IMPROVING THE SELECTIVITY OF ENGINEERED PROTEASE INHIBITORS: OPTIMIZING THE P2 PRIME RESIDUE USING A VERSATILE CYCLIC PEPTIDE LIBRARY

Simon J. de Veer[†], Conan K. Wang[‡], Jonathan M. Harris[†], David J. Craik[‡], Joakim E. Swedberg[‡,*]

[†]Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane QLD 4059, Australia

[‡]Institute for Molecular Bioscience, The University of Queensland, Brisbane QLD 4072, Australia

*Corresponding author: Dr Joakim Swedberg, Institute for Molecular Bioscience, The University of Queensland, Brisbane QLD 4072, Australia. Email: j.swedberg@imb.uq.edu.au

ABSTRACT

Laskowski inhibitors are receiving growing interest as engineering templates for designing reversible serine protease inhibitors. When optimizing binding interactions between the inhibitor and the target protease, many studies focus on the non-primed segment of the inhibitor's binding loop (encompassing the contact β-strand). However, there are currently few methods for screening residues on the primed segment. In this study, we designed a versatile, synthetic inhibitor library (based on the cyclic peptide, sunflower trypsin inhibitor-1) for characterizing the binding specificity of various proteases at the P2 prime (P2′) position. Screening the library against thirteen different serine proteases, including trypsin, chymotrypsin, matriptase, plasmin, thrombin, four kallikrein-related peptidases, and several clotting factors, revealed that each displayed unique P2′ preferences. Further, using this information to modify the P2′ residue in several existing engineered inhibitors yielded new variants that showed considerably improved selectivity, reaching up to 7,000-fold selectivity over certain off-target proteases. Our study demonstrates the importance of the P2′ residue in Laskowski mechanism inhibition, and unveils a new method for screening P2′ substitutions that will benefit future inhibitor engineering studies.

INTRODUCTION

Serine proteases are established therapeutic targets in a spectrum of medical conditions, including coagulation and cardiovascular diseases, metabolic disorders, inflammation, and infectious diseases[1]. Accordingly, the pharmaceutical industry has had a longstanding interest in developing therapeutic molecules that can modulate the activity of one or more proteases, with most attention directed towards protease inhibitors[2]. These drug discovery programs have provided a number of major successes, and an array of inhibitors targeting different proteases are currently used in the clinic[2], including several new drugs on the market[3]. Traditionally, protease-based drug development has favored classical small molecule inhibitors as this approach often yields suitable lead compounds that have excellent pharmacokinetic properties. However, small molecule inhibitors typically have relatively few binding determinants, which can lead to challenges when translating drug leads from the laboratory to the clinic. First, the risk of limited selectivity is often higher for inhibitors that have fewer binding determinants. Off-target inhibition can cause negative side-effects in patients, and was regarded to be a contributing factor in the failure of first generation matrix-metalloproteinase inhibitors in clinical trials for preventing cancer metastasis[4]. Second, the emergence of drug resistance is more likely for inhibitors that have fewer binding determinants as only a small number of mutations are required for the target protein to evade drug binding. This is particularly relevant to therapies for rapidly evolving diseases, and is an ongoing concern for treating HIV[5] and hepatitis C[6] infections. To overcome these challenges, there is now growing interest in engineering naturally occurring peptide and small protein inhibitors that interact via larger binding surfaces and are capable of greater selectivity[7].

Laskowski (or standard mechanism) inhibitors[8, 9] are a major class of serine protease inhibitors that fall between small molecule drugs and large protein biologics (in terms of molecular weight). These inhibitors are found in all forms of life, and operate by an intriguing mechanism that involves the protease cleaving then re-synthesizing the inhibitor's reactive site bond[10, 11]. Despite sharing a common mode of action, Laskowski inhibitors show immense variety in sequence and structure, and segregate into at least 19

different families that display unique characteristics[9]. This is an advantage for drug development, as it provides a series of engineering templates that function in the same way but are structurally diverse, ranging in size from circular peptides (as small as 1.5 kDa) to larger protein domains (6-8 kDa). Indeed, only one structural feature is strictly conserved across all inhibitor families: the conformation of the inhibitor's binding loop (termed the canonical loop)[8, 12]. This segment is pre-configured to match the geometry of the serine protease active site, and displays a variable sequence that is complementary to the substrate specificity profile of intended protease targets (spanning both primed (P′) and non-primed (P) binding subsites: Schechter-Berger nomenclature[13], Figure 1A).

As with many protein-protein interactions, the sequence of the inhibitor's binding loop has a major role in determining which proteases the inhibitor will be effective against. Therefore, engineering the sequence of the canonical loop is one of the main strategies for directing Laskowski inhibitors to new protease targets. New sequence combinations can be identified by several different methods, with those capable of screening both the non-primed and primed residues generally providing greater success. To date, the most widely used approach has been phage display, which is one of only a few methods where all residues of the canonical loop are accessible for optimization. Phage display was first applied to select potent inhibitors of neutrophil elastase[14], then later used to develop the first engineered Laskowski inhibitor to reach the clinic, Ecallantide. This drug is a selective plasma kallikrein inhibitor based on the first Kunitz domain of tissue factor pathway inhibitor[15] and is used to relieve inflammation in hereditary angioedema[16].

Alternatively, the template inhibitor can be engineered by rational approaches that focus on substituting a pre-determined variant sequence into the binding loop. This strategy has been successfully used to engineer novel inhibitors for furin[17], HCV NS3 protease[18] and coagulation proteases[19] from diverse Laskowski inhibitor templates. More recently, synthetic approaches have been developed to identify new binding sequences by screening combinatorial[20] or non-combinatorial[21] libraries of tetrapeptide substrates to characterize the non-primed specificity of the target protease. We have previously used this strategy to

guide the design of inhibitor variants for several different members of the kallikrein-related peptidase (KLK) family, including KLK4[21, 22] and KLK14[23, 24]. The success of the substrate-guided approach derives from observations that for all Laskowski inhibitors, the non-primed residues of the canonical loop interact with the protease to form a β-sheet (spanning from P1 to at least P3) that replicates the binding of a conventional substrate (Figure 1A-B)[9, 25]. As different proteases usually display a unique non-primed specificity profile[20], modifying the sequence of the inhibitor's contact β-strand often leads to marked changes in its potency and selectivity. Serine proteases also display diversity across their primed binding sites that can be targeted in inhibitor engineering studies. Moreover, for Laskowski inhibitors, the primed segment is involved in interactions that are specific to the inhibition mechanism, including the religation reaction, that are not taken into account when screening substrates. Consequently, there is a need to develop new, broadly-applicable strategies for screening the primed residues of the canonical loop to increase the coverage of chemistry-based methods for optimizing the binding sequence.

In this study, we developed a versatile, inhibitor-focused approach for screening preferences at a critical residue in the primed segment, P2′. This strategy was based on a novel, broad-range protease inhibitor that we recently produced by engineering the fourteen amino acid cyclic peptide, sunflower trypsin inhibitor-1 (SFTI-1, Figure 1C-D)[24]. Using the engineered broad-range inhibitor as a starting point, we designed and synthesized a library of twenty P2′ variants to screen against diverse serine proteases. The library screen revealed unique P2′ preferences for thirteen different proteases as well as an overall preference for Ile, which is present at P2′ in many naturally occurring inhibitors, including SFTI-1. Additionally, we used the newly identified P2′ preferences to modify known matriptase, KLK14 and thrombin inhibitors, generating new variants that displayed improved selectivity. Overall, our study presents a versatile cyclic peptide library that makes it possible to rapidly screen P2′ preferences for diverse serine proteases, and demonstrates that substituting the P2′ residue in Laskowski inhibitors, such as SFTI-1, is effective for modulating the selectivity of engineered inhibitor variants.

RESULTS

Design, synthesis and characterization of a P2′ diverse inhibitor library. To devise a new screening method for optimizing interactions between the primed segment of Laskowski inhibitors and various protease targets, we first studied structures of diverse protease/inhibitor complexes. Examining residues in the primed segment (P1′-P4′) revealed that interactions formed by the P1′, P3′ and P4′ residues varied between intramolecular and intermolecular interactions across different inhibitor families. However, the P2′ residue consistently extended out from the inhibitor core and engaged a well-defined pocket (S2′) within the protease active site (Figure 1E-F). This suggested that the P2′ residue was an ideal target for substitution with the aim of changing the inhibitor's potency and selectivity. Therefore, we set out to characterize the P2′ specificity of different proteases by designing a synthetic inhibitor library based on the naturally occurring cyclic peptide, SFTI-1 (compound **1**, Table 1). With only fourteen amino acids, SFTI-1 (and engineered SFTI variants) can be readily produced by solid phase peptide synthesis, including inhibitors containing non-natural amino acids[26, 27]. Additionally, SFTI-1 has relatively simple architecture, comprising a canonical loop with minimal scaffolding, and thus is an excellent model system for examining how various P2′ substitutions affect the inhibitor's activity against different protease targets.

The SFTI variant that we selected for generating the inhibitor library was a broad-range variant (**2I**, compound **2** with Ile at P2′) that we recently engineered[24], where residues 2, 12 and 14 of SFTI-1 were substituted to re-configure the inhibitor's intramolecular hydrogen bond network, and P1 Lys was replaced with Arg (Figure 2A). These replacements gave rise to an inhibitor that forms minimal side chain interactions with the protease's non-primed binding sites (except for the S1 subsite) as the P2 and P4 residues participate in the inhibitor's hydrogen bond network, and the P3 residue forms a disulfide bond with Cys_{11}. As **2I** has relatively few selectivity determinants but a prominent hydrogen bond network, it effectively inhibits a number of proteases, including trypsin and several members of the kallikrein-related peptidase family[24]. Using **2I** as a template, we produced a library of twenty individually

Comment [A1]: Primed is technically correct but sounds a bit discordant. Perhaps use " examining residues on the prime side of the scissile bond"?

synthesized inhibitors where the P2′ residue (Ile7) was substituted (Figure 2B) with each of the naturally occurring amino acids (excluding Cys to prevent formation of disulfide bond isomers), or biphenylalanine (BiP, single letter code: B) as an example of a non-natural amino acid.

