Directed Evolution of Canonical Loops and Their Swapping between Unrelated Serine Proteinase Inhibitors Disprove the Interscaffolding Additivity Model

Eszter Boros [1], Fanni Sebák [2,3], Dávid Héja [1], Dávid Szakács [1], Katalin Zboray [1], Gitta Schlosser [4], András Micsonai [1], József Kardos [1], Andrea Bodor [2] and Gábor Pál [1]

*1 - **Department of Biochemistry,** ELTE Eötvös Loránd University, Pázmány Péter sétány 1/C, H-1117 Budapest, Hungary*
*2 - **Laboratory of Structural Chemistry and Biology,** Institute of Chemistry, ELTE Eötvös Loránd University, Pázmány Péter sétány 1/A, H-1117 Budapest, Hungary*
*3 - **Doctoral School of Pharmaceutical Sciences,** Semmelweis University, Üllői út 26, H-1085 Budapest, Hungary*
*4 - **Department of Analytical Chemistry,** Institute of Chemistry, ELTE Eötvös Loránd University, Pázmány Péter sétány 1/A, H-1117 Budapest, Hungary*

***Correspondence to Gábor Pál:** Pázmány Péter sétány 1/C, H-1117 Budapest, Hungary. gabor.pal@ttk.elte.hu*
https://doi.org/10.1016/j.jmb.2018.12.003
Edited by Sachdev Sidhu

Abstract

Reversible serine proteinase inhibitors comprise 18 unrelated families. Each family has a distinct representative structure but contains a surface loop that adopts the same, canonical conformation in the enzyme–inhibitor complex. The Laskowski mechanism universally applies for the action of all canonical inhibitors independent of their scaffold, but it has two nontrivial extrapolations. Intrascaffolding additivity states that all enzyme-contacting loop residues act independently of each other, while interscaffolding additivity claims that these residues act independently of the scaffold. These theories have great importance for engineering proteinase inhibitors but have not been comprehensively challenged. Therefore, we tested the interscaffolding additivity theory by hard-randomizing all enzyme-contacting canonical loop positions of a Kazal- and a Pacifastin-scaffold inhibitor, displaying the variants on M13 phage, and selecting the libraries on trypsin and chymotrypsin. Directed evolution delivered different patterns on both scaffolds against both enzymes, which contradicts interscaffolding additivity. To quantitatively assess the extent of non-additivity, we measured the affinities of the optimal binding loop variants and their binding loop-swapped versions. While optimal variants have picomolar affinities, swapping the evolved loops results in up to 200,000-fold affinity loss. To decipher the underlying causes, we characterized the stability, overall structure and dynamics of the inhibitors with differential scanning calorimetry, circular dichroism and NMR spectroscopy and molecular dynamic simulations. These studies revealed that the foreign loop destabilizes the lower-stability Pacifastin scaffold, while the higher-stability Kazal scaffold distorts the foreign loop. Our findings disprove interscaffolding additivity and show that loop and scaffold form one integrated unit that needs to be coevolved to provide high-affinity inhibition.

Introduction

Reversible serine protease inhibitors arose through convergent evolution yielding 18 non-homologous families. These families have distinct representative structures, but their mode of action is universal. Each inhibitor displays a convex surface loop, which becomes inserted into the substrate-binding cleft of the cognate enzyme. Regardless of the overall inhibitor structure, the conformation of this loop is essentially the same, that is, canonical. The canonical loop conformation is stabilized by the rest of the protein called the inhibitor scaffold.

In pioneering studies, the Laskowski laboratory established the molecular mechanism of action of these inhibitors, which became often referred to as the Laskowski mechanism. They focused their research on ovomucoid inhibitors of the Kazal family.

Over the years, they isolated ovomucoid third domains from 153 avian species [1–3] and characterized their interaction with different serine proteases [4]. Later, they used turkey ovomucoid third domain (OMTKY3), as a model, identified 10 enzyme-contacting positions, generated all single point mutants and characterized the binding of all 191 variants to 6 serine proteases [5–7]. They also determined the affinity of eglin c (potato I family) P1 variants (Schechter and Berger nomenclature that describes canonical loop (or substrate) positions Pn–Pn′, with the scissible bond located between P1–P1′, and corresponding Sn–Sn′ binding sites on the enzyme [8]) to the same set of enzymes [9]. Based on these data, they concluded that (i) besides the P1, other contact residues may affect affinity and specificity of the inhibitors, but non-enzyme-contacting residues do not exert any effect on binding, and (ii) the P1 side chain acts independently of the rest of the reactive loop and even of the entire inhibitor scaffold, and therefore, binding energy contribution of the P1 is governed by additivity.

Additivity has been a central topic of protein engineering since directed mutagenesis became feasible and the concept of the double mutation cycle was introduced by Fersht and his coworkers [10–12]. In general terms, additivity applies when the free energy change caused by multiple amino acid replacements at different positions equals the sum of individual free energy changes corresponding to the single amino acid replacements. This requires that the individual positions act independently.

The term intrascaffolding additivity coined by the Laskowski group was based on their observation that identical P1 residue substitutions in different ovomucoid third domain orthologs cause near identical free energy changes when binding to the same enzyme [6]. They introduced the term interscaffolding additivity upon finding that analogous P1 residue replacements on two unrelated inhibitors, OMTKY3 and eglin c, had similar effects on enzyme binding [9]. Based on the universal mode of canonical inhibitor action, they suggested that interscaffolding additivity should apply for every canonical inhibitor family, all enzyme-contacting positions and all 20 amino acid residues [9], and stated that the scaffold does not directly affect the binding affinity of the inhibitor [13].

Based on association constants measured for OMTKY3 single mutants, they constructed a sequence to reactivity algorithm [7,14] to predict the affinity of any Kazal inhibitor to six proteases. Later, they narrowed the scope of the algorithm to a subset of Kazal sequences (those having either P2 Thr or P1′ Glu or both) based on potential non-additivity effects predicted by structural data [7,15]. Then, they applied the algorithm to predict the presumably strongest Kazal inhibitors against porcine pancreatic elastase and *Streptomyces griseus* protease B, respectively [16].

The theory of interscaffolding additivity was based on the characterization of single P1 mutants. It has never been assessed experimentally by a more general approach that could test cooperativity between positions. Some subsequent studies did not support universality of the interscaffolding additivity model [9,17,18], but no general conclusion for the applicability of this theory was drawn.

We became interested in this topic when in a directed protein evolution study we observed clear non-additivity within the Pacifastin family of canonical inhibitors [19], which disproved universality of the less general intrascaffolding additivity model. This finding turned our interest toward the more general theory of interscaffolding additivity, which, if applies, would dictate that in directed evolution experiments the same binding loop sequence pattern should be selected on unrelated scaffolds against the same serine protease. In contrast, we found that markedly different binding sites were evolved on sunflower trypsin inhibitor *versus Schistocerca gregaria* protease inhibitor 2 (SGPI-2) when these inhibitors were selected against two complement lectin pathway proteinases, mannan-binding lectin associated serine proteinase (MASP)-1 and MASP-2 [20,21]. Nevertheless, the 14-amino-acid sunflower trypsin inhibitor is not a representative model of canonical inhibitors, and the highly specific MASP enzymes are not the archetypes of serine proteinases either.

Therefore, we designed this study to specifically test interscaffolding additivity by using two classical model enzymes and two typical, unrelated canonical inhibitors. We evolved optimal bovine trypsin-binding and bovine chymotrypsin-binding inhibitory loops in the structural context of the Kazal family serine protease inhibitor Kazal type 1 (SPINK1) and the Pacifastin family SGPI-2 scaffolds. We obtained characteristically different binding loop sequence patterns, which clearly demonstrated a lack of interscaffolding additivity. To quantitatively assess the magnitude of this non-additivity and decipher the underlying causes, we designed individual variants including the optimal binding loop variants and their binding loop-swapped versions, and characterized their affinity, stability, solution structure and dynamics.

Results

SGPI-2 and SPINK1 inhibitor variants were selected for binding to bovine trypsin and chymotrypsin by phage display

We subjected the canonical loop of two unrelated inhibitors, SGPI-2 from the Pacifastin family and SPINK1 from the Kazal family (Fig. 1), to directed evolution. We determined the residue preferences at

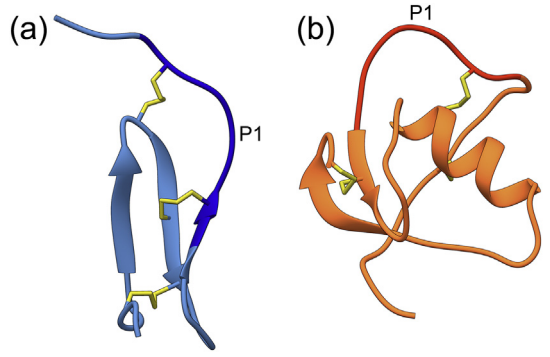

Fig. 1. Overall structure of (a) SGPI-2 and (b) SPINK1. The inhibitors are oriented to highlight the main structural elements of the scaffolds. Disulfide bridges are shown in yellow, and the reactive loop between the P4–P4′ amino acids is highlighted with dark blue for SGPI-2 (PDB ID: 2XTT) and red for SPINK1 (PDB ID: 1TGS). The position of the P1 side chain is also marked. The molecular graphics were drawn with Chimera.

each individual binding position in the form of characteristic sequence patterns (Fig. 2), which defined the corresponding consensus sequences.

Both SGPI-2 and SPINK1 were fused to the p8 coat protein and displayed monovalently on the surface of M13 bacteriophage. In both inhibitors, six positions of the canonical loop were fully randomized such that the structurally important Cys residues were preserved. The size of the resulting inhibitor-phage libraries, which comprised 7.00×10^8 SGPI-2 and 6.25×10^8 SPINK1 variants, approached the theoretical library diversity of $32^6 = 1.07 \times 10^9$. Both libraries were selected for binding to bovine trypsin and chymotrypsin in separate experiments. Three rounds of sorting were performed resulting in over 200-fold enrichment values in each case. Target binding of clones from the third round was tested in phage ELISA.

The enzyme-selected sequence patterns are markedly scaffold-dependent

The number of individual target-binding clones identified by DNA sequencing was 49 and 52 from trypsin and chymotrypsin, respectively, in the case of SGPI-2 and 56–56 from both trypsin and chymotrypsin in the case of SPINK1 (Supplementary Table 1).

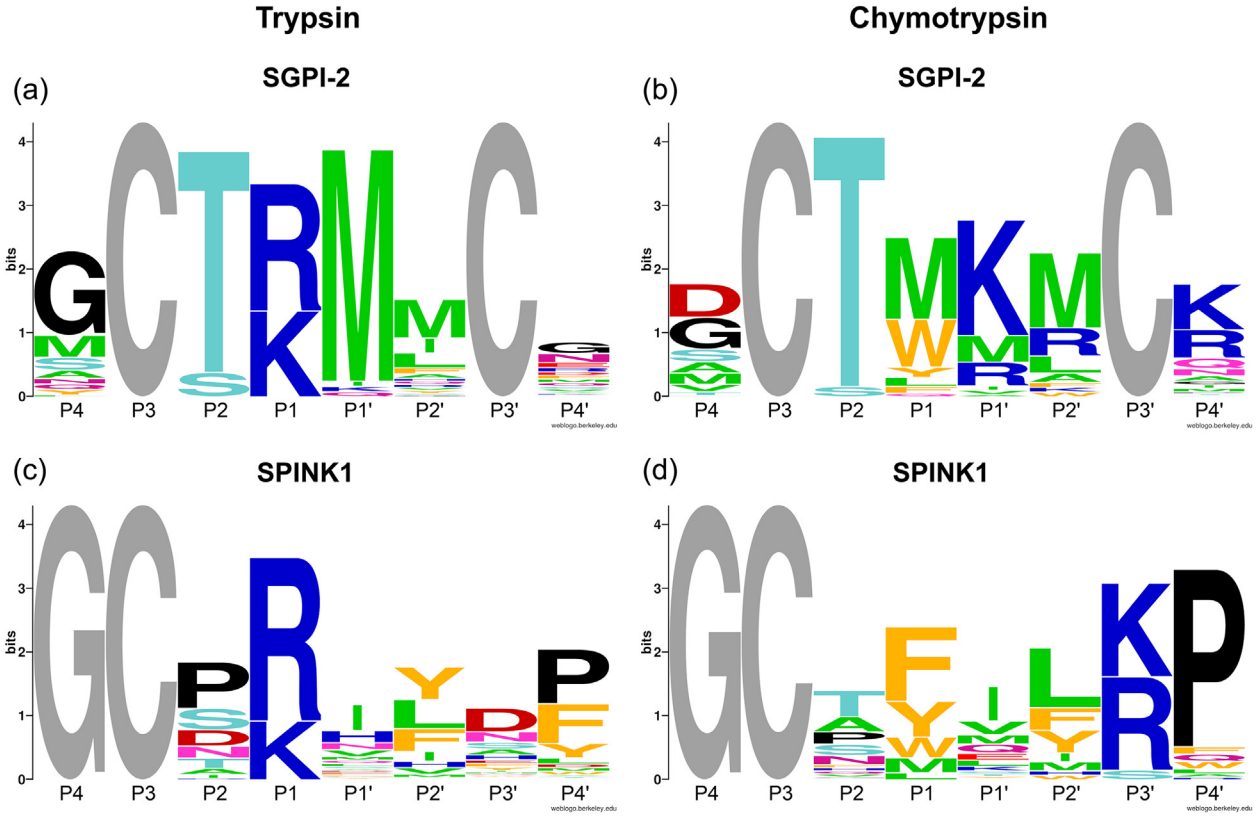

Fig. 2. Sequence logos generated by the WebLogo program representing the P4–P4′ positions of the phage display-selected inhibitors. Sequence logo representation of (a) bovine trypsin-selected SGPI-2 variants, (b) bovine chymotrypsin-selected SGPI-2 variants, (c) bovine trypsin-selected SPINK1 variants and (d) bovine chymotrypsin-selected SPINK1 variants. Amino acids are represented by their one letter code. The overall height of a stack of letters indicates the degree of conservation. Individual symbol height shows the codon-normalized relative frequency of the residue. Maximal symbol height corresponds to completely conserved amino acids, that is, Cys residues and the P4 Gly in SPINK1 shown in gray.

