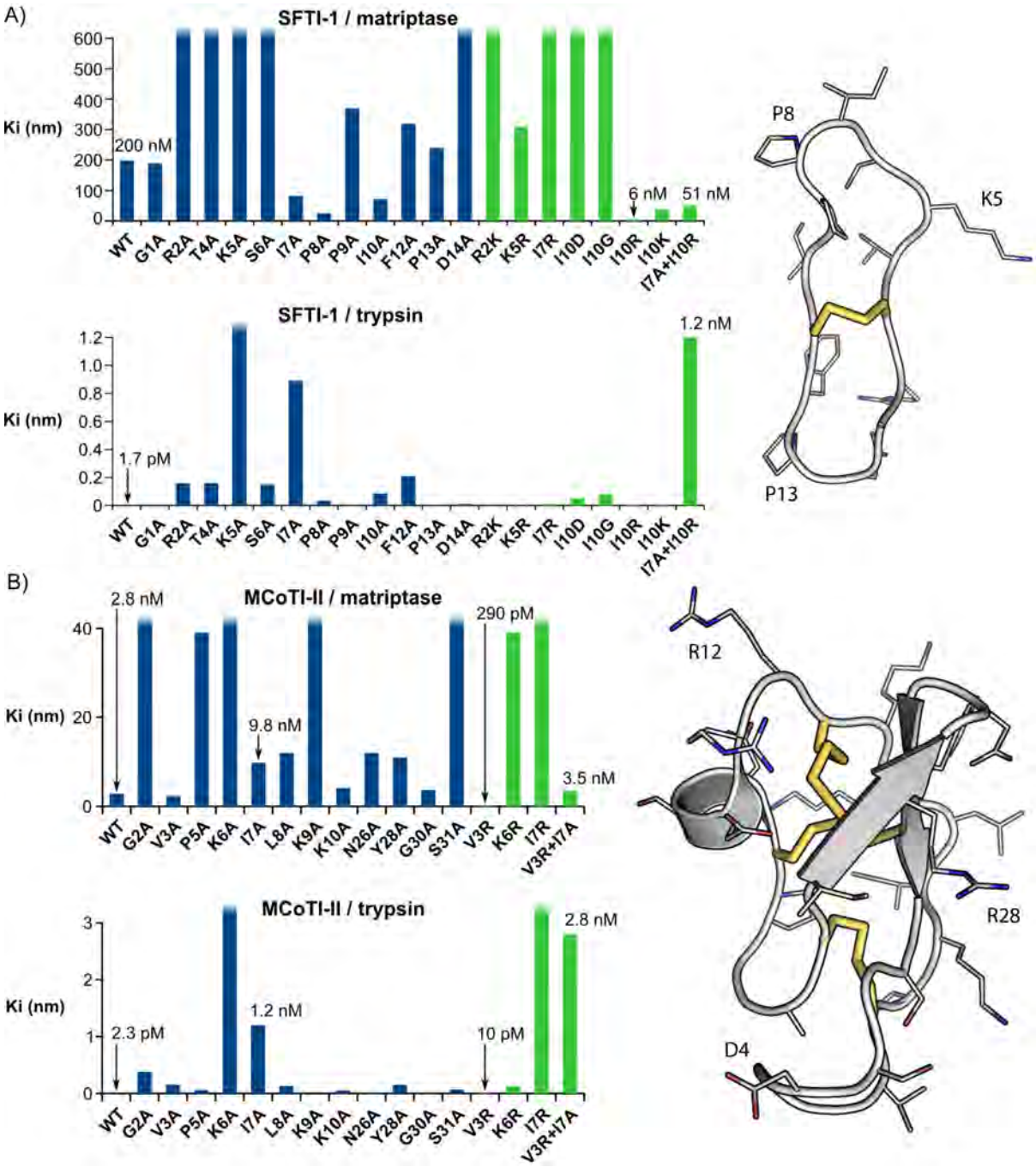

Figure 5

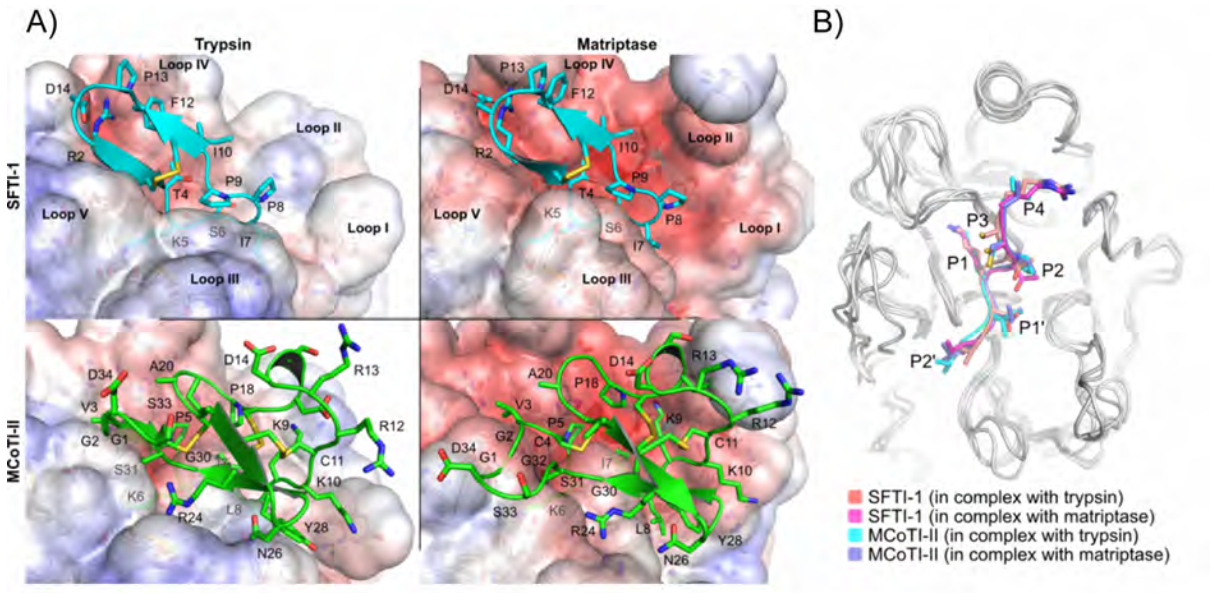