Peptides were synthesized by standard Fmoc synthesis on acid labile 2-chlorotrityl chloride resin. Protected peptides were liberated from the solid support before cyclization in solution using HATU (1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate) as the activator. After protecting group removal and an initial purification by reverse phase HPLC, disulfide bonds were formed before further reverse phase HPLC purification to achieve purities above 95% as determined by analytical UPLC (Table 1 and Figure S1, Supporting Information).

The purified peptides were subsequently characterized by NMR. For SFTI-1 and **2I**, we previously showed that the inhibitor's 1D ^{1}H-NMR spectrum exhibits sharp lines and well-dispersed amide resonances that are distributed across the 7–9 ppm range with down-field shifted αH signals that are indicative of a β-sheet structure[24, 28]. Similar to that of **2I**, the 1D ^{1}H-NMR spectra of the P2′ variants showed sharp lines and well-dispersed signals (Figure S2, Supporting Information), indicating that they were well-folded. In addition, the number of signals (from backbone amides and side chain protons) present in the 7–9 ppm region of the spectra suggested that the peptides adopt one main conformation in solution. Of all the variants, **2P** displayed the greatest conformational flexibility, with the presence of a set of additional low intensity signals in its 1D ^{1}H-NMR spectrum that suggest the presence of at least one other conformation; this is expected as the replacement of an Ile with the conformationally restricted Pro would likely have some effect on the backbone structure. In general though, the NMR data indicate that the SFTI-1 variants were well-folded and rigid.

Comment [A2]: Folding usually refers to tertiary structure – perhaps "geometrically stable" would be better here?

Comment [A3]: Lot of use of "conformation" here. Maybe replace one with "arrangement"

Exploring the versatility of the inhibitor library against diverse serine proteases. The first proteases screened against the P2′ inhibitor library were the classical serine proteases, trypsin and chymotrypsin. In a previous study, we showed that **2I** was a potent inhibitor of trypsin (K_i = 0.7 nM)[24], and thus each of the twenty inhibitor variants was assayed against trypsin at a low nanomolar range (25 nM). Examining

trends across the library revealed that substituting the P2′ residue led to marked changes in the inhibitor's activity against trypsin (Figure 3). The strongest inhibition was provided by branched hydrophobic residues, Ile (80% inhibition) and Val (75%), followed by Arg (65%) and Leu (52%). By contrast, variants containing Gly or Pro showed essentially no activity whereas acidic residues, Asp and Glu, also generated poor inhibitors. These findings were largely consistent with a previous study that used a minimized Bowman-Birk loop peptide (similar to residues 3-11 of the peptides in this study) to examine the effect of P2′ substitutions on trypsin inhibition[29], although the overall level of inhibition in our study was several orders of magnitude higher.

The inhibitor library was also able to provide insight into the P2′ specificity of chymotrypsin, even though all library peptides contained Arg as the P1 residue which is not favored by chymotrypsin[20]. While assays with chymotrypsin required much higher inhibitor concentrations (25 μM), the data generated was useful as the purpose of this screen was to compare the relative activity of each variant at the same concentration. Overall, the P2′ residues that provided the strongest inhibition were similar to trypsin, including hydrophobic residues, Val (76% inhibition) and Ile (60%), and Arg (57%) (Figure 3). However, aromatic residues were also strongly preferred, including Trp (63%) and Phe (52%). To explore the structural basis for these differences in the P2′ specificity profiles for chymotrypsin and trypsin, we used molecular dynamics (MD) simulations to examine binding interactions in various protease/inhibitor complexes. These analyses revealed that the S2′ pocket of chymotrypsin is larger than trypsin, allowing bulky P2′ residues (Trp) to be accommodated in the space left by Thr151 (Figure S3A), Supporting Information). By contrast, trypsin contains Tyr at residue 151, which is able to form cation-π interactions with P2′ Arg (Figure S3B), Supporting Information), but leaves less space for larger resides, including Trp.

To further test the versatility of the inhibitor library, we screened the catalytic domain of matriptase, a type II transmembrane serine protease. Matriptase attracts significant interest as a therapeutic target in many epithelial cancers[30], and has been the focus of previous inhibitor design studies, including several

based on SFTI-1[26, 27, 31]. Across the P2′ inhibitor library, three variants out-performed the remaining peptides, namely Val (54% inhibition), Ile (41%) and Asn (37%) (Figure 3). Additionally, Asp (29%) appeared to be better tolerated by matriptase than seen previously for trypsin and chymotrypsin (comparing Asp to the optimal residue for each protease). Examining interactions between matriptase and **2N** (Figure S3C, Supporting Information) indicated that Asn formed favorable interactions within the S2′ pocket by engaging His143 (His752) – chymotrypsin numbering is used throughout and where relevant, specific protease numbering is given in brackets. In **2D**, Asp formed similar interactions with His143 (His752). Collectively, these findings demonstrate that the P2′ inhibitor library is effective against diverse serine proteases and has the capacity to discern a unique specificity profile for individual enzymes.

Closely-related proteases from the kallikrein-related peptidase family show distinct P2′ preferences. Next, we used the P2′ inhibitor library to examine specificity divergence among four members of the largest known multi-gene serine protease family, the kallikrein-related peptidases (KLKs). KLK proteases are involved in a wide range of physiological functions[32] and are under investigation as therapeutic targets for several indications, including hormone-dependent cancers (prostate and ovarian cancers[33]) and skin diseases[34, 35]. The non-primed (P1-P4) specificity of several KLKs has been studied using combinatorial peptide libraries[36], and later by non-combinatorial library screens[21, 23] in order to guide engineering of the SFTI-1 β-strand. Optimizing the P2′ residue provides a further opportunity to improve the potency and selectivity of engineered Laskowski inhibitors targeting different KLKs.

Screening the inhibitor library against KLK4, KLK5, KLK7 and KLK14 revealed that each protease had a unique P2′ specificity profile (Figure 3). KLK4 and KLK14 displayed similar preferences, with the most effective variant being Val (92% and 86% inhibition, respectively), followed by Ile, Asn and Arg. However, KLK4 also favored aromatic residues (BiP and Phe) but not acidic residues (Glu and Asp), whereas the converse was observed for KLK14. Similarly, the most effective inhibitor variant was identical for both KLK5 and KLK7 (Asn) which provided 93% and 97% inhibition, respectively. Polar

residues were also favored by both proteases, although residues such as His, Gln and Ser provided higher inhibition for KLK7 (compared to Val and Ile), but not KLK5. By contrast, the preference for aromatic residues (BiP, Tyr, Trp and Phe) appeared more pronounced for KLK5.

We further investigated the distinct P2′ preferences observed for each kallikrein protease by performing MD simulations. These analyses revealed that the structural basis for KLK4's preference for Arg at P2′ was different to that previously seen for trypsin. Rather than interacting with an aromatic residue, Arg7 formed a salt bridge with Glu39 in the KLK4 S2′ pocket (Figure S4A, Supporting Information). Additionally, KLK5's strong preference for aromatic residues could be explained by T-oriented π interactions involving Phe150 (Figure S4B, Supporting Information). For KLK7, the most favored residue (Asn) appeared to form a hydrogen bond with the backbone carbonyl of Leu40 (Figure S4C, Supporting Information). Finally, the preference for Lys at P2′ for KLK14 seemed to involve cation-π interactions with Tyr150 (Figure S4D, Supporting Information). These findings illustrate that even closely-related proteases (from the same multi-gene family) show differences in their P2′ specificity profile that can be traced to specific points of sequence diversity in the S2′ pocket.

Profiling the P2′ specificity of proteases from the coagulation and fibrinolysis cascade. The coagulation and fibrinolysis cascade was among the first proteolytic pathways to be targeted in major drug discovery programs[2]. Dysregulation of this proteolytic cascade is a common theme in many diseases, including inflammation, cancer and cardiovascular diseases. Overactive blood coagulation (thrombosis) is the most frequent cause of myocardial infarction and stroke[37]. Additionally, both coagulation and fibrinolysis need to be controlled during surgery to reduce bleeding and prevent activation of a systemic inflammatory response[38]. The broad-spectrum Laskowski inhibitor, aprotinin, has been used to inhibit proteases that drive both coagulation and fibrinolysis, but it was discontinued in 2008 after 15 years of clinical use following a controversial safety trial[39]. Consequently, there has recently been a renewed effort to develop inhibitors that selectively target key serine proteases in the coagulation and fibrinolysis pathways.

The major direct contributor to coagulation is thrombin, which activates platelets and cleaves fibrinogen to generate fibrin[40]. Screening the P2′ inhibitor library against thrombin (Figure 3) revealed a clear preference for Tyr (62% inhibition) followed by Asn (49%) and Phe (41%). Molecular modeling indicated that the Tyr aromatic ring was accommodated by the wide and flat hydrophobic portion of the S2′ pocket, while the hydroxyl group formed hydrogen bonds with Arg73 (Arg431) and Asn143 (Asn506) (Figure S5B, Supporting Information). The actions of thrombin and the coagulation cascade are countered by the fibrinolysis pathway, where plasmin acts to disperse blood clot by degrading fibrin. We have previously studied the non-primed (P1-P4) specificity of plasmin using a non-combinatorial peptide library[41], and screening preferences at P2′ using the inhibitor library revealed that Lys (60%) was the most favored residue, followed by Ile (50%), Val (25%) and Arg (23%). Molecular modeling suggested that P2′ Lys can form a salt bridge with Glu143 (Glu685) deep within the plasmin S2′ pocket (Figure S5A, Supporting Information) while the shallow part of the pocket is hydrophobic, which suits Ile. These findings were consistent with a previous study where plasmin's preference for basic residues at P2′ was used to engineer a selective plasmin inhibitor based on tissue factor pathway inhibitor-2[42].