Based on these sequences, codon-normalized logos were generated that illustrate the amino acid preferences of the enzymes and the level of conservation at each randomized reactive loop position (Fig. 2).

Interestingly, both enzymes selected clearly different sequence patterns on the SGPI-2 *versus* SPINK1 scaffold. The trypsin-selected logos show scaffold-dependent differences at the P2, P1′, P2′ and P4′ positions (Fig. 2(a) and (c)). For example, trypsin selected almost exclusively a Met at the P1′ of SGPI-2, while it had no preference for any particular residue at the same position of SPINK1. The most notable differences between the chymotrypsin-selected logos (Fig. 2(b) and (d)) are located at the P1 and at the P1′–P4′ region. The most characteristic phenomenon is the different spatial distribution of positively charged amino acids on the two scaffolds. On SGPI-2, these scatter across P1′, P2′ and P4′, while on SPINK1, these selectively occur at P3′. We will get back to this phenomenon in Discussion.

Interestingly, irrespective of the target enzyme, Thr is dominant at the P2 of SGPI-2. The P2 position in the Pacifastin family is strictly conserved by natural evolution as Thr that has two important functions: it participates in a loop-stabilizing intramolecular H-bond network through its hydroxyl, while its methyl group binds into the shallow S2 pocket of the target enzymes [22] as detected in several atomic resolution structures [21,23–27]. This conservation was recapitulated by the *in vitro* evolution. In contrast, the P2 on the SPINK1 scaffold does not have an analogous structural role. The selection is dictated solely by the small S2 pocket of trypsin and chymotrypsin, and accordingly, mostly small residues (Gly, Ala, Ser, Pro, Thr, Asp) were selected here.

Pro was the most frequently selected both by trypsin and chymotrypsin at P4′ of SPINK1 and is a conserved residue in the Kazal family. Consequently, the *in vitro* evolution of SPINK1 recapitulated natural evolution by conserving a P4′ Pro on both enzymes corroborating the important structural role of this position in the Kazal family.

The marked differences between the logos demonstrate that the same enzyme prefers residues of different chemical character, size and shape at equivalent positions of the two different scaffolds. This clearly demonstrates that the optimal enzyme-binding loop sequence approximated by the consensus of the sequence logos is scaffold-dependent, which in turn indicates a lack of interscaffolding additivity.

Affinities of Sequence Logo-Derived Recombinant Inhibitor Variants Were Determined

We aimed to quantitate non-additivity by measuring the binding affinities of the four (two scaffolds on two targets) evolved consensus binding loop variants and by comparing these data to those of the four binding loop-swapped versions. As we proceeded with the study, more questions emerged that required two additional SGPI-2 and four additional SPINK-1 variants. So, along with the wild-types, altogether 16 variants were constructed, produced, isolated and tested. Equilibrium inhibitory constants (K_I) of the variants were determined for trypsin or chymotrypsin (Table 1).

The name, binding loop sequence and a short description of these variants are listed in Table 1. The SGPI-2 scaffold-based variants are named SG1-6, while the SPINK1 scaffold-based ones as SP1-8. The consensus carrying variants are named SG1, SG4, SP1 and SP5, while the loop-exchanged variants are SG2, SG5, SP2, and SP6. In the loop-exchanged SPINK1 variants, the P3′ Cys of the donor SGPI-2 loop was substituted with Ala to prevent the occurrence of an unpaired cysteine. Two additional SPINK1 variants (SP3 and SP7) were generated by introducing a P4′ Pro in the loop-exchanged variants to test whether it contributes to affinity and/or stability. Finally, single point mutants of the wild-type (SG3, SG6, SP4 and SP8) were also created, in which the original P1 was replaced with the consensus evolved residue selected by trypsin or chymotrypsin.

Affinity toward the target enzymes was improved by directed evolution

Human SPINK1, previously named as PSTI (pancreatic secretory trypsin inhibitor), is a tight-binding trypsin inhibitor that prevents premature trypsin activation in the pancreas [28]. Albeit weakly, it also inhibits chymotrypsin [29]. SGPI-2 is a potent inhibitor of bovine chymotrypsin [30], but it does not inhibit bovine trypsin. Specificity conversion by phage selection yielded high-affinity chymotrypsin-inhibiting SPINK1 and trypsin-inhibiting SGPI-2 variants. Moreover, directed evolution increased the already high affinity of SPINK1 on trypsin 5-fold as demonstrated by the SP1 variant. On the other hand, the already high chymotrypsin-binding affinity of SGPI-2 could only be marginally improved by the consensus loop represented by SG4 variant.

Swapping the consensus binding loop between scaffolds is detrimental to binding

According to the interscaffolding additivity theory, grafting the consensus binding loop from the original scaffold onto another one should transfer the same affinity along with the loop. Interscaffolding non-additivity can be quantified as the difference of two affinity values, both provided by the same evolved consensus binding loop, but in one case in the context of the original scaffold, while in the other case in the context of the unrelated recipient scaffold. We found

Table 1. Affinity and thermostability of the SGPI-2 and SPINK1 variants

Variant	Description	P4–P3–P2–P1–P1′–P2′–P3′–P4′	K_I (nM)	T_m (°C)
Trypsin-inhibiting variants				
SG1	Trypsin-selected consensus SGPI-2 variant	Gly-Cys-Thr-Arg-Met-Met-Cys-Gly	0.010	71.3[a]
SG2	Loop-exchanged trypsin-inhibiting SGPI-2 variant	Gly-Cys-Pro-Arg-Ile-Tyr-Cys-Pro	372	–
SG3	Trypsin-binding P1 variant of wild-type SGPI-2	Ala-Cys-Thr-Arg-Lys-Ala-Cys-Pro	0.4	83.8[a]
SPINK1	Wild-type SPINK1	Gly-Cys-Thr-Lys-Ile-Tyr-Asp-Pro	0.010	96.3[b]
SP1	Trypsin-selected consensus SPINK1 variant	Gly-Cys-Pro-Arg-Ile-Tyr-Asp-Pro	0.002	100.2[b]
SP2	Loop-exchanged trypsin-inhibiting SPINK1 variant	Gly-Cys-Thr-Arg-Met-Met-Ala-Gly	2.00	64.2[b]
SP3	Loop-exchanged trypsin-inhibiting SPINK1 variant with P4' Pro	Gly-Cys-Thr-Arg-Met-Met-Ala-Pro	0.80	84.4[b]
SP4	Trypsin-binding P1 variant of wild-type SPINK1	Gly-Cys-Thr-Arg-Ile-Tyr-Asp-Pro	0.016	95.9[b]
Chymotrypsin-inhibiting variants				
SGPI-2	Wild-type SGPI-2	Ala-Cys-Thr-Leu-Lys-Ala-Cys-Pro	0.027	76.4[a]
SG4	Chymotrypsin-selected consensus SGPI-2 variant	Asp-Cys-Thr-Met-Lys-Met-Cys-Lys	0.022	75.4[a]
SG5	Loop-exchanged chymotrypsin-inhibiting SGPI-2 variant	Gly-Cys-Thr-Phe-Ile-Leu-Cys-Pro	0.101	64.0[a]
SG6	Chymotrypsin-binding P1 variant of wild-type SGPI-2	Ala-Cys-Thr-Met-Lys-Ala-Cys-Pro	0.110	82.0[a]
SP5	Chymotrypsin-selected consensus SPINK1 variant	Gly-Cys-Thr-Phe-Ile-Leu-Lys-Pro	0.0005	83.7[b]
SP6	Loop-exchanged chymotrypsin-inhibiting SPINK1 variant	Asp-Cys-Thr-Met-Lys-Met-Ala-Lys	15	69.8[b]
SP7	Loop-exchanged chymotrypsin-inhibiting SPINK1 variant with P4' Pro	Asp-Cys-Thr-Met-Lys-Met-Ala-Pro	126	89.1[b]
SP8	Chymotrypsin-binding P1 variant of wild-type SPINK1	Gly-Cys-Thr-Phe-Ile-Tyr-Asp-Pro	0.054	95.8[b]

SGPI-2 variants and SPINK1 variants are named SG and SP, respectively. Inhibition constants (K_I) are provided in nanomolar units.
[a] Determined by spectroscopy.
[b] Determined by differential scanning calorimetry.

that loop transfer weakened the interaction in all four instances. Affinity loss was the largest, 186,000-fold, when the trypsin-binding optimized SPINK1 loop was transferred onto the SGPI-2 scaffold (SP1 *versus* SG2). The other three cases resulted in affinity drops ranging from 202-fold to 682-fold (Table 1).

Low performance of the transferred loop can also be quantified by comparing the affinity of the inhibitor carrying its own phage-optimized consensus to that of a variant having the same scaffold, but carrying a consensus evolved on another scaffold. This measures how much the foreign loop underperforms the one optimized in the context of the recipient scaffold. In 3 of the 4 cases, the affinity difference determined this way was 3 to 4 orders of magnitude. The most significant drop in affinity belongs again to SG2. It carries the consensus trypsin-binding loop selected on SPINK1 and inhibits the enzyme 37,000-fold more weakly than the consensus SG1 variant.

These results verify that differences of the sequence logos represent scaffold-dependent preferences of the target enzymes at analogous positions of unrelated scaffolds. The observed several orders of magnitude affinity drops prove that optimal enzyme-binding loop sequences are inhibitor scaffold specific, which disproves the interscaffolding additivity theory.

Restoration of P4′ Pro in loop-exchanged SPINK1 inhibitors does not significantly alter affinity

Natural evolution conserved a P4′ Pro in Kazal inhibitors [31], and directed evolution recapitulated the same results on both target enzymes. This Pro was lost in both binding loop-swapped SPINK1 variants, SP2 and SP6. In the corresponding SP3 and SP7 variants, we reconstituted this P4′ Pro to test whether it would enhance affinity.

Restoring P4′ Pro improved binding of SP3 to trypsin only 2.5-fold, while the equivalent substitution in SP7 resulted in an 8.4-fold affinity drop on chymotrypsin. A possible explanation is that the removal of the P4′ Lys eliminates an advantageous electrostatic interaction with Asp64 of chymotrypsin [32], which cannot be fully compensated by the stabilizing effect of Pro. In conclusion, restitution of the P4′ Pro had only minor effects on affinity, indicating that the several orders of magnitude affinity reduction of loop-exchanged SPINK1 inhibitors was not due to the loss of this conserved residue.

Introducing phage-optimized P1 residue into tight-binding wild-type does not improve affinity

As the interscaffolding additivity model was based mostly on single P1 mutant OMTKY3 variants, we created two single-mutant variants of both wild-type inhibitors each carrying a trypsin- (SG3 and SP4) or chymotrypsin-selected (SG6 and SP8) consensus P1 residue (see Table 1).

As expected, when the wild-type P1 was unsuitable for trypsin or chymotrypsin inhibition, introducing the phage-optimized P1 boosted affinity resulting in a subnanomolar SGPI-2 trypsin inhibitor and a tight-binding SPINK1 chymotrypsin inhibitor. However, when the wild-type inhibitor was already a tight

binder of the target protease, replacing the wild-type P1 with the phage-selected consensus residue caused a slight (1.6-4.1-fold) decrease in affinity. Although the effects are small, they show that the energetic contribution of the P1 is context dependent, which suggests a low level of intrascaffolding non-additivity.

