Mapping protein selectivity landscapes using multi-target selective screening and next-generation sequencing of combinatorial libraries

Si Naftaly1, Itay Cohen1, Anat Shahar2, Alexandra Hockla3, Evette S. Radisky3 & Niv Papo1

Characterizing the binding selectivity landscape of interacting proteins is crucial both for elucidating the underlying mechanisms of their interaction and for developing selective inhibitors. However, current mapping methods are laborious and cannot provide a sufficiently comprehensive description of the landscape. Here, we introduce a novel and efficient strategy for comprehensively mapping the binding landscape of proteins using a combination of experimental multi-target selective library screening and in silico next-generation sequencing analysis. We map the binding landscape of a non-selective trypsin inhibitor, the amyloid protein precursor inhibitor (APPI), to each of the four human serine proteases (kallikrein-6, mesotrypsin, and anionic and cationic trypsins). We then use this map to dissect and improve the affinity and selectivity of APPI variants toward each of the four proteases. Our strategy can be used as a platform for the development of a new generation of target-selective probes and therapeutic agents based on selective protein–protein interactions.

1 Department of Biotechnology Engineering and the National Institute of Biotechnology in the Negev, Ben-Gurion University of the Negev, Beer-Sheva, Israel.
2 The National Institute for Biotechnology in the Negev (NIBN), Beer-Sheva, Israel.
3 Department of Cancer Biology, Mayo Clinic Comprehensive Cancer Center, Jacksonville, Florida 32224, USA. These authors contributed equally: Si Naftaly, Itay Cohen. Correspondence and requests for materials should be addressed to N.P. (email: papo@bgu.ac.il)
A defining characteristic of protein–protein interactions (PPIs) is the binding selectivity landscape of the interacting proteins\(^1\)-\(^3\), which relates the amino acid sequence to the affinity of a protein toward its target. Comprehensively mapping this landscape is crucial both for understanding the mechanisms and evolutionary origins of selective PPIs and for protein engineering purposes, e.g., for designing selective binders and/or inhibitors for target proteins\(^4\)-\(^7\). The binding selectivity landscape of each protein in a certain PPI is characterized by the interfacial residues of the protein, such that point mutating these residues can help determine the contribution of each residue to target selectivity—or in a protein with a broad selectivity spectrum to the selectivity of the protein to each of its putative targets individually. The binding selectivity landscape usually comprises of four types of key interface residues. Hot-spot residues are a few\(^8\) interface residues that are highly relevant for a specific PPI, i.e., they contribute almost 75% of the total free energy of binding (\(\Delta \Delta G_{\text{bound}}\)) of the protein to its partner\(^9\)-\(^11\). Mutating hot-spot residues therefore decreases the affinity of the protein to a specific partner—not necessarily to others. Cold-spot residues\(^1\)-\(^3\),\(^12\)-\(^14\) are interface residues occupied by suboptimal amino acids, such that mutating them increases the binding affinity of the protein to a specific partner. Selectivity-switch residues\(^1\)-\(^5\),\(^16\) are interface residues in which a point-mutation simultaneously decreases the affinity of the protein to one partner and increases its affinity to another. Finally, correlated-selectivity residues\(^1\)-\(^7\),\(^18\) are interface residues that work together to increase the selectivity of the protein to one specific partner. Such residues are especially difficult to characterize with conventional methods because only a double mutation (one mutation in each residue) can change the affinity of the protein to a certain partner.