We were also able to use the P2′ inhibitor library to screen proteases in the coagulation cascade that lie upstream of thrombin and regulate its activation. Interestingly, factor IXa tolerated both basic and acidic residues at P2′, with the strongest inhibition provided by Glu (54%) followed by Arg and Lys (43%). This anomaly was resolved by modeling analyses, which revealed that Glu was able to form a salt bridge with Arg142 (Arg358) (Figure S5C, Supporting Information), whereas basic residues formed a different salt bridge with Asp39 (Asp249). Factor Xa has a narrow S2′ pocket and was the only protease that did not tolerate Ile (Leu and Trp were also disfavored). Rather, most acidic, aromatic or small residues were accepted, including Asp (31%) and Tyr (26%). Molecular modeling suggested P2′ Asp formed several hydrogen bonds, including a salt bridge with Arg142 (Arg366) (Figure S5D, Supporting Information), while Tyr formed a hydrogen bond with Glu39 (Glu257). The inhibitor library also revealed several differences in the P2′ specificity for factor XIa and factor XIIa. The most preferred P2′ residue for factor

XIa was Val (65%), followed by Ile (47%) and Asp (34%), whereas factor XIIa favored aromatic and branched hydrophobic residues, with the strongest inhibition provided by Trp (87%), closely followed by Ile (85%), Val (80%), BiP (80%) and Tyr (75%). Factor XIa has a narrow and hydrophobic S2′ pocket, which ideally fits Val (Figure S5E, Supporting Information), while Arg39 (Arg413) located at the far end of the pocket enables hydrogen bonding with acidic residues. By contrast, factor XIIa has a much larger S2′ pocket with Tyr150 (Tyr377) at its base, allowing for aromatic ring interactions with Trp (Figure S5F, Supporting Information) or Tyr.

Exploiting P2′ specificity differences to engineer new inhibitor variants that show improved selectivity. Previous studies have used SFTI-1 to design new inhibitors for a diverse array of protease targets[21, 24, 26, 27, 31, 43, 44]. These engineered variants were developed using a range of different strategies, but in several cases, the P2′ residue (Ile) was not targeted for refinement. Therefore, to explore whether the P2′ specificity data from our study can be used in the design process to improve the selectivity of engineered inhibitors, we studied the effect of substituting Ile7 in existing SFTI variants that were developed to inhibit matriptase, KLK14 and thrombin.

As with members of the inhibitor library, we acquired 1D ^{1}H-NMR spectra to confirm that the new engineered inhibitors (Table 1) had folded correctly. As shown in Figure S2 (Supporting Information), the spectra showed sharp lines and well-dispersed signals, indicating that the engineered inhibitors were well-folded. The intensity and number of amide resonances in the spectra suggests that most of the inhibitors adopt one main conformation in solution, with the exception of **4**, which adopts multiple conformations. To investigate this further, we acquired 2D homonuclear TOCSY and NOESY spectra of **4**. Based on the presence of NOEs between the Hα/Hδ of a Pro and the Hα of the preceding residue, we found that the conformational heterogeneity was due to *cis-trans* isomerization of the Asp7-Pro8 and Pro8-Pro9 peptide bonds.

Matriptase. To date, the most common target in SFTI-based engineering studies has been the cancer-related protease, matriptase[26, 27, 31]. In one study, a high affinity SFTI variant for matriptase was developed

using a truncated inhibitor scaffold (spanning residues 1-12) with a non-cyclic peptide backbone and an amidated C-terminus[27] (Compound **3**, Table 2). However, an ongoing challenge in designing SFTI-based inhibitors for matriptase has been achieving a suitable degree of selectivity. Indeed, **3** was reported to show only 6-fold selectivity for matriptase over trypsin[27]. Our inhibitor library screen indicated that matriptase and trypsin had divergent P2′ specificity profiles (Figure 3). In particular, matriptase seemed to tolerate Asp at P2′ reasonably well (compared to Ile), whereas Asp was poorly favored by trypsin. With the aim of improving the selectivity of **3** for matriptase over trypsin, we synthesized **3** and the corresponding Asp7 variant, **4**. Screening both compounds against matriptase revealed that Asp7 was tolerated almost as well as Ile7 with only a slight decrease in activity observed for **4** (Table 2). Modeling **4** in complex with matriptase indicated that Asp7 formed a favorable hydrogen bond with His143 (His752) (Figure 4A), that was similar to that seen for P2′ Asn. However, equivalent assays with trypsin revealed that the Asp7 variant showed a 350-fold decrease in activity compared to **3** (P2′ Ile), which improved the inhibitor's selectivity to 33-fold (under the assay conditions in our study, we found that **3** was a more potent inhibitor of trypsin than matriptase). Further, **4** showed 196-fold selectivity over plasmin due to lower off-target activity compared to **3**, while neither variant **3** nor **4** showed any inhibition of thrombin up to 50 μM (Table 2).

KLK14. A further therapeutic application where SFTI variants are currently under development is for inhibiting kallikrein proteases implicated in skin diseases[24, 45]. In a recent study, we designed a potent and selective inhibitor for KLK14 (**5**, Table 1) by engineering the SFTI-1 contact β-strand (P4-P1 residues) as well as residues 12 and 14[24]. This variant is a potent KLK14 inhibitor (K_i = 2.0 nM) and shows approximately 180-fold selectivity over KLK5 and trypsin. Our data from the P2′ library screen indicated that Lys7 was reasonably well tolerated by KLK14, but was strongly disfavored by related skin proteases, KLK5 and KLK7 (Figure 3). This suggested that replacing Ile7 with Lys could further increase the selectivity of **5**. The P2′ Lys variant (**6**) showed a slight decrease in activity against KLK14 (K_i = 7.0 nM, Table 2), but had dramatically lower activity against KLK5 and KLK7 (K_i > 50,000 nM), which

represents more than 7,000-fold selectivity over these proteases. Molecular modeling of **6** in complex with KLK14 indicated that Lys7 was involved in cation-π interactions with Tyr150 (Tyr158) (Figure 4B) as seen with **2K** in the initial library screen. Additionally, **6** showed 8.4-fold weaker inhibition of trypsin compared to **5** (457-fold selectivity), but maintained reasonable activity against KLK4 (Table 2), which was consistent with data from the inhibitor library screen (Figure 3).

Broad-range kallikrein inhibitors. In addition to engineering selective inhibitors for skin kallikrein proteases, there is growing interest in developing multi-target inhibitors where a single variant shows potent activity against KLK5, KLK7 and KLK14[46]. We found that the starting compound in the P2′ inhibitor library (**2I**) was a potent inhibitor of all three kallikrein proteases (Table 2). A further consideration is whether the multi-target inhibitor should also block matriptase as this protease has been implicated in activating KLK5 and KLK7 in skin diseases[47]. However, deficient matriptase activity is also implicated in causing a skin disease (ichthyosis hypotrichosis syndrome)[48] and thus, a broad-range inhibitor showing activity against matriptase may carry a risk of negative side-effects. In assays with matriptase, **2I** showed weak activity (K_i = 15,700 nM), corresponding to at least 930-fold selectivity for KLK5, KLK7 and KLK14. Yet, when the remaining P2′ variants in the library screen were considered, it appeared that replacing Ile_7 with Asn (**2N**) could provide a more effective KLK5, KLK7 and KLK14 inhibitor without compromising the variant's low activity against matriptase. Whereas **2N** showed a slight decrease in activity against KLK5 and KLK14 compared to **2I**, inhibition of KLK7 was improved by over 20-fold (Table 2). Additionally, **2N** showed weaker activity against matriptase (K_i = 37,300 nM), which equates to over 7,000-fold selectivity.

Thrombin. Activation of thrombin is a key step in the coagulation pathway, and thus a considerable amount of research has focused on the development of potent and selective thrombin inhibitors[40]. Although SFTI-1 is a poor inhibitor of thrombin (K_i = 5,000 nM), a previous study showed that substituting the P1 Lys residue with Arg (**7**) could improve the inhibitor's activity by almost 6-fold (K_i = 860 nM)[26]. Our findings from the P2′ inhibitor library revealed that Tyr7 (62%) led to stronger inhibition

of thrombin than Ile7 (39%). Additionally, Tyr7 appeared to be less favored by plasmin and matriptase compared to Ile. Screening **7** and the corresponding P2′ Tyr variant (**8**) against thrombin revealed that replacing Ile7 with Tyr led to a further increase in activity (Table 2). Modeling **8** in complex with thrombin revealed similar interactions involving P2′ Tyr (Figure 4C) as seen previously with **2Y**. Further, comparing inhibition of other proteases by **7** and **8** revealed that **8** showed 5-fold lower activity against trypsin and plasmin, and more than 90-fold lower activity against matriptase. While **8** remains a very potent trypsin inhibitor and may have limited potential for therapeutic development, these proof-of-concept findings indicate that P2′ Tyr can be incorporated into a future SFTI-based inhibitor for thrombin to improve its potency and selectivity.

DISCUSSION AND CONCLUSIONS

In this study, we developed a new library of cyclic peptide inhibitors that can be used to characterize the P2′ specificity of different serine proteases in order to improve the selectivity of engineered inhibitor variants. Screening the inhibitor library against thirteen different proteases revealed several common preferences but ultimately, a unique P2′ specificity profile was identified for each enzyme. These effects were explored further using MD simulations where we found that the distinct P2′ preferences for each protease were often linked to regions of sequence diversity in the S2′ binding pocket. Finally, we used the newly identified P2′ specificity profiles to further optimize existing engineered inhibitors based on SFTI-1, demonstrating that our data from the inhibitor library screen is transferrable to other SFTI variants. Together with existing substrate-guided approaches, the strategy described in this study now makes it possible to use synthetic chemistry-based methods to screen key residues on both the primed and non-primed sides of the inhibitor's binding loop, allowing engineered variants to achieve higher potency and selectivity.