Scaffold-disrupted inhibitors lose their function and become substrates

To further clarify the role of the inhibitor scaffold, we analyzed the functional consequences of native structure disruption by reducing the disulfide bridges. The consensus (SG4 and SP5) and the loop-exchanged (SG5 and SP6) chymotrypsin-binding variants were treated by the strong reducing agent, TCEP. Inhibitory potency of the reduced and native proteins was compared in enzymatic assays containing large (40- to 100-fold) molar excess of the inhibitors (Supplementary Fig. 1). Native conformation inhibitors completely blocked chymotrypsin activity, while the reduced forms provided only 0 to 12% inhibition. UPLC-MS analysis demonstrated that a 10-min incubation with chymotrypsin left the native inhibitors intact, while 100% of the reduced molecules were cleaved into peptide fragments (Supplementary Fig. 2). Importantly, the P1–P1′ scissile bond was cleaved in all reduced forms. This finding clearly demonstrates that a native conformation scaffold is essential for the inhibitory function.

Substitutions in the reactive loop affect thermostability of the inhibitors

To better understand the molecular background of the observed non-additivity effects, biophysical and structural experiments were carried out by circular dichroism (CD) spectroscopy, differential scanning calorimetry (DSC), NMR spectroscopy and molecular dynamics (MD) simulations.

Thermostability of the variants was determined by monitoring thermal denaturation either by far UV CD spectroscopy or by DSC (Table 1, Supplementary Figs. 3 and 4). Having a T_m of 96.3 °C, SPINK1 was found to be more stable than the smaller SGPI-2 having a T_m of 76.4 °C. The consensus variants, except SP5, have T_m values within a 6 °C range of the wild-type. In contrast, all loop-exchanged variants show a marked change in thermostability compared to their respective wild-types and consensus variants, as follows. There is an over 10 °C drop in the T_m values of two loop-exchanged variants, SG5 and SP6, while the other two variants, SG2 and SP2, did not exhibit cooperative denaturation. CD spectra taken at 10 °C intervals (Supplementary Fig. 5(a–d)) show that their structure becomes gradually disordered upon heating.

Introducing Pro at P4′ of the loop-exchanged variants resulted in T_m values nearly 20 °C higher than those of their parent molecules. This stability increase is in line with observations that Pro at the edges of the reactive loop can contribute to loop rigidity, which in turn can increase the affinity. This could be the case for SP3. However, as demonstrated by SP7, extra stability is not necessarily accompanied by more potent inhibition. Altogether, data on the affinity and thermostability of loop-exchanged SPINK1 P4′ variants imply that P4′ Pro primarily provides stability instead of directly affecting enzyme-binding.

Substituting the P1 with the most frequently selected residue did not change thermostability of SPINK1 variants. In contrast, both SGPI-2 P1 variants, SG3 and SG6, have higher than wild-type T_m.

CD measurements suggest unfolding of β-stranded regions in loop-exchanged SGPI-2 variants

Loop-exchange-induced structural changes were studied using far UV CD spectroscopy. CD spectra of the wild-types, consensus variants and loop-exchanged variants were measured for each enzyme-scaffold pair (Fig. 3(a–d)) and the data were analyzed using BeStSel [33] (Table 2). One should consider that accurate estimation of the secondary structure content of small, disulfide-bonded proteins based on CD spectra is challenging as their structures can tolerate moderate distortions.

Compared to the wild-type SPINK1, SP1 showed a decrease in the β-sheet content, while the α-helix content was increased. This observation, together with SP1 having the highest T_m, suggests that the loop optimization had an effect on the scaffold and increased its stability. In the case of the phage-display-optimized SG1, a 3% increase in antiparallel β-sheet content was observed.

In contrast, CD spectra of both loop-exchanged SGPI-2 variants are different from those of the wild-types and the consensus variants. Grafting the trypsin-selected SPINK1 loop on the SGPI-2 scaffold yielded SG2 that has markedly less β-sheet content compared to the consensus variant SG1, accompanied by the appearance of a short α-helical region. The 37,200-fold drop of trypsin-binding affinity of SG2 compared to SG1 might be due to these observed structural changes.

Significant structural adjustments also occurred when the chymotrypsin-binding SPINK1 loop was placed on the SGPI-2 scaffold generating SG5: the ratio of β-sheet content decreased, while only the ratio of turns and "others" increased, indicating that a part of the molecule, possibly one β-strand, became disordered. Interestingly, this readily detectable structural change is paired with only a modest 4.6-fold reduction in affinity (Table 1).

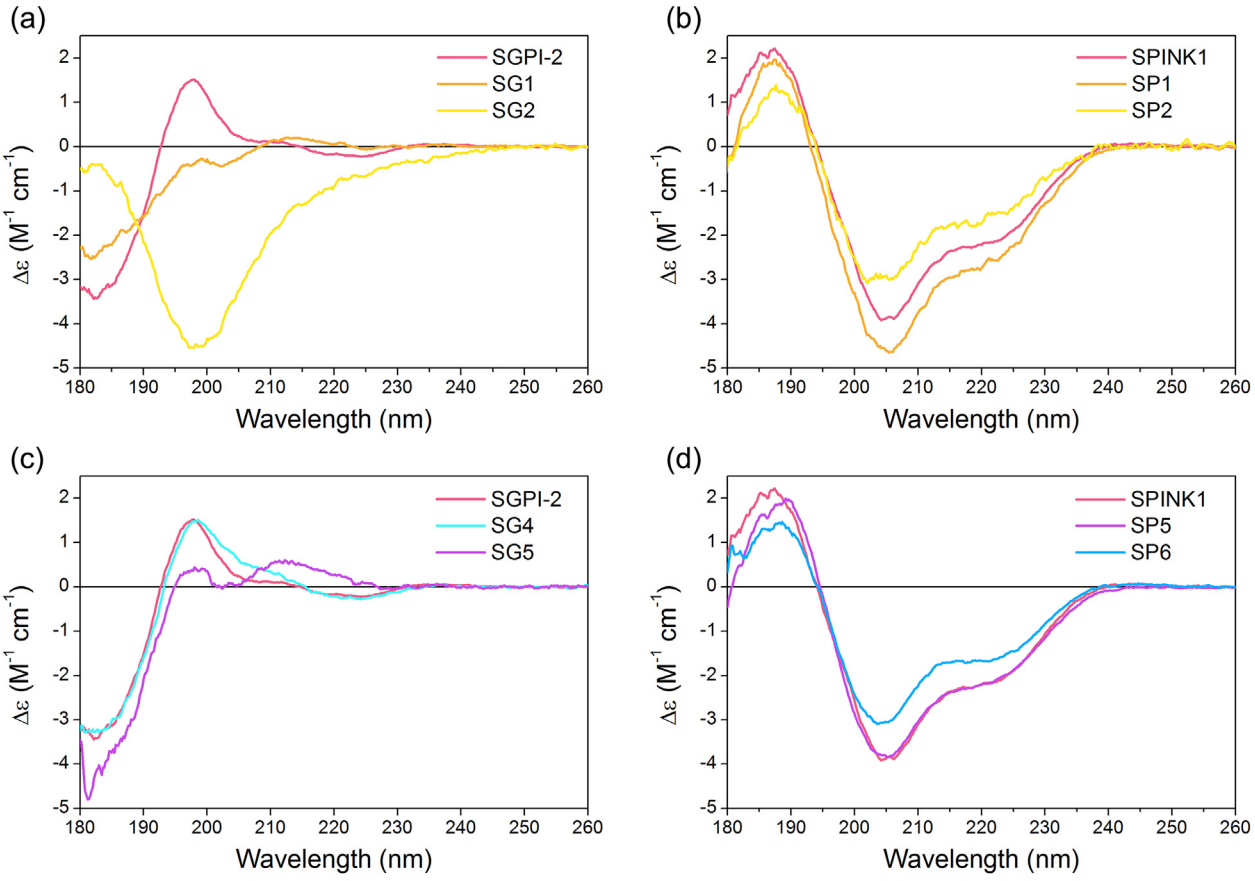

Fig. 3. Structural differences between the wild-type proteins, the consensus variants and the loop-exchanged variants monitored by CD spectroscopy. CD spectra of trypsin-binding SGPI-2 variants (a), trypsin-binding SPINK1 variants (b), chymotrypsin-binding SGPI-2 variants (c) and chymotrypsin-binding SPINK1 variants (d).

In contrast, incorporating the SGPI-2-based binding loops had no effect on the antiparallel β-sheet content in SPINK1, but affected its α-helix, which is linked to the loop by a disulfide bridge at P3 (Fig. 1). Intriguingly, we can observe a correlation between the α-helix content and the T_m values reflecting the stability (Table 2, r = 0.87, Pearson correlation coefficient). Nevertheless, the inhibitory activity of the loop-exchanged SPINK1 variants, SP2 and SP6, is reduced by 3-4 orders of magnitude.

Table 2. Estimated proportions of secondary structure elements in the variants obtained by BeStSel analysis of the CD spectra

Variant	Estimated secondary structure content (%)					
	α-Helix	Antiparallel β	Parallel β	Turn	Others	T_m
SPINK1	16.1	29.0	0.0	14.1	40.7	96.3
SP1	20.0	21.6	0.0	12.1	46.4	100.2
SP2	9.5	27.7	0.7	14.7	47.4	64.2
SP5	11.7	28.9	0.0	13.5	43.4	83.7
SP6	13.2	29.1	0.0	15.5	42.2	69.8
SGPI-2	3.2	28.5	0.0	17.7	50.8	76.4
SG1	4.2	31.8	0.9	16.1	47.1	71.3
SG2	7.3	22.7	0.0	19.1	50.8	–
SG4	2.0	30.2	0.0	18.5	49.3	75.4
SG5	1.5	23.2	0.0	20.8	54.5	64.0

Melting temperatures from thermal denaturation experiments are also presented.

In summary, accomodating a binding loop evolved on a different scaffold caused structural changes that are readily detected by CD spectroscopy. The less stable SGPI-2 structure suffered more extensive changes than the more stable SPINK1 structure. Nonetheless, there is no simple relationship between the degree of structural changes and the extent of affinity loss.

NMR spectroscopy localizes the structural change in loop-exchanged variant SG5

NMR analysis was performed for unlabeled SG4 and SG5 variants. Secondary chemical shifts and corresponding Secondary Structure Propensity (SSP) scores [34] were calculated from the assigned Hα chemical shifts. A plot against the amino acid sequence reveals the secondary structural motifs (Fig. 4(a)).