Methods for mapping protein selectivity landscapes typically include mutating candidate residues and testing the resulting changes in affinity\(^4\),\(^19\). Despite considerable advancements in recent years\(^8\),\(^20\)-\(^22\), currently available methods still demonstrate several caveats that hinder our ability to develop, inter alia, selective inhibitors for clinically important proteins. For instance, alanine scanning and similar classical approaches\(^23\)-\(^29\) can test only a subset of all possible mutants, are time-consuming and laborious, require protein purification and binding affinity measurements for each mutant, and most importantly, focus chiefly on hot-spot residues. Modern approaches can overcome some of these caveats by employing protein library display and sorting technologies, which rapidly explore all possible (hot- and cold-spot) mutations and qualitatively map the contribution of each residue to the affinity of a protein toward its target\(^30\)-\(^37\). However, as only several hundred clones of the sorted libraries are ultimately sequenced, these methods do not comprehensively characterize the entire library and, to date, they cannot identify correlated-selectivity residues. A more recent approach employed next-generation sequencing (NGS) to guide protein and synthetic small-molecule optimization\(^4\)-\(^7\),\(^38\)-\(^42\), effectively improving the binding affinity and selectivity, and generating binding epitopes of de novo\(^6\)-\(^7\),\(^19\),\(^40\)-\(^41\),\(^43\),\(^44\). However, this and most other currently available approaches generate high-affinity (but not necessarily selective) binders, and, in the few studies designed to generate selective binders\(^7\),\(^40\),\(^41\),\(^43\),\(^44\), the methodology was limited to improving discrimination between only two target proteins that have different binding epitopes. Significantly, some broad-spectrum proteins may have many potential targets with binding epitopes that have high-sequence homology and structural similarity.

In the current study, we present a novel, single-step approach for comprehensively mapping the binding selectivity landscape of proteins (including hot-spot, cold-spot, selectivity-switch, and correlated-selectivity residues) using a combination of experimental multi-target selective library screening and in silico NGS analysis. To test our approach in a real-life context, we chose to map the binding selectivity landscape of a broad-spectrum trypsin inhibitor, namely, the human amyloid protein precursor inhibitor (APPi; a member of the human Kunitz-domain family of serine protease inhibitors\(^4\)-\(^5\)), to each of the four human serine proteases—kallikrein-6 (KLK6), mesotrypsin, anionic trypsin, and cationic trypsin—all of which share high-sequence homology and structural similarity. Then, we used this landscape to improve both the selectivity and affinity of APPi variants to each protease, which we evaluated through inhibition studies using the purified proteins.

We recently used a yeast-surface display (YSY) platform—a powerful directed evolution protein engineering technology\(^30\),\(^46\)-\(^50\)—to generate APPi-3M\(^51\); a triple-mutant APPi (M17G/I18F/F34V) whose affinity to mesotrypsin, anionic trypsin, and cationic trypsin is comparable \([K_i = 89.8 \pm 0.23 \text{ pM}, 1.47 \pm 0.02 \text{ pM}, \text{ and } 4.96 \pm 0.25 \text{ pM}]\) for mesotrypsin, anionic trypsin, and cationic trypsin, respectively\(^51\), whereas its affinity to KLK6 is lower by three orders of magnitude \([K_i = 1.06 \pm 0.12 \text{ nM}]\). These features render APPi-3M an optimal model scaffold for engineering binding selectivity; its lack of selectivity toward mesotrypsin, anionic trypsin, and cationic trypsin is a good starting point for manipulating its relative selectivity, while its lower selectivity toward KLK6 makes it a good target for engineering selectivity switches.

Our study design is demonstrated in Supplementary Figure 1. We began by generating a yeast-displayed APPi-3M library including clones with single-residue random mutations in the binding interface (i.e., in the APPi binding loop) and clones with multiple-residue random mutations both within and beyond the binding interface (i.e., in the APPi scaffold and binding loop). Then, we divided the four proteases into combinatorial pairs (six combinations) and sorted the YSD APPi-3M library for variants with differential selectivity toward each protease in each pair. We then used NGS to sequence these fractions and analyzed them computationally. Consequently, the sorted APPi-3M mutant library fractions were rich in affinity- and selectivity-enhancing mutations; of these, we identified the most highly selective APPi mutations based on their ability to inhibit—as soluble proteins—each of the four proteases. To the best of our knowledge, this is the first report of a platform that can provide such a rich PPI binding selectivity landscape.