This study has first demonstrated that the P2′ residue has considerable potential for modulating the potency and selectivity of engineered Laskowski inhibitors. In existing structures of serine proteases, the S2′ pocket is typically well-defined and has a unique configuration in different proteases due to variation in sequence or loop conformation, hinting that interactions involving the P2′ residue are important to the enzyme's binding specificity. Additionally, studies on mesotrypsin have shown that interactions involving the S2′ subsite and P2′ residue have a major role in the protease's notorious resistance to Laskowski inhibitors (including aprotinin)[49, 50]. However, until now, there has been a lack of systematic, broadly-applicable methods for screening the P2′ specificity profile of target proteases (in the context of protease inhibition) to identify suitable modifications that can be harnessed in engineering studies. Using our library of P2′ inhibitor variants, we found that a range of residues with different biophysical properties were favored by various proteases, to the extent that almost all of twenty P2′ residues included in the study were highly preferred by at least one of the thirteen enzymes screened (Figure 3). Moreover, using these data to guide substitution of the P2′ residue in a series of existing engineered inhibitors (for matriptase, KLK14 and thrombin) yielded new variants that showed marked improvements in selectivity. Interestingly, the P2′ substitutions were found to be effective regardless of whether the initial library screen was performed at low (25 nM) or high concentration (10-25 μM), further demonstrating the versatility of the method. Additionally, the slight decrease in potency for the new KLK14 and matripase variants was fully consistent with data from the library screen, as Ile (present in the starting variant) was more favored than new residue (Lys and Asp, respectively) and the substitution was made with the intention of improving the inhibitor's selectivity.

Comment [A4]: I think notable rather than notorious.

The findings presented here also provide new insights into the significance of Ile at P2′ in SFTI-1 and its prevalence in structurally-related inhibitors from the Bowman-Birk family. SFTI-1 has been found to inhibit a number of proteases with reasonably high affinity, including trypsin, matriptase, cathepsin G, KLK4, KLK5, KLK14[21, 24, 26, 51]. Across all P2′ residues included in the inhibitor library, we found that Ile was one of the most preferred residues for almost all proteases screened (Figure 3), together with Val,

which is structurally similar to Ile. Additionally, Ile is found at P2′ in almost half of the 190 Bowman-Birk inhibitor sequences (from different species) listed in the MEROPS database[52] (Figure 5A). In an earlier study, Leatherbarrow and colleagues suggested that Ile may be prevalent at P2′ in Bowman-Birk inhibitors either to favor broad-range inhibition or to allow the binding loop to adopt an optimal conformation for interaction with the protease[29]. Our findings across thirteen diverse serine proteases provide new evidence that supports the former hypothesis (broad-range inhibition). Indeed, Ile appeared to be dispensable at P2′ as although it was favored by most proteases, we identified chemically diverse P2′ residues that performed as well or better than Ile for each protease screened, suggesting that Ile is not a structural requirement for Bowman-Birk inhibitors. Interestingly, branched hydrophobic residues (Val, Ile and Leu) are also found at P2′ in over 80% of serine protease pro-segments that undergo cleavage during zymogen activation (Figure 5B). This suggests that broad-range Laskowski inhibitors and their protease targets have both capitalized on the widespread preference for branched hydrophobic residues at P2′ for the purpose of generating binding motifs that are tolerated by many different serine proteases. By extension, the broad tolerance of Ile and Val at P2′ is also significant in the context of inhibitor engineering, as retaining these types of residues at P2′ is likely to limit the selectivity of new inhibitor variants.

Comment [A5]: "similar" is a bit too strong. They are both branched chain but have very different surface areas and degrees of freedom. Perhaps "chemically reminiscent" instead

Considering the broader aspects of the Laskowski mechanism and serine protease catalysis, the primed segment of the binding loop has an especially important role in the inhibitor's activity, as seen in this study. During serine protease catalytic processing, once the scissile bond has been cleaved, the primed segment becomes the initial leaving group, whereas the non-primed segment remains covalently bound to the protease via an ester bond (acyl-enzyme complex)[53]. For Laskowski inhibitors, the positioning of the neo-N-terminus (P1′ residue) appears to be critical for efficient religation as it is proposed to engage the catalytic His residue (in place of the hydrolytic water molecule) to initiate the resynthesis reaction[54, 55]. Therefore, the degree of complementarity between the primed segment of the inhibitor and the protease will probably have a strong influence on whether cleavage of the reactive site bond leads to hydrolysis or

religation (as seen previously with mesotrypsin[49, 50]). Since these interactions are specific to the inhibitory mechanism, it may not be as effective to use substrate-based screens to identify primed segment substitutions as the set of sequences that are cleaved rapidly will also include those that generate efficient leaving groups. Indeed, although existing P2′ cleavage specificity data for trypsin, thrombin and factor Xa[56] shows a number of similarities with our inhibitor library data, the overall level of correlation is only modest. However, by using a substrate-guided approach to optimize the non-primed residues, followed by an inhibitor-based screen to refine the primed segment, we were able to overcome this problem and developed a potent KLK14 inhibitor, for example, that showed more than 7,000-fold selectivity over closely-related skin proteases, KLK5 and KLK7.

The potency and selectivity of the new inhibitor variants described in this study compare favorably with those developed by existing methods, including phage display. While phage display has major advantages relating to the number of variants that can be screened, its ability to scan both primed and non-primed residues simultaneously, and it is entirely inhibitor-based, our synthetic chemistry approach is versatile, allows validation that each library member is present (in the same concentration), and focuses solely on variants that bind within the active site. Thus, selection of variants that are not effective inhibitors of the target protease, as seen in a previous phage display study involving thrombin[15], is avoided. However, there are several factors that should be considered when transferring this study's findings to other engineered inhibitors. First, all variants in the inhibitor library contain Arg as the P1 residue. Therefore, we cannot exclude the possibility that certain residues may become more or less favored (due to co-operativity effects) when the P1 (or an adjacent) residue is changed. For example, in an engineered variant containing Phe at P1, we have noticed that P2′ Tyr becomes highly favored by chymotrypsin. Second, although the geometry of the binding loop is conserved across all Laskowski inhibitors, how well the specificity trends identified in this study transfer to scaffolds from other inhibitor families remains to be investigated. Here, it is important to note that in SFTI-1 (and many other Bowman-Birk inhibitors), the primed segment has a unique conformation as the P2′ residue is followed immediately by two Pro

residues (or in some cases, one Pro and a small residue). Thus, the inhibitor backbone bends away (forming an anti-parallel β-strand with the P4-P1 residues) without engaging the conventional S3′ or S4′ binding sites.

In conclusion, our study demonstrates that the P2′ residue has a strong influence on Laskowski inhibitor potency and selectivity, and unveils a novel, inhibitor-based library screen to guide optimization of the P2′ residue in future engineering studies. We anticipate that the P2′ specificity profiles determined in this study will help to improve the potency and/or selectivity of engineered SFTI variants targeting many of the proteases we screened, many of which have pharmaceutical indications. Additionally, the inhibitor library is itself a valuable resource that, once synthesized and validated, can be screened against further proteases under investigation as targets for inhibitor design. Indeed, this strategy contributes to making more of the canonical loop accessible to optimization by synthetic chemistry-based methods, and when combined with a systematic approach for refining the non-primed binding sequence, provides a promising blueprint for engineering potent and selective serine protease inhibitors.

EXPERIMENTAL SECTION

Protein Expression and Protein Sources

Recombinant KLK proteases used in this study were expressed in zymogen (pro) form in *Pichia pastoris* strain X-33 (KLK5 and KLK7) or *Spodoptera frugiperda* Sf9 insect cells (KLK4 and KLK14), then purified and activated as described recently[24]. Active site titration was performed using either α_2-antiplasmin (Sino Biological) (KLK5), α_1-antitrypsin (Sigma-Aldrich) (KLK7) or 4-methylumbelliferyl-p-guanidinobenzoate (MUGB, Sigma-Aldrich) (KLK4 and KLK14). Active KLKs were supplemented with 20% (*v*/*v*) glycerol to allow storage at –80°C with minimal losses of activity. Remaining proteases were obtained from commercial distributers: β-trypsin, α-chymotrypsin, plasmin and factor XIa (Sigma-Aldrich); matriptase (R&D Systems); α-thrombin, factor IXa-β, factor Xa and factor α-XIIa (Molecular Innovation).

Peptide Synthesis

Cyclic inhibitors (including the P2′ inhibitor library and protease-targeted variants) were synthesized on 2-chlorotrityl chloride resin (0.8 mmol/g) while linear inhibitors were synthesized on Rink amide MBHA resin (0.8 mmol/g) using a Symphony automated peptide synthesizer (Protein Technologies, Inc). For coupling reactions, Fmoc N-protected amino acids (4 equiv) were activated with 4 equiv *O*-(6-chlorobenzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate (HCTU) and 8 equiv N,N-diisopropylethylamine (DIPEA) in DMF (2 × 5 min reactions per residue). The Fmoc protecting group was subsequently removed using 30% piperidine in DMF (2 × 3 min). Assembled peptides (retaining side-chain protecting groups) were cleaved from the resin using 1% TFA in DCM, and collected by precipitation in diethyl ether. Head-to-tail cyclization was performed in DMF (50 mL per 0.1 mmol peptide) with 4 equiv 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate (HATU) and 8 equiv DIPEA at room temperature for 3 hours. To remove DMF and activators, the post-cyclization solution was mixed with an equal volume of DCM, and then washed with

5 volumes of H_2O (2-3 times). The organic phase was recovered, solvent was removed by rotary evaporation, and side-chain protecting groups were removed by cleavage using TFA/triisopropylsilane/H_2O (96:2:2) (20 mL per 0.1 mmol peptide). Fully deprotected peptides were collected by precipitation in diethyl ether. Formation of the intramolecular disulfide bond was achieved by vigorous stirring at room temperature (3 hours) in 0.1 M sodium bicarbonate buffer (pH 8.5) containing 10 µM oxidized glutathione (total volume: 100 mL per 0.1 mmol peptide).