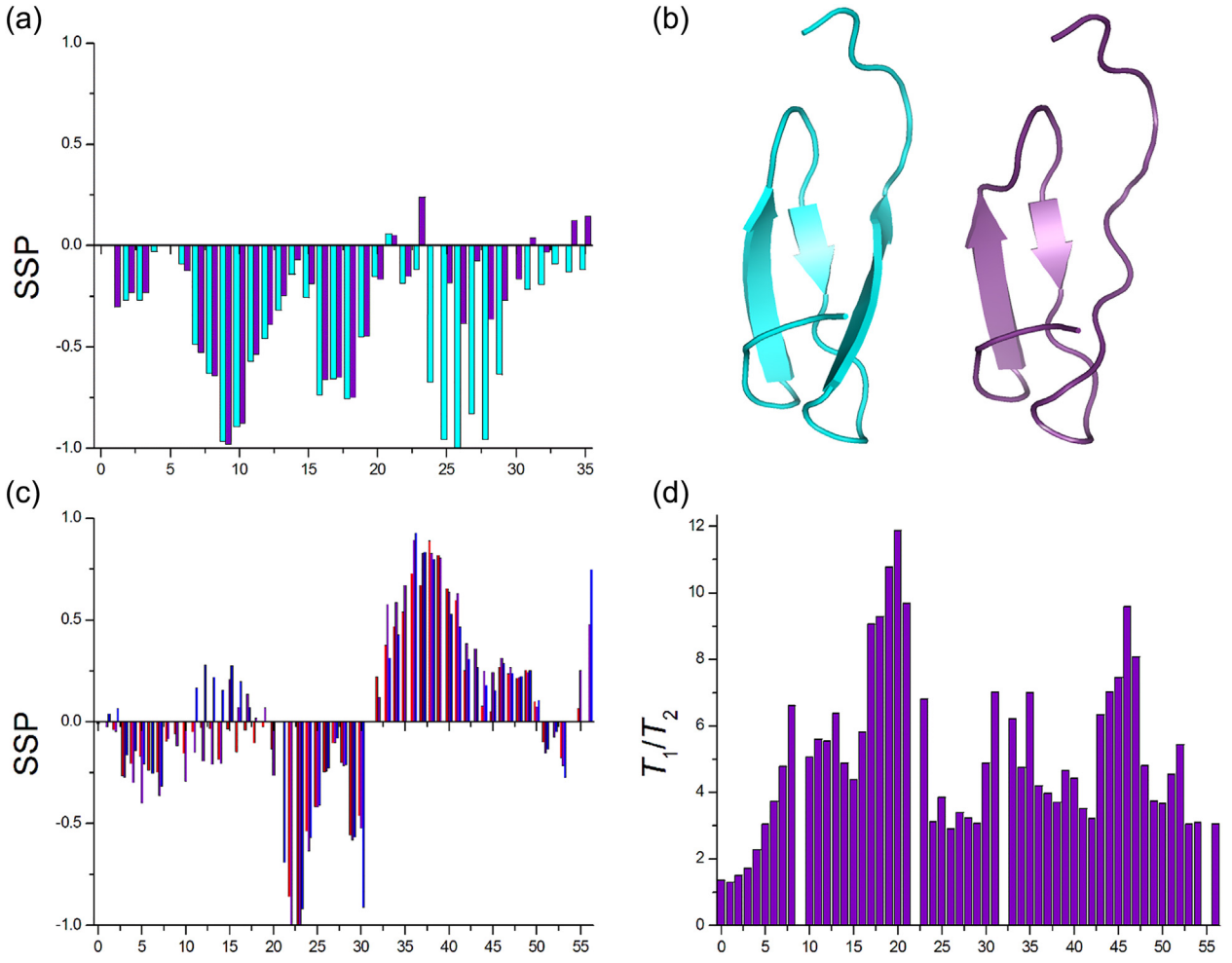

Fig. 4. Structural differences between SGPI-2 and SPINK1 variants detected by NMR spectroscopy. Calculated SSP scores (a) and corresponding tentative structures (b) (based on PDB ID: 1KGM) for SG4 (cyan) and SG5 (violet). SSP scores of SPINK1 variants (c): wild-type SPINK1 (red), SP5 (violet), SP6 (blue) and backbone relaxation data (T_1/T_2 ratio) for SP5 (d). Increased negative SSP values suggest β-strand formation, while increased positive tendency reveals an α-helix. We defined β-strands for regions with SSP < −0.5, indicating that 50% of the conformational ensemble adopts a β-structure at that position [34].

For SG4, the G7–D12, T16–R18 and K24–T29 regions are β-strands and the K13–N15 and S21–G23 parts show turn-like behavior. These motifs are in accordance with previously published data for the structure of wild-type SGPI-2 [23], which consists of a rigid hydrophobic core, 3 antiparallel β-strands (between residues F10–K11, T16–C19 and A26–C28) forming a β-sheet and flexible loops and turns connecting the β-strands.

In contrast, the structure of the loop-exchanged SG5 variant is different from that of the wild-type and SG4. Replacement of the hydrophilic MKMCK loop region of SG4 with the more hydrophobic FILCP sequence in SG5 leads to the disappearance of the third β-strand and the C-terminus becomes unstructured (Fig. 4(a), (b)). These results are consistent with the decrease in β-sheet content indicated by CD spectroscopy. Comparison of the SSP scores for SG4 and SG5 shows no significant structural changes in the N-terminal region; however, the second β-strand of SG5 is slightly less structured than the corresponding region in SG4.

In conclusion, NMR spectroscopy localized and identified the structural changes indicated by CD spectroscopy, namely, the unfolding of the third β-strand in the loop-exchanged variant SG5.

Loop exchange does not alter the overall structure of the SPINK1 scaffold

CD spectroscopy analysis showed very similar secondary structure contents for the wild-type SPINK1, the consensus variant SP5 and the loop-exchanged SP6, while their K_i values for the inhibition of chymotrypsin differ by several orders of magnitude. In order to decipher the reasons for this phenomenon,

structural and dynamical residue-based analyses were performed. Full assignment is provided for wild-type SPINK1 and SP5, while it is 80% complete for SP6 due to significant line broadening (see Supplementary Fig. 6 for the $^{1}H–^{15}N$ HSQC spectra).

Wild-type SPINK1, SP5 and SP6 present similar SSP variations (Fig. 4(c)). The N-terminus (D1–P22), which hosts the enzyme-binding loop, is mostly unstructured, although the L3–A7 region shows a turn-like behavior, followed by a disordered section in the C9–P22 part. For SP5, regions V23–G25 and N29–T30 display two β-strands, and an α-helix is detectable in the V36–R44 region. Based on X-ray crystallography of PSTI-3 [35] and a previous NMR study of PSTI-7 [36], there is a third β-strand in the I50–S53 region, where we only found a minor turn-like propensity for all three studied variants, and this is also valid for analysis of the PSTI-7 data. The structure of the loop-exchanged SP6 variant is similar to that of SP5, but a slightly weaker helical content is detected. In both cases, increased line broadening is observed in the loop part, and in the case of SP6, also at the beginning of the α-helix.

In conclusion, the structures of the three SPINK1 variants are similar, neither the binding loop nor the scaffold shows significant structural changes. This is in accordance with the analysis of the CD spectra of these variants.

Backbone dynamics reveal conformational motions in the canonical loop

As no structural changes were observed in the SPINK1 scaffold to explain the differences in inhibitory effect, ^{15}N backbone dynamic studies were performed on SP5.

T_1 and heteronuclear nOe values provide information on fast scale motions (ns–ps), T_2 values report on slow time scale motions including slower conformational exchanges (Supplementary Fig. 7(a–c)) [35,36]. The motional anisotropy is eliminated if we monitor the T_1/T_2 ratio (Fig. 4(d)). The higher-than-average values, found for the binding loop region and for the end part of the α-helix (K43–S47), are indicative of residues undergoing chemical or conformational exchange.

The experimentally determined T_1 values are fairly constant along the polypeptide chain (average: 0.52 s), with the exception of the highly mobile N-terminus (Supplementary Fig. 7(a)). A disulfide bridge restricts the flexibility of the C-terminus. These variations are reflected in the heteronuclear nOe values as well (Supplementary Fig. 7(c)), where the lowest numbers reveal the most flexible parts of the molecule: the N-terminus and the disordered C9-K21 part, which contains the canonical loop.

In order to highlight the residues undergoing conformational exchange, the relaxation data were further analyzed using the Lipari–Szabo model-free formalism [37,38] for axial anisotropy case. Since there is high structural similarity between Kazal family members, we used the available SPINK1 variant structure [32] to calculate the S^2 order parameters (see Supplementary Fig. 7(d)) and the rotational correlation time (3.1 ns). The obtained S^2 order parameters are lowest for the mobile regions: C9–K21, which includes the canonical loop and Q45–I48, while the highlighted residues are characterized by R_{ex} (1/*s*) values—in accordance with their involvement in conformational motion.

In conclusion, NMR data show that although the loop exchange did not alter the SPINK1 scaffold, it might affect the mobility of the binding loop, where conformational motions are detected. In order to test this possibility and highlight the differences between the flexibility of the variants, MD simulations were performed.

MD simulations highlight differences of the dynamics of SPINK1 variants

All the SPINK1 and SGPI-2 variants were subjected to 1-μs MD in explicit water model. The global SPINK1 scaffold was found to be stable in all variants (Fig. 5(a) and Supplementary Figs. 8 and 9). The three characteristic β-strands (residues 26–28, 33–34 and 53–56) remained unaltered throughout the simulations. The α-helix (residues 37–47) was clearly present in all variants; however, variant-dependent fluctuations in its length were observed. The helix was most stable in wild-type SPINK1 and its SP8 variant (Supplementary Fig. 9), while in SP1, SP4 and SP5, it was more dynamic. The α-helix showed the highest level of fluctuation in the loop-exchanged SP2 and SP6 variants, sometimes resulting in complete helix unfolding. This observation is consistent with the decreased melting temperatures (Table 1) and lower α-helix content (Table 2) of SP2 and SP6. Moreover, NMR experiments also detected signs of helix destabilization upon loop exchange.

The flexibility of the polypeptide chain was calculated as described earlier [40]. Generally, we found that the six N-terminal residues showed highly increased mobility compared to the rest of the molecule in all SPINK1 variants (Fig. 5(a)). This observation is supported by the NMR experiments, which showed that the seven N-terminal residues of SP5 exhibit increased dynamics. To a lesser extent, the C-terminal part of the α-helix also showed increased dynamics (Fig. 5(b), (c)), which was also confirmed for SP5 by NMR.

Comparison of the average structures of the SPINK1 variants shows that the largest conformational differences are mapped to the reactive loop (Fig. 5(a)), which might explain the large diversity of the binding affinity values. Wild-type SPINK1 forms a relatively stable 3_{10}-helix at residues 16–18

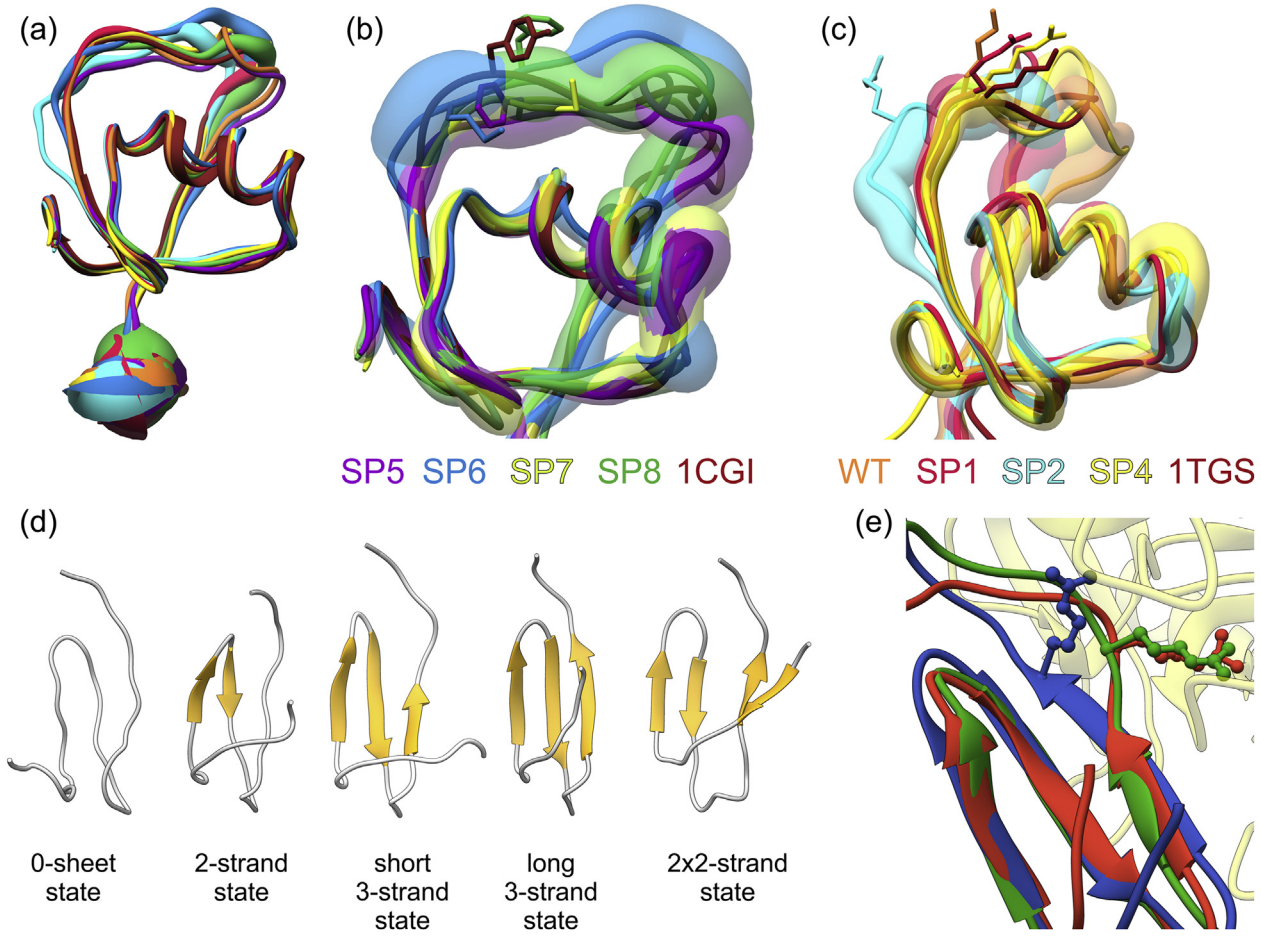

Fig. 5. MD simulations of SPINK1 and SGPI-2 variants. (a) Superimposed representative structures of the SPINK1 variants presented by tube models. The diameter of the tube reflects the β-factor values derived from the local fluctuations (RMSF values) [39]. (b) Representative structures of chymotrypsin specific SPINK1 variants with tubes representing the volume explored by the structural ensemble during the simulations. The X-ray structure of the SPINK1–chymotrypsinogen complex (PDB ID: 1CGI) and the structural models of the variants were superimposed. P1 side chains are represented by sticks. (c) Comparison of the loop conformations of trypsin-specific SPINK1 variants to that of the available SPINK1–trypsinogen complex (PDB ID: 1TGS). P1 side chains are shown with sticks. Note that the molecules in panels b and c are presented in different orientations for better visibility. (d) Representative conformers of SGPI-2 variants differing in the number and extent of β-strands. (e) SGPI-2 third β-strand conformers superimposed to the enzyme–inhibitor complex structure. Green: enzyme–inhibitor complex, red: short third β-strand, blue: long third β-strand state. P1 side chains are shown as ball and stick.