**Results**

**Selecting APPi variants with improved selectivity.** We began by generating a library of APPi-3M clones using both site-directed random mutagenesis of the APPi-3M binding loop (residues 11–18, except invariant Cys-14) and error-prone PCR amplification of the entire coding sequence. This design yielded a ‘naïve’ APPi-3M library of 3.5 million variants, each with 0–2 amino acid mutations. Then, using our YSD system, we expressed each of these variants on the surface of yeast cells and used fluorescence-activated cell sorting (FACS) to quantify their binding to each of the four (soluble) serine proteases (mesotrypsin, KLK6, anionic trypsin, and cationic trypsin). We introduced the yeast-displayed naïve library to pairs of serine proteases, each labeled with a different fluorescent dye (Alexa Fluor-650 or Alexa Fluor-488; i.e., a pairwise selective screen, Fig. 1a), at concentrations optimized for each pair to achieve an equivalent distribution of staining intensities (Fig. 1b). The library was sorted to isolate \(~1\) million variants per sorted fraction (sorting gate), with increased selectivity toward each of the four serine protease targets versus its paired protease. Subsequent FACS analyses showed clear enrichment of the binding
selectivity to one target protease and increased the selectivity toward another.

Our analysis revealed that residue 15 in APPI-3M is a general hot spot for binding human serine proteases, as all mutations in this residue, except a substitution to Lys, decreased its binding affinity toward all four proteases (Fig. 2 and Supplementary Figure 3). The analysis also revealed two clear cold spots: most mutations in residue 13 increased binding selectivity toward mesotrypsin versus all other proteases (Fig. 2a and Supplementary Figure 3A, dashed line), while most mutations in residue 17 increased binding selectivity toward KLK6 versus all other proteases (Fig. 2b and Supplementary Figure 3B, dashed line). These two selectivity-switch residues (13 and 17) enable a selectivity shift from three proteases toward a single, different protease (either mesotrypsin or KLK6). In addition, most mutations in residue 17 increased the selectivity of APPI-3M toward anionic trypsin and cationic trypsin as compared with mesotrypsin (Fig. 2c, d and Supplementary Figure 3C and D, lower dashed line), while most mutations in residues 11 and 18 increased the selectivity toward anionic trypsin and cationic trypsin as compared with KLK6 (Fig. 2c, d and Supplementary Figure 3C and D, upper dashed line). For example, we found that residues 11 and 17 are selectivity switches for mesotrypsin and KLK6, respectively (Fig. 2, Supplementary Figure 3 and Table 1), as mutating the residue in position 11 (originally Thr) from His to Ile (Fig. 2a, d, white arrows) switched the selectivity from anionic trypsin to mesotrypsin by a factor of $69 \times 10^3$ and mutating the residue in position 17 (originally Gly) from Glu to Arg (Fig. 2b, c, black arrows) switched the selectivity from cationic trypsin to KLK6 by a factor of $7 \times 10^3$. 

Mapping correlated-selectivity residues. Next, we turned to identify the effects of double-amino acid substitutions in APPI-3M on the selectivity toward each of the four serine proteases. The first steps in this process (quality filtration and integration, translation, alignment, and enrichment ratio calculations) were similar to those described above for single-amino acid analyses. Most double-mutant APPI-3M variants that increased the selectivity toward one serine protease versus all others increased the selectivity toward KLK6 [note that the affinity of the parental
APPI-3M to KLK6 was two orders of magnitude lower than to anionic and cationic trypsin and one order of magnitude lower than to mesotrypsin\cite{a}, and these variants had mutations in residues 11 and 17 (Supplementary Table 1). To elucidate the effects of correlated residues and of residues 11 and 17 (Fig. 3c), in particular on the selectivity toward KLK6, we predicted the total effect of each pair of mutated residues (i.e., the effect of all mutations in these two residues; see Methods) and illustrated the results as heatmaps (Fig. 3a and Supplementary Figure 4). Variants in which both residues 11 and 17 were mutated demonstrated an increased selectivity (enrichment ratio >1) toward KLK6 versus the three other proteases. Therefore, we generated additional heatmaps to estimate the effect of specific pair residues (all pair combinations of residues 11 and 17, Fig. 3b). These heatmaps (Fig. 3b) revealed that many combinations of double-amino acid substitutions increased the selectivity toward KLK6 versus the three other proteases, including a combination of either Val, Ala, Pro, or Ser at residue 11 with either Ala, Arg, or Ser at residue 17 (enrichment ratio >1). For instance, the combination of Val at residue 11 and Arg at residue 17 increased the total selectivity of APPI-3M toward KLK6 by a factor of $4 \times 10^9$ (calculated as the multiplication of the three relative selectivities: $-59 \times 10^{3}$-fold versus mesotrypsin, $-364$-fold versus cationic trypsin, and $-170$-fold versus anionic trypsin; Supplementary Table 2). Similarly, the combination of Ser at residue 11 and Arg at residue 17 increased the total selectivity by a factor of $7 \times 10^7$ ($-37 \times 10^{3}$-fold versus mesotrypsin, $-24$-fold versus cationic trypsin, and $-85$-fold versus anionic trypsin).