Peptide *para*-nitroanilide (pNA) substrates were synthesized using 2-chlorotrityl chloride resin that had been derivatized with *para*-phenylenediamine (Sigma Aldrich), as previously described[57]. Assembly of linear peptides and liberation of protected peptides from the solid support was performed as above. Oxidation of the C-terminal *para*-aminoanilide group was carried out in solution using 6 equiv Oxone (Sigma Aldrich), as previously described[57]. Side-chain protecting groups were removed by TFA cleavage as described for SFTI variants.

Peptide Purification and Mass Spectrometry Analysis

SFTI variants and peptide-pNA substrates were purified by reverse phase HPLC (Shimadzu Prominence) using a 5 µm ZORBAX Extend-C18 PrepHT column (21.2 × 250 mm) and a linear gradient of 10% acetonitrile/0.05% TFA to 50% acetonitrile/0.05% TFA. Peptide purity (> 95%) was confirmed by UPLC using a 5 µm Agilent 300 SB C18 column (2.1 × 50 mm) at 50°C with mobile phases as above (Figure S1, Supporting Information). Peptide masses were determined by electrospray ionization mass spectroscopy (Shimadzu Prominence, Table 1).

1D NMR

Unless otherwise stated, peptide TFA salts were dissolved in H_2O:D_2O (9:1) at a concentration of 1.2 mM (pH 3-4) including 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) as an internal standard (2 µg/mL, referenced at 0.00 ppm). ^{1}H NMR spectra were collected on a Bruker Avance-600 MHz spectrometer at

298 K. For TOCSY and NOESY experiments, a mixing time of 80 ms and 200 ms was used, respectively. Spectra were processed with TOPSPIN 2.1 and further analyzed with CCPNMR 2.2.2 if required.

Inhibition Assays

Inhibition assays were performed in 96-well low-binding plates using 250 µL assay buffer (0.1 M Tris pH 8.0, 0.1 M NaCl and 0.005% Triton X-100) and 100 µM peptide-pNA substrates. Enzyme concentrations, protease-specific buffer additives and substrates are listed in Table S1 (Supporting Information). Proteolytic activity was measured by monitoring release of the pNA moiety using a plate reader spectrophotometer (λ = 405 nm, reading interval: 10 s, assay time course: 300 s), and the level of inhibition was calculated by comparing kinetic rates to control assays without inhibitor. Substrate constants (Michaelis–Menten) and inhibition constants (Morrison K_i) were determined by non-linear regression using GraphPad Prism 6.0.

Molecular Modeling

For proteases where their structure in complex with SFTI-1 has not been determined by x-ray crystallography, protease/SFTI complexes were generated by overlay of the trypsin/SFTI-1 complex (PDB ID: 1SFI) with existing structures for each protease: chymotrypsin (PDB ID: 1CHO), KLK5 (PDB ID: 2PSX), KLK7 (PDB ID: 2QXI), plasmin (PDB ID: 3UIR), thrombin (PDB ID: 3VXE), factor IXa (PDB ID: 3LC3), factor Xa (PDB ID: 2JKH), factor XIa (PDB ID: 3SOR). Additionally, homology models were created for KLK14 and factor XIIa using SWISS-MODEL[58] and the templates KLK5 (PDB ID: 2PSX) and hepatocyte growth factor activator (PDB ID: 2R0L), respectively. For KLK4 and matriptase, existing protease/SFTI complexes were used (PDB ID: 4KEL and 3PF8). Systems were solvated with TIP3P water and neutralized by Na^+/Cl^- counter ions to a final concentration of 100 mM in VMD 1.9.1[59]. This generated systems of approximately 22,000-25,000 atoms including 6,000-7,000 water molecules.

Each protease/inhibitor complex was equilibrated using a stepwise relaxation procedure. First, all heavy atoms were harmonically restrained with a force constant of 2 kcal/(mol $Å^2$) before a conjugate gradient minimization was performed (500 steps) using NAMD 2.9[60] and CHARMM27 force fields parameters. This was followed by heating to 298 K and simulating 500 ps under NPT conditions with periodic boundary conditions. A Langevin thermostat with a damping coefficient of 0.5 ps^{-1} was used to maintain the system temperature and the system pressure was maintained at 1 atm using a Langevin piston barostat. The particle mesh Ewald algorithm was used to compute long-range electrostatic interactions at every time step and non-bonded interactions were truncated smoothly between 10-12 Å. Hydrogen bonds were constrained by the SHAKE algorithm (or the SETTLE algorithm for water). For the second stage, the restraints were retained on the protease and inhibitor α-carbons only (1 ns), while all constraints were released in the third stage (1 ns). Production runs (10 ns) were performed under NVT conditions with otherwise identical force field and simulation parameters as above. Coordinates were saved every 100 simulation steps producing 50,000 frames per trajectory.

Sequence Alignments

The Bowman-Birk inhibitor (family I12) sequence alignment was performed using all 190 sequences listed in the MEROPS database v9.12[52] (date accessed: 9 April 2015). For the human serine protease alignment, all sequences (139) were extracted from the UniProt proteome database (http://www.uniprot.org/proteomes/) (date accessed: 9 April 2015). Sequences were aligned using Clustal Omega 1.2[61], with only serine protease sequences found to contain a canonical pro-sequence included in the alignment (112 out of 139). Graphics illustrating sequence diversity were generated using WebLogo 3.4 (http://weblogo.threeplusone.com/). The structure-based sequence alignment of 13 serine proteases in Figure S6 (Supporting Information) was generated using Chimera 1.8.1 and the PDB structures described above.

Supporting Information Available: Contains compound characterization (analytical UPLC traces and 1D NMR spectra), molecular modeling, structure-based protease sequence alignment, and protease-

specific kinetic assay conditions. This material is available free of charge via the Internet at http://pubs.acs.org.

Corresponding Author Information: Dr Joakim Swedberg; Phone, +61 (0)7 3346 2021; Email, j.swedberg@imb.uq.edu.au; Address, Institute for Molecular Bioscience, The University of Queensland, Brisbane QLD 4072, Australia.

Acknowledgement: We thank Dr Peta Harvey (The University of Queensland) for assistance with NMR experiments. This study was funded by the Australian National Health and Medical Research Council (NHMRC) [Grant 1059410]. David Craik is a NHMRC Senior Principal Research Fellow [Grant 1026501] and Joakim Swedberg is a NHMRC Early Career Fellow [Grant 1069819].

Abbreviations Used: KLK, kallikrein-related peptidase; MD, molecular dynamics; pNA, *para*-nitroanilide; SFTI-1, sunflower trypsin inhibitor-1

REFFERENCES

(1) Bachovchin, D. A., and Cravatt, B. F. (2012) The pharmacological landscape and therapeutic potential of serine hydrolases, *Nat. Rev. Drug Discov. 11*, 52-68.

(2) Drag, M., and Salvesen, G. S. (2010) Emerging principles in protease-based drug discovery, *Nat. Rev. Drug Discov. 9*, 690-701.

(3) Mullard, A. (2014) 2013 FDA drug approvals, *Nat. Rev. Drug Discov. 13*, 85-89.

(4) Coussens, L. M., Fingleton, B., and Matrisian, L. M. (2002) Matrix metalloproteinase inhibitors and cancer: trials and tribulations, *Science 295*, 2387-2392.

(5) Rosenbloom, D. I., Hill, A. L., Rabi, S. A., Siliciano, R. F., and Nowak, M. A. (2012) Antiretroviral dynamics determines HIV evolution and predicts therapy outcome, *Nat. Med. 18*, 1378-1385.

(6) Rong, L., Dahari, H., Ribeiro, R. M., and Perelson, A. S. (2010) Rapid emergence of protease inhibitor resistance in hepatitis C virus, *Sci. Transl. Med. 2*, 30ra32.

(7) Scott, C. J., and Taggart, C. C. (2010) Biologic protease inhibitors as novel therapeutic agents, *Biochimie 92*, 1681-1688.

(8) Laskowski, M., Jr., and Kato, I. (1980) Protein inhibitors of proteinases, *Annu. Rev. Biochem. 49*, 593-626.

(9) Laskowski, M., Jr., and Qasim, M. A. (2000) What can the structures of enzyme-inhibitor complexes tell us about the structures of enzyme substrate complexes?, *Biochim. Biophys. Acta 1477*, 324-337.

(10) Finkenstadt, W. R., and Laskowski, M., Jr. (1965) Peptide bond cleavage on trypsin-trypsin inhibitor complex formation, *J. Biol. Chem. 240*, 962-963.

(11) Finkenstadt, W. R., and Laskowski, M., Jr. (1967) Resynthesis by trypsin of the cleaved peptide bond in modified soybean trypsin inhibitor, *J. Biol. Chem. 242*, 771-773.

(12) Bode, W., and Huber, R. (1992) Natural protein proteinase inhibitors and their interaction with proteinases, *Eur. J. Biochem. 204*, 433-451.

(13) Schechter, I., and Berger, A. (1967) On the size of the active site in proteases. I. Papain, *Biochem. Biophys. Res. Commun. 27*, 157-162.