(Supplementary Fig. 8(a)), which is also present in SP5. However, the other variants only show traces of it for short fractions of time and the loop-exchanged SP6 completely lacks this structural feature (Supplementary Fig. 9(a), (b)).

Loop transfer to the SPINK1 scaffold distorts the canonical loop conformation dislocating the P1 side chain

The X-ray structure of the SPINK1–chymotrypsinogen complex (PDB ID: 1CGI) and the representative structures of the chymotrypsin-specific SP5, SP6 and SP8 variants were superimposed (Fig. 5(b)). The P1 side chains of the tight-binding inhibitor variants SP5 and SP8 are located close to that of the wild-type complex and oriented properly, while the P1 of loop-exchanged SP6 and SP7 are shifted either to the location of P1′ or toward P2 of the complex, respectively, suggesting that the loop conformations of SP6 and SP7 are inappropriate for high-affinity binding to the enzyme.

Comparing the loop conformations of trypsin-specific SPINK1 variants to that of the available SPINK1–trypsinogen complex (PDB ID: 1TGS), we found that the P1 side chains of the wild-type SPINK1 and the phage-optimized SP1 and SP4 variants, which are all high-affinity inhibitors, are located in close proximity to the P1 of the complex (Fig. 5(c)). In contrast, the reactive loop of the loop-

exchanged SP6 variant is completely distorted, and the P1 side chain is located far from the favorable position presented in the complex, explaining its poor affinity for the enzyme.

Loop-exchange shifts the distribution of SGPI-2 conformers toward low-affinity forms with altered β-sheet content

MD simulations of the SGPI-2 variants showed highly fluctuating, dynamic structures. We could observe characteristic conformers differing in the number of β-strands, namely structures having no β-strands, two or three β-strands forming one β-sheet, and four β-strands forming two β-sheets (Fig. 5(d)). Generally, the structures of the variants were fluctuating between these states during the MD simulations. The distribution between these conformers, shown in Supplementary Figures 10, 11 and 12, highly depended on the variants.

Assuming that the different conformers exhibit different stability, a function between the distribution of the conformers upon the trajectories and the melting temperatures of the variants was calculated (Supplementary Fig. 12) in a way, that each conformer has a characteristic melting temperature, and the T_m of the SGPI-2 variant is the linear combination of the conformers' melting temperature weighted by the partitioning times. High correlation was found between the distributions of the conformers and the melting temperatures determined by CD spectroscopy, showing that the three-β-strand state is the most stable one with an apparent melting temperature of 88 °C.

Time-averaged secondary structure content was also calculated from the MD trajectories for the SGPI-2 variants (Supplementary Fig. 12). The results correlate well with CD spectroscopy, showing that the loop-exchanged variants SG2 and SG5 have the lowest β-sheet and highest disordered contents, which is also in agreement with the NMR results on SG5.

To assess the inhibitory potential of the SGPI-2 variants, we compared the different conformer structures of the variants to the X-ray structure of the picomolar affinity SGPI-1–trypsin complex (PDB ID: 2XTT). We found that the three-β-strand conformer having a short C-terminal β-strand is the most compatible structure for binding to the enzyme (Fig. 5(e)), and it has a canonical loop conformation matching the one in the tight-binding complex. Thus, we may expect that the portion of time a variant spends in this compatible state should correlate with the value of the inhibition constant. A longer than wild-type third β-strand distorts the loop conformation possibly resulting in weaker binding affinity. The two-strand form, although not ideal, might be flexible enough to accommodate to the enzyme surface upon binding and can still result in relatively high affinity. The four-β-strand form is entirely incompatible with enzyme binding, and the other extreme, the β-strand-free form might be too disordered for high-affinity binding. These assumptions fit the MD results well explaining the low-affinity of SG2 for trypsin and, compared to wild-type SGPI-2 and the SG4 variant, weaker chymotrypsin-binding affinity of SG5 and SG6.

Discussion

The interscaffolding additivity model was well founded by the following observations: (i) the 18 non-related inhibitor scaffolds are evolved to carry a loop having identical, canonical conformation in enzyme complexes; (ii) the canonical loop provides the majority of the binding contacts; and (iii) all 18 families have the same mechanism of action.

The first studied inhibitor–proteinase interactions, for example, bovine pancreatic trypsin inhibitor/trypsin, were described by the lock and key model. Later, it was shown for several cases including the SPGI-2/chymotrypsin interaction that the binding loop is flexible and the canonical conformation becomes stabilized only upon binding to the enzyme [24,41,42]. Also, several proteinases, for example, MASP-2 [21], undergo conformational changes upon complex formation. Nevertheless, along with the intrascaffolding and interscaffolding additivity models, the lock and key model also remained widely accepted. Although the Laskowski group discovered several exemptions from their own additivity rules [6,7,9,15,17,43,44], some backed by structural explanations [43], and while others also reported similar cases [45–47], these were considered as exceptions to the general rule.

In order to settle this uncertainty, we decided to test the validity of the additivity principle by systematically studying the whole binding loop. As the combinatorial sequence space for the six loop positions cannot be tested through individual variants, we used combinatorial mutagenesis and directed protein evolution. We selected SGPI-2 and SPINK1 variants (Fig. 1) for binding to trypsin or chymotrypsin. Affinity selection resulted in picomolar inhibitors and preserved wild-type-like stability and structural features of the proteins (Table 1).

The first important result was that *in vitro* evolution yielded characteristically different, that is, scaffold dependent sequence patterns on the two inhibitors against both enzymes (Fig. 2). Loop swapping allowed for quantitative assessment of this observed interscaffolding non-additivity. The largest non-additivity (186,000-fold affinity drop) was measured when the trypsin-selected consensus SPINK1 loop was transferred onto the SGPI-2 scaffold (SP1 *versus* SG2). This transfer replaced the conserved P2 Thr of SGPI-2 with a Pro erasing the critically

important H-bond network between the loop and the scaffold resulting in non-cooperative unfolding of the SG2 variant.

We introduced the term *direct interscaffolding non-additivity* for this and similar trivial cases when a position has dual roles: (i) binding to the enzyme and (ii) establishing an intramolecular, scaffold-dependent contact.

A good control to this was when the chymotrypsin-selected SPINK1 loop from SP5 was grafted onto SGPI-2 resulting in an only 202-fold affinity drop. In the corresponding SG5 variant, the P2 Thr and the cooperative unfolding mechanism were preserved. The observed affinity drop might be due to the replacement of the optimal P1′ Lys with a disfavored Ile on SGPI-2.

In the simplest case, each side chain from each canonical loop position enters a separate dedicated binding site. However, neither trypsin nor chymotrypsin has separate S1′ and S3′ binding sites as these form a combined S1′/S3′ site that can be reached by side chains both from the P1′ and P3′ positions [48]. For binding to this site, SGPI-2 can use only its P1′, as its P3′ is occupied by a disulfide-forming Cys conserved in the Pacifastin family. At the SGPI-2 P1′, trypsin selected a Met, which is also optimal for trypsin in small synthetic substrates [48], while chymotrypsin preferred the above-mentioned P1′ Lys, which is also the optimal residue for chymotrypsin in small substrates [48]. On SPINK1, the P1′ and P3′ were simultaneously randomized and could compete for the S1′/S3′ binding site. From the two enzymes, only chymotrypsin showed clear positional preference by strongly selecting a Lys or Arg at the P3′ site and being indifferent for residue types at the P1′. It means that on the SPINK1 scaffold, the P3′ is better positioned for binding to the chymotrypsin S1′/S3′ site. The positively charged P3′ residue interacts with Asp64 of chymotrypsin as shown in crystal structures of SPINK1 variants in complex with the enzyme [32]. A similar interaction is seen for another Kazal inhibitor, OMTKY3 [49], and for the unrelated eglin c [50]. When the chymotrypsin-selected SGPI-2 consensus was transferred to SPINK1 (SG4 *versus* SP5) resulting in a 682-fold affinity drop, the SPINK1 optimum P3′ Lys was replaced with an Ala (note that SGPI-2 has a disulfide-forming Cys here), and in a sense, this Lys was relocated from the preferred P3′ to the disfavored P1′ position.

Altogether, this means that the P1′ could freely evolve on both inhibitors, and yet, they utilized it in a scaffold-dependent manner. We call this phenomenon *indirect interscaffolding non-additivity*, as the scaffold-dependent use of the P1′ site is due to a nontrivial combination of three phenomena: (i) a combined binding site on the enzyme that can be reached from more than one loop positions, (ii) enzyme-dependent preference between these loop positions and (iii) scaffold-dependent loop stabilization through one of these positions.

The combination of CD and NMR spectroscopy, DSC and MD simulations provided a consistent model on how directed evolution and subsequent loop exchange affected the overall structure, stability and dynamics of the proteins. CD and NMR spectroscopy showed that directed evolution conserved the wild-type structure in the phage-optimized proteins (SP1 being an exemption with increased α-helix and decreased β-sheet content), loop exchange marginally perturbed the β-sheet of SPINK1 affecting only its α-helix but significantly decreased the β-sheet-content of the SGPI-2 structure up to the point where the third, C-terminal β-strand has been lost. These results are consistent with the thermostability measurements that found SPINK1 more stable than SGPI-2.

Thermodynamic measurements also demonstrated that even though only the α-helix content has changed in the SPINK1 structure, loop exchange caused a marked drop in the stabilities of the corresponding variants, which could be reversed by reinstituting the structurally important P4′ Pro. On the other hand, re-stabilization did not improve inhibitory efficiency.

NMR backbone dynamics revealed conformational motions of the SPINK1 binding loop and the C-terminal part of the α-helix, which can influence affinity. The MD results were consistent with the CD spectroscopy, stability and NMR data and provided additional details about the mechanisms leading to affinity drop upon loop exchange. In essence, MD simulations showed that on the SGPI-2 scaffold, loop exchange dramatically destabilized the β-sheet structure generating a wide variety of non-wild-type conformations incompatible with high-affinity enzyme binding and decreasing thermodynamic stability. In conclusion, the loop, evolved on the SPINK1 scaffold, destabilized the inherently less stable SGPI-2 scaffold.

In contrast, the loop, evolved on the SGPI-2 scaffold, did not alter the overall structure of the high-stability SPINK1 scaffold. Instead, the foreign loop was forced into a position incompatible with high-affinity enzyme binding. The P14′ Gly/Ala mutation in OMTKY3 was reported to cause a similar hinged rotation of the reactive loop relative to the scaffold accompanied with intrascaffolding non-additivity effects [43]. Moreover, it was already reported that loop transfer between Kunitz family members can result in loss of proper intramolecular loop-scaffold stabilizing contacts converting the inhibitor to substrate [45,46].

In this study, we revealed an incompatibility of wild-type scaffolds and “non-self” loops, in which the functional outcome was invariably low affinity, while the structural outcome depended on the inherent stability of the scaffold. Our findings clearly show that

scaffold and binding loop are not autonomous units that could be recombined without negative energetic effects. Instead, the two units cooperate, and this cooperation is scaffold specific. For wild-type inhibitors, the cooperation is inherently achieved by natural evolution, which coevolves these two parts. The same applies for directed evolution, which, even when focusing only on the binding loop region, evolves it in the context of the scaffold and can therefore generate stable, high-affinity variants, many times with significantly altered specificity [20,21,51–61].

We used loop transfer as a research tool to quantitate interscaffolding non-additivity. However, loop transfer is also an important practical tool in developing therapeutic monoclonal antibodies. Humanized antibodies are generated by CDR transfer from mouse antibodies to a human scaffold. This, in most cases, results in significant affinity drop and is therefore usually followed or even bypassed by structure-based directed evolution using phage display, approaches and technologies recently acknowledged by a Nobel prize in chemistry for their utmost importance [62–66].

There is an interesting topological relationship between antibody humanization and loop transfer between unrelated canonical inhibitors. In humanization, the donor and acceptor scaffolds are homologous, while the loops exchanged are unrelated representing an intra-scaffold scenario. Structural similarity of the scaffolds should facilitate maintaining the conformation and binding properties of the transferred loops. Moreover, the antibody scaffolds had been evolved to carry many different loops. The case of canonical inhibitors is just the opposite representing an inter-scaffold scenario: loops having almost identical conformations are exchanged between unrelated scaffolds that had been evolved to carry these almost identical loops.