Validating the selectivity changes using soluble inhibitors. To validate the results of the NGS computational analysis, we...
generated and purified the soluble forms of APPI-3M variants, in which the mutation was located at selectivity-switch residues on the APPI loop (Supplementary Figure 5). These variants included the mutations T11H (for which the NGS analysis predicted a selectivity switch from anionic trypsin to mesotrypsin), T11I (predicted switch from mesotrypsin to anionic trypsin), G17R (predicted switch from cationic trypsin to KLK6), and G17E (predicted switch from KLK6 to cationic trypsin) (Table 1). Then, we evaluated the affinity of these four purified APPI-3M variants to each serine protease by measuring the degree to which they inhibit the ability of each protease to hydrolyze its substrate [benzyloxycarbonyl–Gly–Pro–Arg–p–nitroanilide (Z-GPR-pNA) for mesotrypsin, anionic trypsin, cationic trypsin, and BOC–FSR–MSC for KLK6]. We determined the inhibition constant ($K_i$) of each of these interactions by quantifying the slow tight binding behavior (Supplementary Figure 6). The experimental results indeed correlated well with those of the NGS analysis (Table 2).

Positions 11 and 17 in the APPI-3M sequence are correlated. As residues in positions 11 and 17 of the APPI-3M sequence increased the selectivity of APPI-3M toward KLK6, we elucidated the interactions between different amino acids at these positions by generating and purifying representative single- and double-mutant APPI-3M variants. We chose the KLK6-selective T11V/G17R and T11S/G17R double-mutant variants (see Fig. 3), and their corresponding single-mutant selectivity-switch variants T11V, T11S, and G17R (see Tables 1 and 2). We tested the affinity of the soluble forms of these five variants to each of the four serine proteases in a competitive inhibition assay (Table 3) and, based on the obtained $K_i$ values, we determined the selectivity of each variant toward KLK6 and compared it with the selectivity of the unmodified APPI-3M (Table 4).

The amino acid substitution that most increased the total selectivity of APPI-3M toward KLK6 was T11V/G17R, followed by G17R and finally, T11S/G17R. The individual substitutions T11V and T11S did not improve the selectivity toward KLK6, rather they somewhat decreased it (Table 4). These results suggest that residues 11 and 17 are correlated-selectivity residues, which act together to increase target selectivity. To further test this hypothesis, we conducted a double-mutant cycle analysis, in which we used the selectivity values of KLK6 with the two double-mutant variants and their single variants (T11V/G17R, T11S/ G17R, T11V, T11S, and G17R, Table 4) to calculate the selectivity strength between two mutated residues (i.e., the coupling energy, $\Delta G_{\text{int}}$, Supplementary Figure 7). Indeed, in both double mutations, the $\Delta G_{\text{int}}$ values were non-zero, indicating that residues 11 and 17 interact with each other to cooperatively affect the selectivity toward KLK6.