(14) Roberts, B. L., Markland, W., Ley, A. C., Kent, R. B., White, D. W., Guterman, S. K., and Ladner, R. C. (1992) Directed evolution of a protein: selection of potent neutrophil elastase inhibitors displayed on M13 fusion phage, *Proc. Natl. Acad. Sci. U. S. A. 89*, 2429-2433.

(15) Markland, W., Ley, A. C., and Ladner, R. C. (1996) Iterative optimization of high-affinity protease inhibitors using phage display. 2. Plasma kallikrein and thrombin, *Biochemistry 35*, 8058-8067.

(16) Cicardi, M., Levy, R. J., McNeil, D. L., Li, H. H., Sheffer, A. L., Campion, M., Horn, P. T., and Pullman, W. E. (2010) Ecallantide for the treatment of acute attacks in hereditary angioedema, *N. Engl. J. Med. 363*, 523-531.

(17) Lu, W., Zhang, W., Molloy, S. S., Thomas, G., Ryan, K., Chiang, Y., Anderson, S., and Laskowski, M., Jr. (1993) Arg15-Lys17-Arg18 turkey ovomucoid third domain inhibits human furin, *J. Biol. Chem. 268*, 14583-14585.

(18) Martin, F., Dimasi, N., Volpari, C., Perrera, C., Di Marco, S., Brunetti, M., Steinkuhler, C., De Francesco, R., and Sollazzo, M. (1998) Design of selective eglin inhibitors of HCV NS3 proteinase, *Biochemistry 37*, 11459-11468.

(19) Grzesiak, A., Krokoszynska, I., Krowarsch, D., Buczek, O., Dadlez, M., and Otlewski, J. (2000) Inhibition of six serine proteinases of the human coagulation system by mutants of bovine pancreatic trypsin inhibitor, *J. Biol. Chem. 275*, 33346-33352.

(20) Harris, J. L., Backes, B. J., Leonetti, F., Mahrus, S., Ellman, J. A., and Craik, C. S. (2000) Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries, *Proc. Natl. Acad. Sci. U. S. A. 97*, 7754-7759.

(21) 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.

(22) Swedberg, J. E., de Veer, S. J., Sit, K. C., Reboul, C. F., Buckle, A. M., and Harris, J. M. (2011) Mastering the canonical loop of serine protease inhibitors: enhancing potency by optimising the internal hydrogen bond network, *PLoS One 6*, e19302.

(23) de Veer, S. J., Swedberg, J. E., Parker, E. A., and Harris, J. M. (2012) Non-combinatorial library screening reveals subsite cooperativity and identifies new high efficiency substrates for kallikrein-related peptidase 14, *Biol. Chem. 393*, 331-341.

(24) de Veer, S. J., Swedberg, J. E., Akcan, M., Rosengren, K. J., Brattsand, M., Craik, D. J., and Harris, J. M. (2015) Engineered protease inhibitors based on sunflower trypsin inhibitor-1 (SFTI-1) provide insights into the role of sequence and conformation in Laskowski mechanism inhibition, *Biochem. J. 469*, 243-253.

(25) Madala, P. K., Tyndall, J. D., Nall, T., and Fairlie, D. P. (2010) Update 1 of: Proteases universally recognize beta strands in their active sites, *Chem. Rev. 110*, PR1-31.

(26) 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.

(27) 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. 11*, 1848-1857.

(28) Korsinczky, M. L., Schirra, H. J., Rosengren, K. J., West, J., Condie, B. A., Otvos, L., Anderson, M. A., and Craik, D. J. (2001) Solution structures by 1H NMR of the novel cyclic trypsin inhibitor SFTI-1 from sunflower seeds and an acyclic permutant, *J. Mol. Biol. 311*, 579-591.

(29) Gariani, T., McBride, J. D., and Leatherbarrow, R. J. (1999) The role of the P2' position of Bowman-Birk proteinase inhibitor in the inhibition of trypsin. Studies on P2' variation in cyclic peptides encompassing the reactive site loop, *Biochim. Biophys. Acta 1431*, 232-237.

(30) Antalis, T. M., Bugge, T. H., and Wu, Q. (2011) Membrane-anchored serine proteases in health and disease, *Prog. Mol. Biol. Transl. Sci. 99*, 1-50.

(31) Quimbar, P., Malik, U., Sommerhoff, C. P., Kaas, Q., Chan, L. Y., Huang, Y. H., Grundhuber, M., Dunse, K., Craik, D. J., Anderson, M. A., and Daly, N. L. (2013) High-affinity cyclic peptide matriptase inhibitors, *J. Biol. Chem. 288*, 13885-13896.

(32) Magdolen, V., Sommerhoff, C. P., Fritz, H., and Schmitt, M. (2012) *Kallikrein-related peptidases. Characterization, regulation, and interactions within the protease web*, Walter de Gruyter GmbH. pp. 1-410, Berlin.

(33) Lawrence, M. G., Lai, J., and Clements, J. A. (2010) Kallikreins on steroids: structure, function, and hormonal regulation of prostate-specific antigen and the extended kallikrein locus, *Endocr. Rev. 31*, 407-446.

(34) de Veer, S. J., Furio, L., Harris, J. M., and Hovnanian, A. (2014) Proteases: common culprits in human skin disorders, *Trends Mol. Med. 20*, 166-178.

(35) Prassas, I., Eissa, A., Poda, G., and Diamandis, E. P. (2015) Unleashing the therapeutic potential of human kallikrein-related serine proteases, *Nat. Rev. Drug Discov. 14*, 183-202.

(36) Debela, M., Magdolen, V., Schechter, N., Valachova, M., Lottspeich, F., Craik, C. S., Choe, Y., Bode, W., and Goettig, P. (2006) Specificity profiling of seven human tissue kallikreins reveals individual subsite preferences, *J. Biol. Chem. 281*, 25678-25688.

(37) Mackman, N. (2008) Triggers, targets and treatments for thrombosis, *Nature 451*, 914-918.

(38) Edmunds, L. H., Jr. (2010) Managing fibrinolysis without aprotinin, *Ann. Thorac. Surg. 89*, 324-331.

(39) Deanda, A., Jr., and Spiess, B. D. (2012) Aprotinin revisited, *J. Thorac. Cardiovasc. Surg. 144*, 998-1002.

(40) Di Nisio, M., Middeldorp, S., and Buller, H. R. (2005) Direct thrombin inhibitors, *N. Engl. J. Med. 353*, 1028-1040.

(41) Swedberg, J. E., and Harris, J. M. (2011) Plasmin substrate binding site cooperativity guides the design of potent peptide aldehyde inhibitors, *Biochemistry 50*, 8454-8462.

(42) Bajaj, M. S., Ogueli, G. I., Kumar, Y., Vadivel, K., Lawson, G., Shanker, S., Schmidt, A. E., and Bajaj, S. P. (2011) Engineering kunitz domain 1 (KD1) of human tissue factor pathway inhibitor-2 to selectively inhibit fibrinolysis: properties of KD1-L17R variant, *J. Biol. Chem. 286*, 4329-4340.

(43) Kocsis, A., Kekesi, K. A., Szasz, R., Vegh, B. M., Balczer, J., Dobo, J., Zavodszky, P., Gal, P., and Pal, G. (2010) Selective inhibition of the lectin pathway of complement with phage display selected peptides against mannose-binding lectin-associated serine protease (MASP)-1 and -2: significant contribution of MASP-1 to lectin pathway activation, *J. Immunol. 185*, 4169-4178.

(44) Dbowski, D., Karna, N., Gowska, A., Stirnberg, M., Gutschow, M., Rolka, K., and Gitlin, A. (2015) Inhibitors of matriptase-2 based on trypsin inhibitor SFTI-1, *ChemBioChem*, (in press) doi: 10.1002/cbic.201500200.

(45) de Veer, S. J., Ukolova, S. S., Munro, C. A., Swedberg, J. E., Buckle, A. M., and Harris, J. M. (2013) Mechanism-based selection of a potent kallikrein-related peptidase 7 inhibitor from a versatile library based on the sunflower trypsin inhibitor SFTI-1, *Biopolymers 100*, 510-518.

(46) Tan, X., Soualmia, F., Furio, L., Renard, J. F., Kempen, I., Qin, L., Pagano, M., Pirotte, B., El Amri, C., Hovnanian, A., and Reboud-Ravaux, M. (2015) Toward the first class of suicide inhibitors of kallikreins involved in skin diseases, *J. Med. Chem. 58*, 598-612.

(47) Sales, K. U., Masedunskas, A., Bey, A. L., Rasmussen, A. L., Weigert, R., List, K., Szabo, R., Overbeek, P. A., and Bugge, T. H. (2010) Matriptase initiates activation of epidermal pro-kallikrein and disease onset in a mouse model of Netherton syndrome, *Nat. Genet. 42*, 676-683.

(48) Basel-Vanagaite, L., Attia, R., Ishida-Yamamoto, A., Rainshtein, L., Ben Amitai, D., Lurie, R., Pasmanik-Chor, M., Indelman, M., Zvulunov, A., Saban, S., Magal, N., Sprecher, E., and Shohat,

M. (2007) Autosomal recessive ichthyosis with hypotrichosis caused by a mutation in ST14, encoding type II transmembrane serine protease matriptase, *Am. J. Hum. Genet. 80*, 467-477.

(49) Salameh, M. A., Soares, A. S., Hockla, A., Radisky, D. C., and Radisky, E. S. (2011) The P(2)' residue is a key determinant of mesotrypsin specificity: engineering a high-affinity inhibitor with anticancer activity, *Biochem. J. 440*, 95-105.

(50) Szmola, R., Kukor, Z., and Sahin-Toth, M. (2003) Human mesotrypsin is a unique digestive protease specialized for the degradation of trypsin inhibitors, *J. Biol. Chem. 278*, 48580-48589.

(51) 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.

(52) Rawlings, N. D., Waller, M., Barrett, A. J., and Bateman, A. (2014) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors, *Nucleic Acids Res. 42*, D503-509.