The results of our study unequivocally demonstrate that interscaffolding additivity, a non-evident extrapolation of the Laskowski mechanism, is not a valid model for canonical inhibitor action. This should be valuable information for those aiming to develop serine proteinase inhibitors with novel specificities either for academic research or therapeutic use.

Materials and Methods

Reagents and bacterial strains

Restriction enzymes were from Thermo Scientific and New England Biolabs, trypsin and chymotrypsin used for phage display selections and phage ELISA were from Sigma Aldrich, and trypsin and chymotrypsin used for K_I determinations were from Sigma Aldrich and Worthington, respectively. DNA sequencing was done by the ABI PRISM BigDye v3.1 Kit. The following *Escherichia coli* strains were used: XL1-Blue (Stratagene), SS320 (Lucigen), CJ236 and SHuffle T7 Express (New England Biolabs). The M13KO7 helper phage was from New England Biolabs.

Library construction

The synthetic gene encoding human SPINK1 (UniProt: P00995) was purchased from Genscript and was subcloned into a phagemid vector to be fused to the M13 p8 coat protein gene. Then, six codons corresponding to binding loop positions P2–P4′ were replaced with stop codons by Kunkel mutagenesis. Next, in a combinatorial mutagenesis step, the stop codons were replaced by NNK codons encoding all 20 amino acids to produce the SPINK1-phagemid library.

The SGPI-2 library was based on the Tag-SGPI-2-pGP8 phagemid vector [19] and created exactly as described in Refs. [21,51,52]. The P4, P2, P1, P1′, P2′ and P4′ positions were fully randomized, while the disulfide-bonded P3 and P3′ Cys residues were kept as wild-type.

The phagemid libraries were electroporated into *E. coli* SS3320 cells and superinfected by M13KO7 helper phage to produce the inhibitor-phage libraries.

Selection of the SGPI-2 and SPINK1 libraries against bovine trypsin and chymotrypsin by phage display

Phage display selections were carried out as described in Ref. [67]. Briefly, bovine trypsin or chymotrypsin was immobilized on MaxiSorp (Nunc) plates. Phage particles were obtained with the PEG/NaCl precipitation method and applied to the blocked wells. Following thorough washing, bound phages were eluted. *E. coli* XL1-Blue was used to amplify the eluted phages to be used in a next round of selection. After three selection and amplification cycles, the libraries were enriched over 100-fold compared to the control obtained on blocked wells containing no target enzymes.

Phage ELISA and sequence analysis

Individual clones from the third round of selection were tested for binding to their cognate enzyme in phage ELISA experiments, and positive clones were sequenced (Supplementary Table 1) as described in Ref. [67]. Unique DNA sequences were used to build codon-normalized amino acid sequence logos (Fig. 2) by the Weblogo program [68] as described in Ref. [51].

Expression and purification of inhibitor variants

Six SGPI-2 and eight SPINK1 variants (Table 1) were designed. The genes encoding these variants

were created using either PCR-based or Kunkel mutagenesis. The variants were produced along with wild-type SGPI-2 and SPINK1 as described earlier [52]. The inhibitors were expressed in *E. coli* SHuffle T7 Express fused to the His-tagged disulfide isomerase DsbC, purified by Ni^{2+} affinity chromatography and cleaved off their fusion partner by tobacco etch virus protease. Afterward, DsbC was allowed to correct any non-native disulfides. His-tagged tobacco etch virus protease and DsbC were removed in a second affinity chromatography step. The inhibitors were purified to homogeneity by reversed phase HPLC.

^{15}N-labeled SP5, ^{15}N-labeled SP6 and ^{13}C,^{15}N-double-labeled SP5 were produced for subsequent NMR measurements. SHuffle T7 Express cells transformed with the appropriate expression plasmid were grown in LB medium to OD_{600} = 0.8, washed two times in 10× concentrated M9 minimal salts and suspended in M9 minimal medium containing ^{15}N NH_4Cl (Cambridge Isotope Laboratories, NLM-467-1) or ^{15}N NH_4Cl and U-13C6 D-glucose (Cambridge Isotope Laboratories, CLM-1396-1), respectively. Expression was induced with 500 μM IPTG and lasted 6 h. Purification of the labeled inhibitors was carried out as already described for the unlabeled proteins.

Determination of protein concentrations

Bovine trypsin and bovine chymotrypsin were active-site titrated according to Jameson *et al*. [69]. The concentration of SPINK1, SP1, SP4 and SG1 was determined by titration against trypsin using *N*-benzoyl-DL-arginine-4-nitroanilide. Wild-type SGPI-2, SG4, SG6, SP5 and SP8 were titrated against chymotrypsin using Suc-Ala-Ala-Pro-Phe-pNA. Because of the low affinity of the remaining variants, their concentration was measured using Micro BCA Protein Assay Kit (Thermo Fisher Scientific) using active-site titrated SP4 and SG4 as internal standards for SPINK1 and SGPI-2 variants, respectively.

Equilibrium binding assays

The binding affinity of the variants for their cognate enzymes was determined by measuring the equilibrium inhibitory constants (K_I) according to Empie and Laskowski [4]. Increasing amounts of inhibitor were pre-incubated with fixed concentrations of the enzyme until equilibrium was reached. Suc-Ala-Ala-Pro-Phe-AMC or Z-Gly-Pro-Arg-AMC was added to measure the residual activity of chymotrypsin or trypsin, respectively. The following equation was fitted to the data: $y = E - (E + x + K - \mathrm{sqrt}((E + x + K)^2 - 4Ex))/2$, where x designates the total inhibitor concentration, y represents the free protease concentration in equilibrium, K is K_I, and E stands for the total protease concentration. K_I values presented here (Table 1) represent the average of three measurements.

Comparative protease inhibition assays using disulfide-reduced and native inhibitors

Chymotrypsin inhibiting variants SG4, SG5, SP5 and SP6 were treated with 50-fold excess of TCEP (Tris(2-carboxyethyl)phosphine) for 20 min in 20 mM Tris–HCl, 10 mM $CaCl_2$ and 0.005% Triton X-100 (pH 8.0) buffer. Two micromolars of treated SG4, SG5 or SP5 was incubated with 50 nM chymotrypsin, and 5 μM of treated SP6 was incubated with 10 nM chymotrypsin for 10 min. Aliquots of the mixtures were removed for subsequent mass spectrometry analysis. The pH of these samples was immediately set to 2.5 with formic acid to stop the enzymatic reactions, and the aliquots were stored at −20 °C. Residual enzyme activity in the mixtures was determined spectrophotometrically by adding Suc-AAPF-pNA to a final concentration of 500 μM. Control reactions contained the same concentrations of untreated inhibitors. For inhibitor samples not treated with TCEP, 100% enzyme activity was defined by using a chymotrypsin containing sample free of TCEP. For TCEP-treated samples, 100% enzyme activity was defined by measuring chymotrypsin activity in the presence of 100 μM TCEP for SG4, SG5 and SP5 or 250 μM TCEP for SP6.

LC-MS analysis of the protease treated inhibitors

LC-MS analysis was performed on a Thermo Scientific Q Exactive Focus, high-resolution and high-mass accuracy, hybrid quadrupole-orbitrap mass spectrometer (Bremen, Germany) using online UPLC coupling. UPLC separation was performed on a Dionex 3000 UPLC system using a Waters Acquity C18 column (2.1 × 150 mm, 1.7 μm). Linear gradient elution (0 min 2% B, 1 min 2% B, 17 min 90% B) with eluent A (0.1% formic acid in water, v/v) and eluent B (0.1% formic acid in acetonitrile/water, 80:20, v/v) was used at a flow rate of 0.3 mL/min at 40 °C. High-resolution mass spectra were acquired in the 300–1600 *m*/*z* range.

CD spectroscopy

CD spectra of the wild-type SPINK1 and SGPI-2, trypsin-binding variants SG1, SG2, SP1 and SP2 and chymotrypsin-binding variants SG4, SG5, SP6 and SP7 were recorded between 180 and 260 nm on a Jasco J-810 spectropolarimeter at 10-nm/min scanning speed with 1-nm band width, 0.2-nm data pitch and 8-s response time. Three scans were accumulated. SG1, SG2, SP1 and SP2 were measured at 2.5-12.2 mg/ml in 20 mM Na-phosphate buffer at pH 7.0, while the wild-type proteins and SG4, SG5, SP5 and SP6 were measured in 10 mM MES and 50 mM NaCl

(pH 6.5) buffer at 4.6-9.7 mg/ml in a 0.001-cm quartz cell. Data were analyzed with the BeStSel program [33].

Besides the high-protein concentration measurements, CD spectra of the latter four mutants were also determined at 0.1 mg/ml concentration in 5 mM MES and 5 mM NaCl (pH 6.5) buffer in a 0.1-cm cuvette. Comparison of the corresponding spectra confirmed that increased protein concentrations, used in the subsequent NMR experiments, do not result in any detectable aggregation or structural change.

CD spectrum of wild-type SPINK1 was also recorded at pH 3.5, 4.5 and 5.5 to test for any structural changes at low pH values to be used in the NMR experiments. Measurements were carried out at 7 mg/ml in 50 mM Na-phosphate buffer in a 0.001-cm quartz cuvette. Comparison of these spectra to the spectrum recorded at pH 6.5 indicates that no structural changes take place between pH 3.5 and 6.5.

Thermostability measurements

CD spectroscopy was used to characterize the thermostability of wild-type SGPI-2 and its variants. Thermal denaturation was monitored at 202 nm in a Jasco J-810 spectropolarimeter equipped with a PTC-514 temperature controller. Protein samples of 0.1 mg/ml in 20 mM Na-phosphate and 20 mM NaCl (pH 6.5) buffer were heated from 10 to 110 °C with a slope of 2 °C/min in a 0.1-cm cuvette. Data were recorded with 2-nm band width, 0.2 °C data pitch and a response time of 16 s. The equation describing two-state thermal denaturation [70] was fitted to the data except for SG2, which did not show cooperative denaturation. ΔC_p was fixed at 1000 J mol^{-1} K^{-1}. T_m values of the variants are provided in Table 1.

Thermal denaturation of wild-type SPINK1 and its variants was not accompanied with a sigmoidal signal change in the far UV CD spectra. Therefore, for these variants, and also for SG2, DSC was used. The experiments were carried out using a MicroCal VP-DSC calorimeter at 0.2–0.5 mg/ml protein concentrations in 20 mM Na-phosphate and 20 mM NaCl (pH 6.5) buffer. The samples were heated from 7 to 125 °C at 1.5 °C/min, cooled down and re-scanned. The denaturation turned out to be irreversible for each variant, which allowed us to use the second runs as references. MicroCal ITC Origin software was used to calculate T_m values by fitting a non-two-state model (Table 1). Aggregation of the proteins at high temperatures thwarted accurate determination of the enthalpy of denaturation.

Structural changes in SG2, SP2 and SP8 induced by increasing the temperature were also followed by recording the CD spectra in the 190- to 260-nm range at 10 °C intervals for SG2 and SP2 and only at three temperatures, 10, 65 and 108 °C, for SP8.

NMR measurements

Typical sample composition of the unlabeled SGPI-2 variants SG4 and SG5 was 1 mM protein in 10 mM MES buffer and 50 mM NaCl at pH 6.5. Similar composition was used for the ^{15}N-labeled SPINK1 variants SP5 and SP6, measured at pH 3.0 and pH 6.5. The 0.5 mM ^{13}C, ^{15}N-labeled SP5 variant was studied at pH 3.1 and 4.1 in 50 mM phosphate buffer. The wild-type SPINK1 variant sample contained 1 mM protein in 50 mM phosphate buffer at pH 5.5 and 3.5. All samples contained 10% D_2O and DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) as an internal chemical shift reference.

NMR spectra were recorded on a Bruker Avance III 700 spectrometer operating at 700.05 MHz using a 5-mm room-temperature TXI, z-gradient probehead, and a Prodigy TCI H&F-C/N-D, z-gradient probehead. ^{1}H chemical shifts were referenced to the internal DSS standard, whereas ^{15}N and ^{13}C chemical shifts were referenced indirectly via the gyromagnetic ratios. Temperature was calibrated against the methanol standard sample; measurements were done at 298 K.