To gain insight into the structural basis of the observed selectivity changes, we attempted to crystallize the APPI-3M-T11V/G17R variant in complex with the increased-selectivity target KLK6 and the reduced-selectivity target mesotrypsin. We were able to obtain a high-resolution crystal structure of the APPI-3M-T11V/G17R variant bound to mesotrypsin (PDB ID: 6GFI; Supplementary Table 3). A structural analysis of this complex revealed that a deleterious steric interaction between the APPI Arg-17 mutation and mesotrypsin Arg-193 pushes Arg-193 into a more buried conformation (Supplementary Figure 8), as previously found in the structures of mesotrypsin bound to wild-type APPI or BPTI Kunitz-type inhibitors. The steric clash and the restriction of Arg-193 to a single buried conformation can explain the reduction in affinity toward mesotrypsin, which is consistent with our prior structure and mutagenesis studies.

The corresponding amino acid that occupies position 193 in KLK6 is Gly (PDB ID: 4D8N); therefore, the lack of a side chain in position 193 of KLK6 is probably more energetically favored (upon binding to APPI-3M-G17R) than that of mesotrypsin Arg-193 (due to the steric clash and the restriction of Arg-193). Efforts to crystallize the APPI-3M-T11V/G17R complex with KLK6 were unsuccessful, and thus the basis for selectivity improvements toward this alternative target, and for cooperativity between APPI residues 11 and 17, remain a subject for future investigations.

**Selective screens are superior to affinity screens.** A significant advantage of our pairwise selectivity screen approach over the traditional sequential affinity screen (a commonly used method, in which the library is sorted against each enzyme separately in a sequential manner) is the ability of our approach to identify, in a single screening step (rather than two sequential affinity screen steps), the top ~5% of clones that are more selective toward one target versus another, even if the absolute affinity of these clones toward both targets is lower than that of the parent variant (in the current study, APPI-3M). To demonstrate that the traditional sequential affinity screens are unable to detect the clones obtained by our pairwise selectivity screens (namely, those with improved selectivity and low affinity), we performed two separate sequential affinity screens, one toward KLK6 and another toward cationic trypsin (Supplementary Figure 9A, B, D, E). As expected, both the sequential affinity and the pairwise selectivity screen approaches were able to identify the G17R mutation as a KLK6 selectivity-improving mutation (Supplementary Table 4), which is consistent with the 1.7-fold improvement in the selectivity toward KLK6 versus cationic trypsin, measured by the enzymatic assay (Supplementary Table 5). In contrast, we were unable to identify the selective G17E mutation by using the sequential affinity approach (Supplementary Table 4), although it was clearly identified using the pairwise selectivity screen between KLK6 and cationic trypsin (Supplementary Table 4), demonstrating a 3.4-fold improved selectivity toward cationic trypsin, as measured by the enzymatic assay (Supplementary Table 5). This discrepancy between the two approaches stems from the fact that the G17E mutation was not in the top ~5% binders in the cationic trypsin and KLK6 sorts due to its weakened affinity toward cationic trypsin and KLK6 relative to the parental molecule APPI-3M (by ~4-fold and ~10-fold, respectively, Supplementary Table 5).

**Upscaling.** Our selective pairwise screening approach can be easily scaled up for multiple target proteins per screen, such that a library

---

Table 1 Selectivity of APPI-3M variants with mutations at selectivity-switch residues toward human serine proteases

| Mutation | Target A         | Target B         | Enrichment ratio, target A | Enrichment ratio, target B | Selectivity $^a$ |
|----------|------------------|------------------|----------------------------|----------------------------|------------------|
| T11H     | Mesotrypsin      | Anionic trypsin  | 0.07                       | 7.39                       | 1                |
| T11I     | Mesotrypsin      | Anionic trypsin  | 5.91                       | 0.01                       | 69.22 × 10^3     |
| G17E     | Mesotrypsin      | Anionic trypsin  | 0.12                       | 4.88                       | 1                |
| G17R     | Mesotrypsin      | Anionic trypsin  | 12.50                      | 0.08                       | 6.65 × 10^3      |