(53) Hedstrom, L. (2002) Serine protease mechanism and specificity, *Chem. Rev. 102*, 4501-4524.

(54) Radisky, E. S., and Koshland, D. E., Jr. (2002) A clogged gutter mechanism for protease inhibitors, *Proc. Natl. Acad. Sci. U. S. A. 99*, 10316-10321.

(55) Zakharova, E., Horvath, M. P., and Goldenberg, D. P. (2009) Structure of a serine protease poised to resynthesize a peptide bond, *Proc. Natl. Acad. Sci. U. S. A. 106*, 11034-11039.

(56) Bianchini, E. P., Louvain, V. B., Marque, P. E., Juliano, M. A., Juliano, L., and Le Bonniec, B. F. (2002) Mapping of the catalytic groove preferences of factor Xa reveals an inadequate selectivity for its macromolecule substrates, *J. Biol. Chem. 277*, 20527-20534.

(57) Abbenante, G., Leung, D., Bond, T., and Fairlie, D. P. (2000) An efficient Fmoc strategy for the rapid synthesis of peptide para-nitroanilidies, *Lett. Pept. Sci. 7*, 347-351.

(58) Guex, N., and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling, *Electrophoresis 18*, 2714-2723.

(59) Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics, *J. Mol. Graph. 14*, 33-38, 27-38.

(60) Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD, *J. Comput. Chem. 26*, 1781-1802.

(61) Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, *Mol. Syst. Biol. 7*, 539.

FIGURES

Figure 1

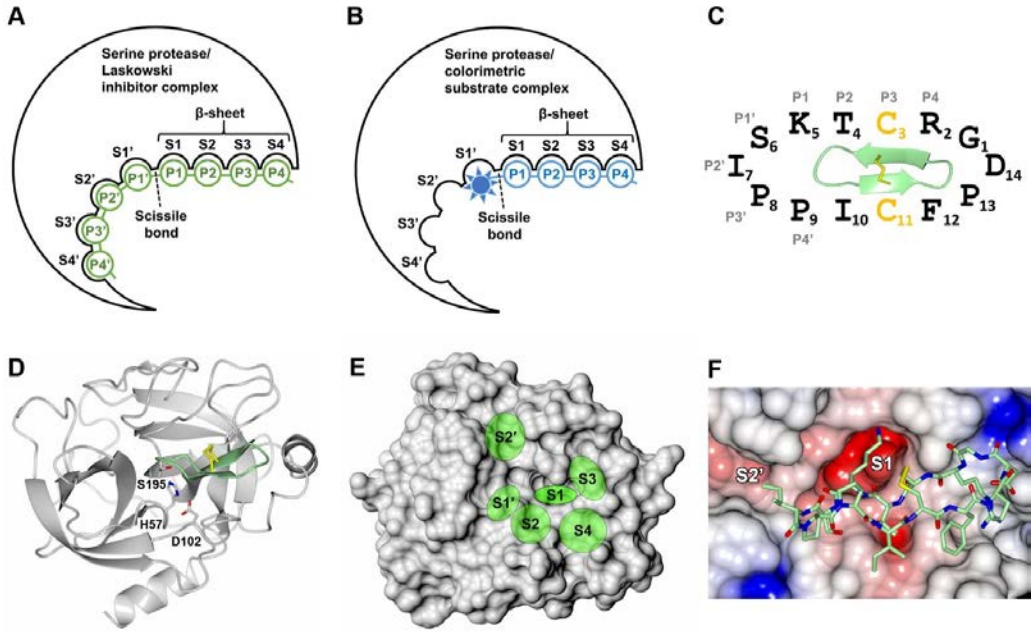

Figure 1: Structural features of the serine protease active site and binding interactions with sunflower trypsin inhibitor-1 (SFTI-1). (A) Schematic diagram showing binding of a Laskowski inhibitor (via the canonical loop) (green) to a serine protease (black) illustrating the substrate binding sites (S4-S4′) and protease binding residues (P4-P4′). (B) Schematic diagram showing binding of a colorimetric tetrapeptide substrate (blue) to a serine protease (black) with interactions between the P4-P1 residues and the S4-S1 sites. The colorimetric reporter group is represented as a symbol (near the S1′ pocket). (C) Schematic drawing of the SFTI-1 sequence. Cysteine residues forming the disulfide bond are highlighted in yellow and the protease binding residues (P4-P4′) are labeled (grey text). Inside the schematic is a ribbon diagram of the solution structure of SFTI-1 (green, PDB ID: 1JBL) with the disulfide bond shown in stick representation (yellow). (D) Ribbon diagram of SFTI-1 (green) in complex with trypsin (grey, PDB ID: 1SFI). The catalytic triad of trypsin (His57, Asp102 and Ser195) and the

disulfide bond of SFTI-1 are shown in stick representation. (E) Molecular surface of trypsin (grey) illustrating the SFTI-1 binding sites S4-S2′ (green). (F) Structure of SFTI-1 (stick representation) bound to trypsin (molecular surface colored by electrostatic potential) highlighting the S1 and S2′ substrate binding sites.

Figure 2

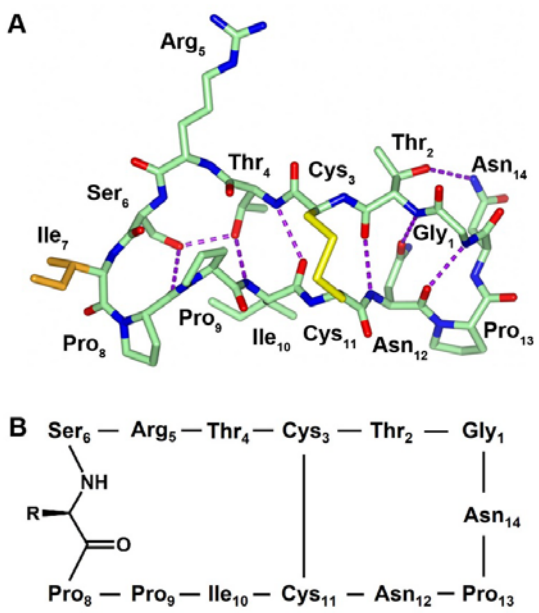

Figure 2: Designing a cyclic inhibitor library for screening the P2′ specificity of diverse serine proteases. (A) The structure of the starting compound **2I** was determined by NMR spectroscopy (BMRB accession code: 21057)[24] and is shown in stick representation. The variable P2′ residue is colored orange, the disulfide bond shown in yellow, intramolecular hydrogen bonds are represented by dashed lines (purple), and remaining atoms are colored as follows: carbon (green), oxygen (red), nitrogen (blue). (B) The sequence of the template for the SFTI library based on compound **2I** showing the atom backbone of the residue 7 and the variable side chain as -R.

Figure 3

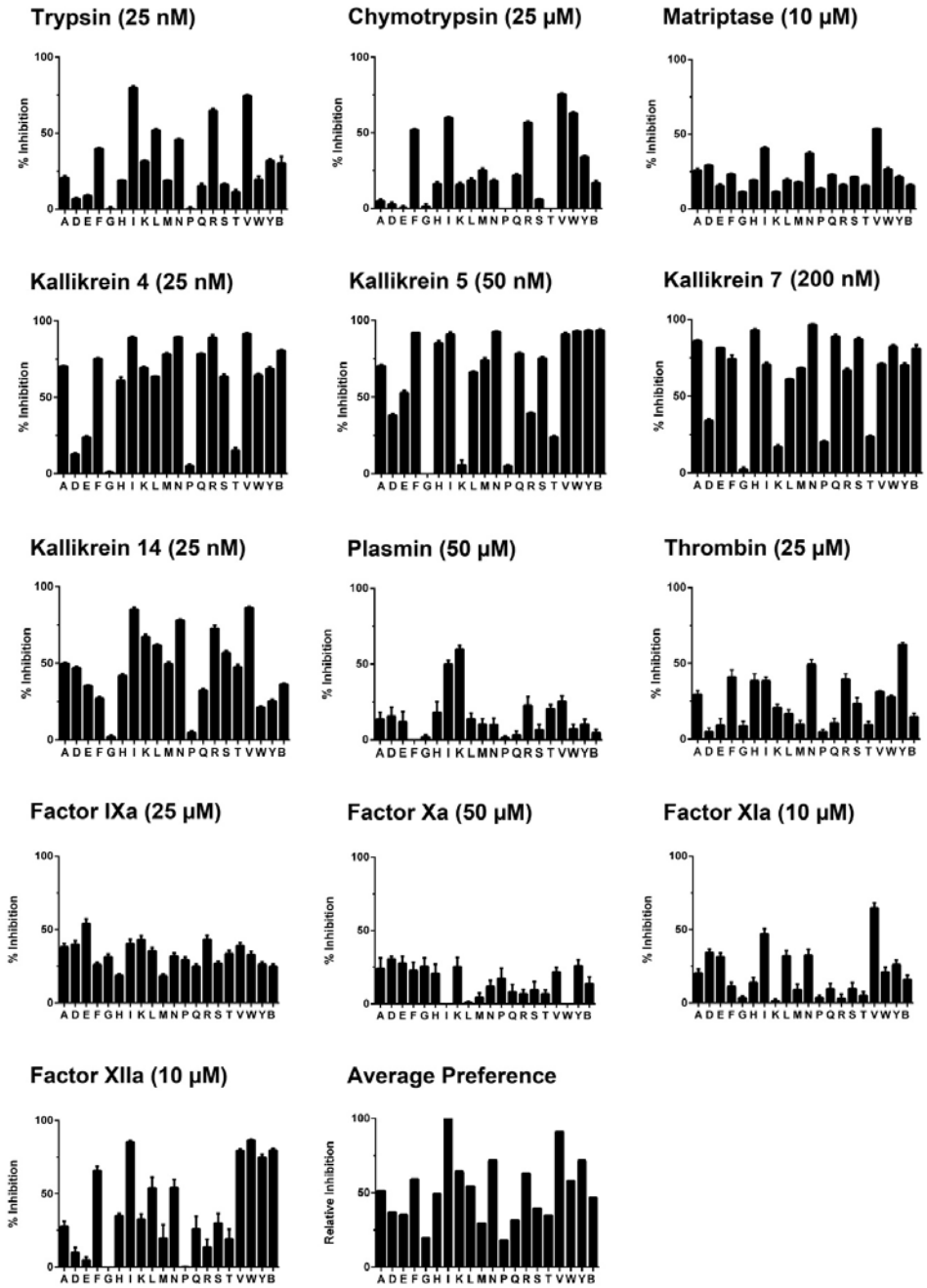
Trypsin (25 nM)
Chymotrypsin (25 µM)
Matriptase (10 µM)
Kallikrein 4 (25 nM)
Kallikrein 5 (50 nM)
Kallikrein 7 (200 nM)
Kallikrein 14 (25 nM)
Plasmin (50 µM)
Thrombin (25 µM)
Factor IXa (25 µM)
Factor Xa (50 µM)
Factor XIa (10 µM)
Factor XIIa (10 µM)
Average Preference
% Inhibition
Relative Inhibition
100
50
0
A D E F G H I K L M N P Q R S T V W Y B