For unlabeled SGPI-2 variants SG4 and SG5, resonance assignment and sequential connectivities were determined from 2D homonuclear (TOCSY, NOESY, COSY) measurements. For the ^{15}N-labeled SPINK1 variants SP5 and SP6, 3D HSQC–TOCSY and HSQC–NOESY (mixing time = 120 ms) were acquired, while for ^{13}C,^{15}N-labeled SP5, sequence-specific assignment of HN, N, C′, Cα, Cβ and further side-chain resonances were obtained from BEST type 3D-HNCA, HN(CO)CA, HNCACB, HN(CO)CACB, HN(CA)CO, HNCO and standard CCCONH measurements. T_1 relaxation time measurements were measured in 11 steps with a variable delay between 0.1 and 4 s, while T_2 relaxation time measurements were measured in 12 steps using delays between 0.017 and 1.018 s; in both cases, one delay has been recorded twice to check data reproducibility. ^{1}H–^{15}N heteronuclear nOe measurements were carried out with and without proton saturation using a 7-s relaxation delay. Relaxation times were evaluated from single exponential decays using peak intensities, and heteronuclear nOes were calculated from the intensity ratios of individual peaks with and without proton saturation. All spectra were processed with TopSpin and analyzed with CARA [71], Sparky [72], Dynamics Center and CCPN software [73]. Backbone relaxation data were further analyzed by FAST-Modelfree [74].

MD simulations

SGPI-2 and SPINK1 variants were subjected to MD simulations as implemented in GROMACS [40] using the AMBER-ff99SB*-ILDNP force field [75].

For SGPI-2 variants, the starting models were prepared on the basis of the PDB ID: 1KGM NMR structure of Gáspári *et al*. [23] using SwissPDBViewer [76] and Chimera for side-chain replacements and initial energy minimization. In the case of the SPINK1 scaffold, the PDB ID: 1HPT X-ray structure [35] served as a basis. For MD, the system was solvated by water molecules with TIP4P parametrization [77]. The total charge of the system was neutralized, and the physiological salt concentration was set by placing Na^+ and Cl^- ions. Energy minimization of starting structures was followed by sequential relaxation of constraints on protein atoms in three steps and an additional NVT step (all for 200 ps) to stabilize pressure. Pne-microsecond trajectories of NPT simulations at 300 K at 1 bar were recorded (collecting snapshots at every 20 ps). The first 100 ns of the simulations were taken as final equilibration, and the 100-ns to 1-μs part was used for evaluation. Secondary structure compositions of the frames of the trajectories were determined by the DSSP algorithm [78]. Molecular graphics were performed with the UCSF CHIMERA package (University of California, San Francisco) [79].

Accession numbers

NMR assignment data have been deposited in the BioMagResBank with accession number 27491 for SG4, 27492 for SG5, 27500 for SP5, 27501 for SP6 and 12020 for wild-type SPINK1.

Acknowledgments

This work was supported by the National Research, Development and Innovation Office/Hungarian Scientific Research

Fund Grants K119386, K120391, K124900 and KH125597; by the European Union and the State of Hungary and co-financed by the European Regional Development Fund within the projects VEKOP-2.3.2-16-2017-00014 and VEKOP-2.3.3-15-2017-00020; and by the Hungarian Ministry of Human Capacities in the frame of the ELTE Institutional Excellence Program (783-3/2018/FEKUTSRAT), as well as by the MedInProt Protein Science Research Synergy Program of the Hungarian Academy of Sciences and by the National Development Agency Grant KMOP-4.2.1/B-10-2011. A.M. was supported by the János Bolyai Scholarship of the Hungarian Academy of Sciences.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2018.12.003.

Received 8 July 2018;
Received in revised form 26 November 2018;
Accepted 6 December 2018
Available online 11 December 2018

Keywords:
directed evolution;
phage display;
molecular recognition;
protein–protein interaction;
SPINK1

Present address: D. Héja, Translational Transplant Research Center, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

Present address: K. Zboray, Plant Protection Institute, Centre for Agricultural Research, Hungarian Academy of Sciences, Martonvásár, Hungary.

Abbreviations used:
MASP, mannan-binding lectin serine protease; OMTKY-3, turkey ovomucoid third domain; SGPI-2, *Schistocerca gregaria* protease inhibitor 2; SPINK1, serine protease inhibitor Kazal type 1.

References

[1] M. Laskowski, I. Kato, W. Ardelt, J. Cook, A. Denton, M.W. Empie, W.J. Kohr, S.J. Park, K. Parks, B.L. Schatzley, Ovomucoid third domains from 100 avian species: isolation, sequences, and hypervariability of enzyme–inhibitor contact residues, Biochemistry 26 (1987) 202–221.

[2] M. Laskowski, I. Apostol, W. Ardelt, J. Cook, A. Giletto, C.A. Kelly, W.Y. Lu, S.J. Park, M.A. Qasim, H.E. Whatley, Amino acid sequences of ovomucoid third domain from 25 additional species of birds, J. Protein Chem. 9 (1990) 715–725.

[3] I. Apostol, A. Giletto, T. Komiyama, W. Zhang, M. Laskowski, Amino acid sequences of ovomucoid third domains from 27 additional species of birds, J. Protein Chem. 12 (1993) 419–433.

[4] M.W. Empie, M. Laskowski Jr., Thermodynamics and kinetics of single residue replacements in avian ovomucoid third domains: effect on inhibitor interactions with serine proteinases, Biochemistry 21 (1982) 2274–2284.

[5] T.L. Bigler, W. Lu, S.J. Park, M. Tashiro, M. Wieczorek, R. Wynn, M. Laskowski, Binding of amino acid side chains to preformed cavities: interaction of serine proteinases with turkey ovomucoid third domains with coded and noncoded P1 residues, Protein Sci. 2 (1993) 786–799.

[6] W. Lu, I. Apostol, M.A. Qasim, N. Warne, R. Wynn, W.L. Zhang, S. Anderson, Y.W. Chiang, E. Ogin, I. Rothberg, K. Ryan, M. Laskowski Jr., Binding of amino acid side-chains to S1 cavities of serine proteinases, J. Mol. Biol. 266 (1997) 441–461, https://doi.org/10.1006/jmbi.1996.0781.

[7] S.M. Lu, W. Lu, M.A. Qasim, S. Anderson, I. Apostol, W. Ardelt, T. Bigler, Y.W. Chiang, J. Cook, M.N.G. James, I. Kato, C. Kelly, W. Kohr, T. Komiyama, T.-Y. Lin, M. Ogawa, J. Otlewski, S.-J. Park, S. Qasim, M. Ranjbar, M. Tashiro, N. Warne, H. Whatley, A. Wieczorek, M. Wieczorek, T. Wilusz, R. Wynn, W. Zhang, M. Laskowski, Predicting the reactivity of

proteins from their sequence alone: Kazal family of protein inhibitors of serine proteinases, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 1410–1415.
[8] I. Schechter, A. Berger, On the size of the active site in proteases. I. Papain, Biochem. Biophys. Res. Commun. 27 (1967) 157–162.
[9] M.A. Qasim, P.J. Ganz, C.W. Saunders, K.S. Bateman, M.N. James, M. Laskowski Jr., Interscaffolding additivity. Association of P1 variants of eglin c and of turkey ovomucoid third domain with serine proteinases, Biochemistry 36 (1997) 1598–1607, https://doi.org/10.1021/bi9620870.
[10] P.J. Carter, G. Winter, A.J. Wilkinson, A.R. Fersht, The use of double mutants to detect structural changes in the active site of the tyrosyl-tRNA synthetase (*Bacillus stearothermophilus*), Cell 38 (1984) 835–840.
[11] A. Horovitz, Non-additivity in protein–protein interactions, J. Mol. Biol. 196 (1987) 733–735.
[12] J.A. Wells, Additivity of mutational effects in proteins, Biochemistry 29 (1990) 8509–8517.
[13] C.A. Kelly, M. Laskowski Jr., M.A. Qasim, The role of scaffolding in standard mechanism serine proteinase inhibitors, Protein Pept. Lett. 12 (2005) 465–471.
[14] M. Laskowski Jr., M.A. Qasim, Z. Yi, Additivity-based prediction of equilibrium constants for some protein–protein associations, Curr. Opin. Struct. Biol. 13 (2003) 130–139.
[15] M.A. Qasim, W. Lu, M. Lu, M. Ranjbar, Z. Yi, Y.-W. Chiang, K. Ryan, S. Anderson, W. Zhang, S. Qasim, M. Laskowski, Testing of the additivity-based protein sequence to reactivity algorithm, Biochemistry 42 (2003) 6460–6466, https://doi.org/10.1021/bi027186u.
[16] M.A. Qasim, L. Wang, S. Qasim, S. Lu, W. Lu, R. Wynn, Z.-P. Yi, M. Laskowski, Additivity-based design of the strongest possible turkey ovomucoid third domain inhibitors for porcine pancreatic elastase (PPE) and *Streptomyces griseus* protease B (SGPB), FEBS Lett. 587 (2013) 3021–3026, https://doi.org/10.1016/j.febslet.2013.07.029.
[17] A. Grzesiak, R. Helland, A.O. Smalås, D. Krowarsch, M. Dadlez, J. Otlewski, Substitutions at the P1′ position in BPTI strongly affect the association energy with serine proteinases, J. Mol. Biol. 301 (2000) 205–217, https://doi.org/10.1006/jmbi.2000.3935.
[18] D. Krowarsch, M. Dadlez, O. Buczek, I. Krokoszynska, A.O. Smalas, J. Otlewski, Interscaffolding additivity: binding of P1 variants of bovine pancreatic trypsin inhibitor to four serine proteases, J. Mol. Biol. 289 (1999) 175–186, https://doi.org/10.1006/jmbi.1999.2757.
[19] B. Szenthe, A. Patthy, Z. Gáspári, A.K. Kékesi, L. Gráf, G. Pál, When the surface tells what lies beneath: combinatorial phage-display mutagenesis reveals complex networks of surface-core interactions in the pacifastin protease inhibitor family, J. Mol. Biol. 370 (2007) 63–79, https://doi.org/10.1016/j.jmb.2007.04.029.
[20] A. Kocsis, K.A. Kékesi, R. Szász, B.M. Végh, J. Balczer, J. Dobó, P. Závodszky, P. Gál, G. Pál, 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. 1950 (185) (2010) 4169–4178, https://doi.org/10.4049/jimmunol.1001819.
[21] D. Héja, V. Harmat, K. Fodor, M. Wilmanns, J. Dobó, K.A. Kékesi, P. Závodszky, P. Gál, G. Pál, Monospecific inhibitors show that both mannan-binding lectin-associated serine protease (MASP)-1 and -2 are essential for lectin pathway activation and reveal structural plasticity of MASP-2, J. Biol. Chem. (2012)https://doi.org/10.1074/jbc.M112.354332.
[22] L. Hedstrom, Serine protease mechanism and specificity, Chem. Rev. 102 (2002) 4501–4524, https://doi.org/10.1021/cr000033x.
[23] Z. Gáspári, A. Patthy, L. Gráf, A. Perczel, Comparative structure analysis of proteinase inhibitors from the desert locust, *Schistocerca gregaria*, Eur. J. Biochem. FEBS 269 (2002) 527–537.
[24] A. Roussel, M. Mathieu, A. Dobbs, B. Luu, C. Cambillau, C. Kellenberger, Complexation of two proteic insect inhibitors to the active site of chymotrypsin suggests decoupled roles for binding and selectivity, J. Biol. Chem. 276 (2001) 38893–38898, https://doi.org/10.1074/jbc.M105707200.
[25] B. Breugelmans, G. Simonet, V. van Hoef, S. Van Soest, J. Vanden Broeck, Pacifastin-related peptides: structural and functional characteristics of a family of serine peptidase inhibitors, Peptides 30 (2009) 622–632, https://doi.org/10.1016/j.peptides.2008.07.026.
[26] B. Szenthe, Z. Gáspári, A. Nagy, A. Perczel, L. Gráf, Same fold with different mobility: backbone dynamics of small protease inhibitors from the desert locust, *Schistocerca gregaria*, Biochemistry 43 (2004) 3376–3384 [doi:10.1021/bi035689+].
[27] K. Fodor, V. Harmat, R. Neutze, L. Szilágyi, L. Gráf, G. Katona, Enzyme:substrate hydrogen bond shortening during the acylation phase of serine protease catalysis, Biochemistry 45 (2006) 2114–2121, https://doi.org/10.1021/bi0517133.
[28] L.J. Greene, M.H. Pubols, D.C. Bartelt, Human pancreatic secretory trypsin inhibitor, Methods Enzymol. 45 (1976) 813–825.
[29] E. Antonini, P. Ascenzi, M. Bolognesi, G. Gatti, M. Guarneri, E. Menegatti, Interaction between serine (pro)enzymes, and Kazal and Kunitz inhibitors, J. Mol. Biol. 165 (1983) 543–558, https://doi.org/10.1016/S0022-2836(83)80219-8.
[30] Z. Malik, S. Amir, G. Pál, Z. Buzás, E. Várallyay, J. Antal, Z. Szilágyi, K. Vékey, B. Asbóth, A. Patthy, L. Gráf, Proteinase inhibitors from desert locust, *Schistocerca gregaria*: engineering of both P(1) and P(1)' residues converts a potent chymotrypsin inhibitor to a potent trypsin inhibitor, Biochim. Biophys. Acta 1434 (1999) 143–150.
[31] T. Lauber, A. Schulz, K. Schweimer, K. Adermann, U.C. Marx, Homologous proteins with different folds: the three-dimensional structures of domains 1 and 6 of the multiple Kazal-type inhibitor LEKTI, J. Mol. Biol. 328 (2003) 205–219.
[32] H.J. Hecht, M. Szardenings, J. Collins, D. Schomburg, Three-dimensional structure of the complexes between bovine chymotrypsinogen A and two recombinant variants of human pancreatic secretory trypsin inhibitor (Kazal-type), J. Mol. Biol. 220 (1991) 711–722.
[33] A. Micsonai, F. Wien, L. Kernya, Y.-H. Lee, Y. Goto, M. Réfrégiers, J. Kardos, Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy, Proc. Natl. Acad. Sci. U. S. A. 112 (2015) E3095–E3103, https://doi.org/10.1073/pnas.1500851112.
[34] J.A. Marsh, V.K. Singh, Z. Jia, J.D. Forman-Kay, Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation, Protein Sci. 15 (2006) 2795–2804, https://doi.org/10.1110/ps.062465306.
[35] H.J. Hecht, M. Szardenings, J. Collins, D. Schomburg, Three-dimensional structure of a recombinant variant of human pancreatic secretory trypsin inhibitor (Kazal type), J. Mol. Biol. 225 (1992) 1095–1103.
[36] W. Klaus, D. Schomburg, Solution structure of a variant of human pancreatic secretory trypsin inhibitor determined by