Figure 6

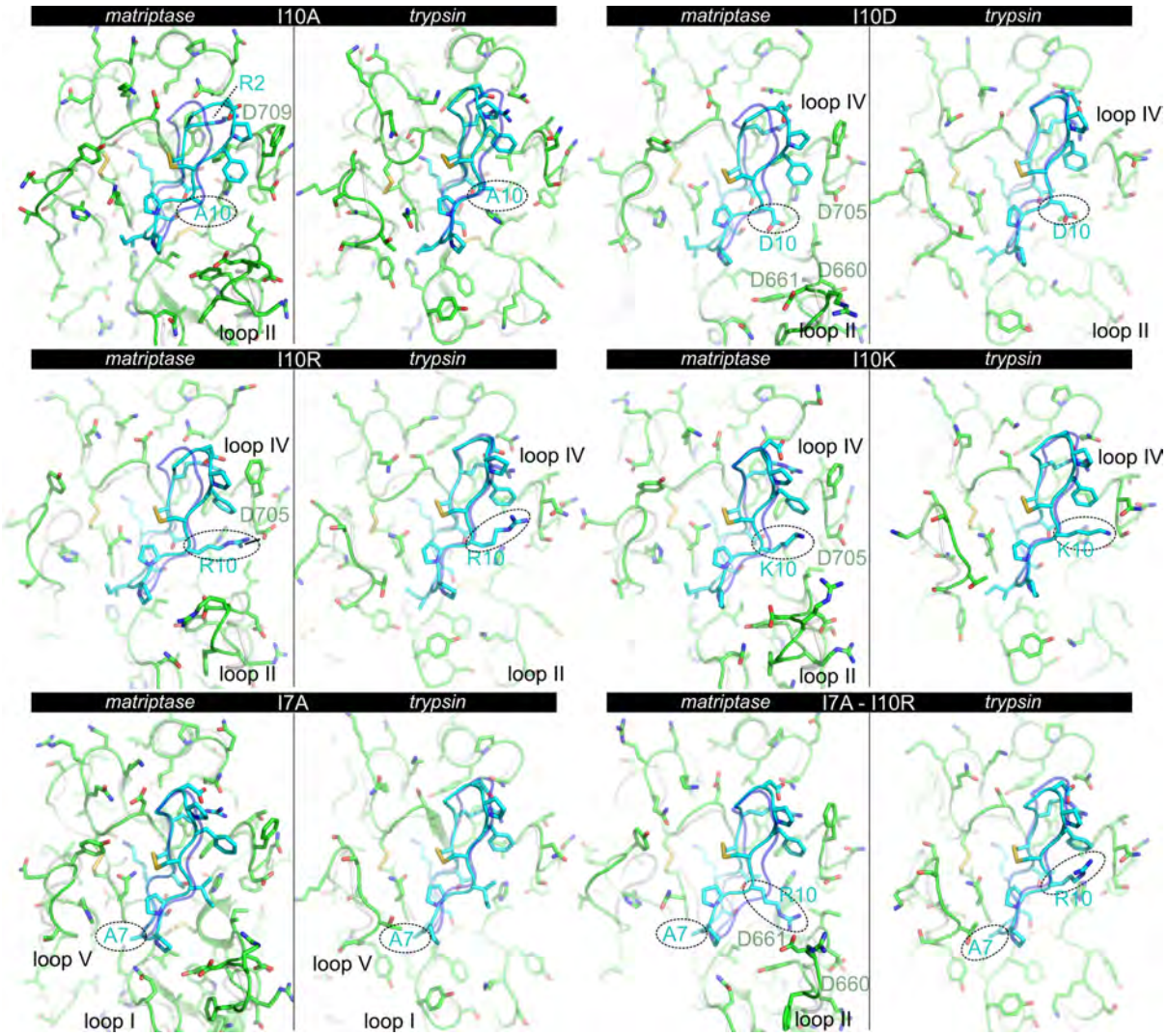

Figure 7