Figure 3: Inhibitory activity of a SFTI-based library with a variable P2′ residue against thirteen serine proteases. The variable residue at the P2′ position is shown on the *x*-axis using single letter code for naturally occurring amino acids (B denotes biphenylalanine). The % inhibition (*y*-axis) for each variant was calculated by comparing kinetic rates to control assays without inhibitor. Data is expressed as mean ± S.E.M. from three independent experiments performed in duplicate. For each protease, the concentration of peptide used to screen the inhibitor library is given in brackets after the protease name. In the bottom row, the average P2′ preference (middle panel) was generated by normalizing the level of inhibition for each residue to Ile7 (set as 100%), and then calculating the mean for each residue across all proteases (excluding FXa since data could not be normalized as the P2′ Ile variant showed almost no inhibition).

Figure 4

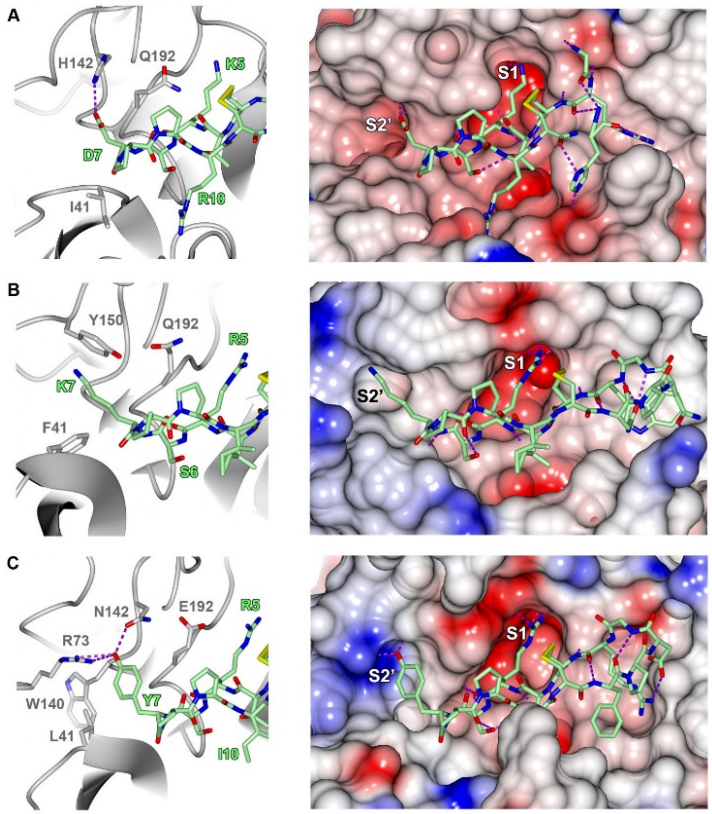

Figure 4: Molecular dynamics simulations reveal new favorable interactions introduced by library-guided P2′ substitutions. Average simulation structures for (A) matriptase/compound **4**, (B) KLK14/compound **6**, and (C) thrombin/compound **8** were calculated from 10 ns simulation trajectories (see methods for full details). Proteases are illustrated as ribbon diagrams (grey, left panels) or molecular surface graphics (colored by electrostatic potential, right panels) and inhibitors are shown in stick representation in both panels. Atom colors used are: carbon (green), nitrogen (blue), oxygen (red) and sulfur (yellow). Hydrogen bonds are represented by purple dashed lines.

Figure 5

Figure 5: Sequence diversity in the canonical loop of Bowman-Birk inhibitors and at the P2′ residue in the pro-segment of human serine protease zymogens. (A) Sequence diversity across the P4-P4′ sites in all Bowman-Birk inhibitors (family I12) currently listed in the MEROPS database[52]. (B) Sequence diversity at the P2′ residue in the pro-segment of all human serine protease zymogens that have a canonical pro-region.

TABLES

Table 1: Sequences, masses and purity of SFTI inhibitor variants

Compound	Peptide sequence[a,b]	Theoretical mass	Determined mass	Purity (%)
1	GRCTKSIPPICFPD	-	-	-
2A	GTCTRS**A**PPICNPN	1410.6	1410.3	99.2
2D	GTCTRS**D**PPICNPN	1454.6	1454.2	96.2
2E	GTCTRS**E**PPICNPN	1468.6	1468.2	99.0
2F	GTCTRS**F**PPICNPN	1486.7	1486.4	98.8
2G	GTCTRS**G**PPICNPN	1396.6	1396.2	98.6
2H	GTCTRS**H**PPICNPN	1476.7	1476.5	99.2
2I	GTCTRSIPPICNPN	1452.7	1452.4	99.9
2K	GTCTRS**K**PPICNPN	1467.7	1467.6	98.3
2L	GTCTRS**L**PPICNPN	1452.7	1452.4	98.5
2M	GTCTRS**M**PPICNPN	1470.7	1470.4	99.8
2N	GTCTRS**N**PPICNPN	1453.6	1453.5	97.3
2P	GTCTRS**P**PPICNPN	1436.6	1436.4	98.1
2Q	GTCTRS**Q**PPICNPN	1467.7	1467.9	99.6
2R	GTCTRS**R**PPICNPN	1495.7	1495.9	98.9
2S	GTCTRS**S**PPICNPN	1426.6	1426.4	99.9
2T	GTCTRS**T**PPICNPN	1440.6	1440.6	99.3
2V	GTCTRS**V**PPICNPN	1438.7	1438.4	99.1
2W	GTCTRS**W**PPICNPN	1525.7	1525.5	96.2
2Y	GTCTRS**Y**PPICNPN	1502.7	1502.4	97.7
2B	GTCTRS**B**PPICNPN[c]	1562.7	1562.4	99.9
3	GRCTKSIPP**R**C**H**-NH_2	1351.6	1351.4	96.4
4	GRCTKS**D**PP**R**C**H**-NH_2	1353.5	1353.3	99.9
5	G**W**C**IR**SIPPIC**N**P**N**	reported in ref. 24	-	-
6	G**W**C**IR**S**K**PPIC**N**P**N**	1564.9	1564.6	97.8
7	GRCT**R**SIPPICFPD	1541.8	1541.5	99.9
8	GRCT**R**S**Y**PPICFPD	1591.8	1591.5	99.9

[a]Residues that are mutated from compound **2I** are shown in bold for compounds **2A-2Y** (the P2′ residue is underlined). All peptides excluding compounds **3** and **4** have a cyclic backbone.
[b]Residues that are mutated from SFTI-1 (**1**) are displayed in bold for compounds **3-8**
[c]B: Biphenylalanine (BiP)

Table 2. Inhibition constants for engineered SFTI variants with P2′ substitutions

Compound	Sequence[a]	K_i (nM)			
		Matriptase	**Trypsin**	**Plasmin**	**Thrombin**
3	GRCTKS<u>I</u>PP**RCH**-NH_2	81.9 ± 6.8	8.0 ± 0.9	14,300 ± 1,400	> 50,000
4	GRCTKS<u>**D**</u>PP**RCH**-NH_2	87.0 ± 9.1	2,830 ± 111	17,100 ± 2,200	> 50,000
		KLK14	**KLK5**	**KLK4**	**Trypsin**
5	G**WCIR**S<u>I</u>PPIC**N**P**N**	2.0 ± 0.1	362 ± 14	9.0 ± 0.9	379 ± 9.2
6	G**WCIR**S<u>**K**</u>PPIC**N**P**N**	7.0 ± 0.5	> 50,000	19.9 ± 0.6	3,200 ± 99
		KLK5	**KLK7**	**KLK14**	**Matriptase**
2I	G**TCTR**S<u>I</u>PPIC**N**P**N**	2.1 ± 0.1	16.8 ± 0.4	0.4 ± 0.02	15,700 ± 2,200
2N	G**TCTR**S<u>**N**</u>PPIC**N**P**N**	5.2 ± 0.2	0.8 ± 0.05	1.2 ± 0.07	37,300 ± 7,500
		Thrombin	**Trypsin**	**Matriptase**	**Plasmin**
7	GRCT**R**S<u>I</u>PPICFPD	381 ± 13	0.03 ± 0.002	546 ± 21	121 ± 7.0
8	GRCT**R**S<u>**Y**</u>PPICFPD	214 ± 13	0.16 ± 0.007	> 50,000	549 ± 31

[a]Residues that are mutated from SFTI-1 (**1**) are displayed in bold font and the P2′ residue is underlined

TABLE OF CONTENTS GRAPHIC