nuclear magnetic resonance spectroscopy, J. Mol. Biol. 229 (1993) 695–706, https://doi.org/10.1006/jmbi.1993.1073.
[37] G. Lipari, A. Szabo, Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity, J. Am. Chem. Soc. 104 (1982) 4546–4559, https://doi.org/10.1021/ja00381a009.
[38] G. Lipari, A. Szabo, Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results, J. Am. Chem. Soc. 104 (1982) 4559–4570, https://doi.org/10.1021/ja00381a010.
[39] G. Bange, N. Kümmerer, C. Engel, G. Bozkurt, K. Wild, I. Sinning, FlhA provides the adaptor for coordinated delivery of late flagella building blocks to the type III secretion system, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 11295–11300, https://doi.org/10.1073/pnas.1001383107.
[40] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M.R. Shirts, J.C. Smith, P.M. Kasson, D. van der Spoel, B. Hess, E. Lindahl, GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit, Bioinformatics 29 (2013) 845–854, https://doi.org/10.1093/bioinformatics/btt055.
[41] Z. Gáspári, P. Várnai, B. Szappanos, A. Perczel, Reconciling the lock-and-key and dynamic views of canonical serine protease inhibitor action, FEBS Lett. 584 (2010) 203–206, https://doi.org/10.1016/j.febslet.2009.11.058.
[42] K. Fodor, V. Harmat, C. Hetényi, J. Kardos, J. Antal, A. Perczel, A. Patthy, G. Katona, L. Gráf, Extended intermolecular interactions in a serine protease-canonical inhibitor complex account for strong and highly specific inhibition, J. Mol. Biol. 350 (2005) 156–169, https://doi.org/10.1016/j.jmb.2005.04.039.
[43] T.-W. Lee, M.A. Qasim, M. Laskowski, M.N.G. James, Structural insights into the non-additivity effects in the sequence-to-reactivity algorithm for serine peptidases and their inhibitors, J. Mol. Biol. 367 (2007) 527–546, https://doi.org/10.1016/j.jmb.2007.01.008.
[44] O. Buczek, K. Koscielska-Kasprzak, D. Krowarsch, M. Dadlez, J. Otlewski, Analysis of serine proteinase–inhibitor interaction by alanine shaving, Protein Sci. 11 (2002) 806–819.
[45] S. Khamrui, S. Majumder, J. Dasgupta, J.K. Dattagupta, U. Sen, Identification of a novel set of scaffolding residues that are instrumental for the inhibitory property of Kunitz (STI) inhibitors, Protein Sci. 19 (2010) 593–602, https://doi.org/10.1002/pro.338.
[46] S. Majumder, S. Khamrui, J. Dasgupta, J.K. Dattagupta, U. Sen, Role of remote scaffolding residues in the inhibitory loop pre-organization, flexibility, rigidification and enzyme inhibition of serine protease inhibitors, Biochim. Biophys. Acta 1824 (2012) 882–890, https://doi.org/10.1016/j.bbapap.2012.04.009.
[47] M.A. Salameh, A.S. Soares, D. Navaneetham, D. Sinha, P.N. Walsh, E.S. Radisky, Determinants of affinity and proteolytic stability in interactions of Kunitz family protease inhibitors with mesotrypsin, J. Biol. Chem. 285 (2010) 36884–36896, https://doi.org/10.1074/jbc.M110.171348.
[48] V. Schellenberger, C.W. Turck, W.J. Rutter, Role of the S′ subsites in serine protease catalysis. Active-site mapping of rat chymotrypsin, rat trypsin, alpha-lytic protease, and cercarial protease from *Schistosoma mansoni*, Biochemistry 33 (1994) 4251–4257.
[49] M. Fujinaga, A.R. Sielecki, R.J. Read, W. Ardelt, M. Laskowski, M.N. James, Crystal and molecular structures of the complex of alpha-chymotrypsin with its inhibitor turkey ovomucoid third domain at 1.8 Å resolution, J. Mol. Biol. 195 (1987) 397–418.
[50] F. Frigerio, A. Coda, L. Pugliese, C. Lionetti, E. Menegatti, G. Amiconi, H.P. Schnebli, P. Ascenzi, M. Bolognesi, Crystal and molecular structure of the bovine alpha-chymotrypsin-eglin c complex at 2.0 Å resolution, J. Mol. Biol. 225 (1992) 107–123.
[51] A. Szabó, D. Héja, D. Szakács, K. Zboray, K.A. Kékesi, E.S. Radisky, M. Sahin-Tóth, G. Pál, High affinity small protein inhibitors of human chymotrypsin C (CTRC) selected by phage display reveal unusual preference for P4′ acidic residues, J. Biol. Chem. 286 (2011) 22535–22545, https://doi.org/10.1074/jbc.M111.235754.
[52] E. Boros, A. Szabó, K. Zboray, D. Héja, G. Pál, M. Sahin-Tóth, Overlapping specificity of duplicated human pancreatic elastase 3 isoforms and archetypal porcine elastase 1 provides clues to evolution of digestive enzymes, J. Biol. Chem. 292 (2017) 2690–2702, https://doi.org/10.1074/jbc.M116.770560.
[53] H.R. Maun, C. Eigenbrot, R.A. Lazarus, Engineering exosite peptides for complete inhibition of factor VIIa Using a protease switch with substrate phage, J. Biol. Chem. 278 (2003) 21823–21830, https://doi.org/10.1074/jbc.M300951200.
[54] L. Kiczak, M. Kasztura, K. Koscielska-Kasprzak, M. Dadlez, J. Otlewski, Selection of potent chymotrypsin and elastase inhibitors from M13 phage library of basic pancreatic trypsin inhibitor (BPTI), Biochim. Biophys. Acta 1550 (2001) 153–163.
[55] C.I. Wang, Q. Yang, C.S. Craik, Isolation of a high affinity inhibitor of urokinase-type plasminogen activator by phage display of ecotin, J. Biol. Chem. 270 (1995) 12250–12256.
[56] M.S. Dennis, A. Herzka, R.A. Lazarus, Potent and selective Kunitz domain inhibitors of plasma kallikrein designed by phage display, J. Biol. Chem. 270 (1995) 25411–25417.
[57] N. Dimasi, F. Martin, C. Volpari, M. Brunetti, G. Biasiol, S. Altamura, R. Cortese, R. De Francesco, C. Steinkühler, M. Sollazzo, Characterization of engineered hepatitis C virus NS3 protease inhibitors affinity selected from human pancreatic secretory trypsin inhibitor and minibody repertoires, J. Virol. 71 (1997) 7461–7469.
[58] A.S. Tanaka, M.M. Silva, R.J. Torquato, M.A. Noguti, C.A. Sampaio, H. Fritz, E.A. Auerswald, Functional phage display of leech-derived tryptase inhibitor (LDTI): construction of a library and selection of thrombin inhibitors, FEBS Lett. 458 (1999) 11–16.
[59] L.R. Ceci, M. Volpicella, Y. Rahbé, R. Gallerani, J. Beekwilder, M.A. Jongsma, Selection by phage display of a variant mustard trypsin inhibitor toxic against aphids, Plant J. 33 (2003) 557–566.
[60] W. Markland, A.C. Ley, S.W. Lee, R.C. Ladner, Iterative optimization of high-affinity proteases inhibitors using phage display. 1. Plasmin, Biochemistry 35 (1996) 8045–8057, https://doi.org/10.1021/bi9526286.
[61] W. Markland, A.C. Ley, R.C. Ladner, Iterative optimization of high-affinity protease inhibitors using phage display. 2. Plasma kallikrein and thrombin, Biochemistry 35 (1996) 8058–8067, https://doi.org/10.1021/bi952629y.
[62] G.P. Smith, Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface, Science 228 (1985) 1315–1317.
[63] A.R.M. Bradbury, S. Sidhu, S. Dübel, J. McCafferty, Beyond natural antibodies: the power of in vitro display technologies, Nat. Biotechnol. 29 (2011) 245–254, https://doi.org/10.1038/nbt.1791.

[64] P.T. Jones, P.H. Dear, J. Foote, M.S. Neuberger, G. Winter, Replacing the complementarity-determining regions in a human antibody with those from a mouse, Nature 321 (1986) 522–525, https://doi.org/10.1038/321522a0.
[65] J.D. Marks, H.R. Hoogenboom, T.P. Bonnert, J. McCafferty, A.D. Griffiths, G. Winter, By-passing immunization. Human antibodies from V-gene libraries displayed on phage, J. Mol. Biol. 222 (1991) 581–597.
[66] L. Riechmann, M. Clark, H. Waldmann, G. Winter, Reshaping human antibodies for therapy, Nature 332 (1988) 323–327, https://doi.org/10.1038/332323a0.
[67] S.S. Sidhu, H.B. Lowman, B.C. Cunningham, J.A. Wells, Phage display for selection of novel binding peptides, Methods Enzymol. 328 (2000) 333–363.
[68] G.E. Crooks, G. Hon, J.-M. Chandonia, S.E. Brenner, WebLogo: a sequence logo generator, Genome Res. 14 (2004) 1188–1190, https://doi.org/10.1101/gr.849004.
[69] G.W. Jameson, D.V. Roberts, R.W. Adams, W.S.A. Kyle, D.T. Elmore, Determination of the operational molarity of solutions of bovine alpha-chymotrypsin, trypsin, thrombin and factor Xa by spectrofluorimetric titration, Biochem. J. 131 (1973) 107–117.
[70] P. Shih, D.R. Holland, J.F. Kirsch, Thermal stability determinants of chicken egg-white lysozyme core mutants: hydrophobicity, packing volume, and conserved buried water molecules, Protein Sci. 4 (1995) 2050–2062.
[71] R.L.J. Keller, The Computer Aided Resonance Assignment Tutorial, Cantina Verlag, Goldau, 2004.
[72] T.D. Goddard, D.G. Kneller, SPARKY 3, University of California, San Francisco, 2000.
[73] W.F. Vranken, W. Boucher, T.J. Stevens, R.H. Fogh, A. Pajon, M. Llinas, E.L. Ulrich, J.L. Markley, J. Ionides, E.D. Laue, The CCPN data model for NMR spectroscopy: development of a software pipeline, Proteins 59 (2005) 687–696, https://doi.org/10.1002/prot.20449.
[74] R. Cole, J.P. Loria, FAST-Modelfree: a program for rapid automated analysis of solution NMR spin-relaxation data, J. Biomol. NMR 26 (2003) 203–213.
[75] A.E. Aliev, M. Kulke, H.S. Khaneja, V. Chudasama, T.D. Sheppard, R.M. Lanigan, Motional timescale predictions by molecular dynamics simulations: case study using proline and hydroxyproline sidechain dynamics, Proteins 82 (2014) 195–215, https://doi.org/10.1002/prot.24350.
[76] N. Guex, M.C. Peitsch, SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling, Electrophoresis 18 (1997) 2714–2723, https://doi.org/10.1002/elps.1150181505.
[77] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935, https://doi.org/10.1063/1.445869.
[78] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, Biopolymers 22 (1983) 2577–2637, https://doi.org/10.1002/bip.360221211.
[79] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera—a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612, https://doi.org/10.1002/jcc.20084.