$^a$Selectivity is defined as the fold change in the enrichment ratio for target A, divided by the fold change in the enrichment ratio for target B.
can be screened against a target of interest (labeled with one type of fluorophore) versus a mixture of competitors (all labeled with the same fluorophore, which is different from the one used for the target of interest). Such an approach is especially useful in the case where there is a single primary target of interest, since it will be completed through only a single sort. To demonstrate the feasibility of such an approach, we performed a competitive sort, in which KLK6 was the primary target of interest (labeled with Alexa Fluor-650) and cationic trypsin, anionic trypsin, and mesotrypsin (each labeled with Alexa Fluor-488) were the competitors (Supplementary Figure 9C, F), and compared the enrichment values to those of our pairwise comparisons. The enrichment ratios of the competitive multi-target screen were highly correlated with those of the pairwise selective screen; in both setups, the top-rated selectivity-improving clones were similar (Supplementary Table 6), both for single mutations (e.g., G17R) and for double mutations (e.g., T11S/G17R and T11V/G17R). In addition, this analysis revealed a clear selectivity cold spot, in which most mutations in residue 17 increased the
Table 2 Changes in the selectivity of APPI-3M variants with mutations at selectivity-switch residues toward human serine proteases

| Mutation   | Target A               | Target B               | Predicted switcha | Ki [pM], target A | Ki [pM], target B | Switch ratio |
|------------|------------------------|------------------------|-------------------|-------------------|-------------------|--------------|
| T11H       | Mesotrypsin            | Anionic trypsin        | A → B             | 302 ± 24          | 4.45 ± 0.23       | 1.6          |
| T11L       |                        |                        | B → A             | 61.2 ± 2.6        | 1.44 ± 0.14       |              |
| Kᵢ Fold changeb | —                      | —                      |                   | 4.93              | 3.09              | —            |
| G17E       | KLK6                   | Cationic trypsin       | A → B             | 464 ± 39          | 85.8 ± 5.9        | 5.7          |
| G17R       |                        |                        | B → A             | 77.4 ± 2.6        | 8.12 ± 0.08       |              |
| Kᵢ Fold changec | —                      | —                      |                   | 59.95             | 10.57             | —            |

*As predicted by the NGS analysis

bResults (means ± SD) were obtained from three independent experiments

cCalculated as the Kᵢ fold change for target A divided by the Kᵢ fold change for target B

Switch ratio = (Ki fold for target A) / (Ki fold for target B)

Table 3 Kᵢ constants of human serine proteases inhibited by various APPI-3M variants

| Mutant   | 2Ki [pM] | Mesotrypsin | Anionic trypsin | Cationic trypsin | KLK6 |
|----------|----------|-------------|----------------|----------------|------|
| Unmodified |         | 98.0 ± 1.0 | 2.26 ± 0.08    | 22.5 ± 0.6     | 362 ± 10 |
| APPI-3M   |         | 494 ± 28   | 0.92 ± 0.07    | 2.37 ± 0.17    | 16.4 ± 0.9 |
| T11V/G17R |         | 1060 ± 30  | 2.98 ± 0.19    | 7.25 ± 0.5     | 124 ± 13  |
| T11S/G17R |         | 581 ± 7    | 1.16 ± 0.09    | 7.63 ± 0.55    | 1000 ± 60 |
| T11S      |         | 65.0 ± 1.0 | 3.72 ± 0.21    | 14.1 ± 0.5     | 378 ± 9   |
| G17R      |         | 676 ± 8    | 3.58 ± 0.16    | 8.12 ± 0.08    | 77.4 ± 2.6 |

*Results (means ± SD) were obtained from three independent experiments

binding selectivity toward KLK6 versus all other proteases (Supplementary Figure 3E). This finding is consistent with those obtained using the pairwise screening approach (Supplementary Figure 3B).

Discussion

We describe a novel strategy for mapping the binding selectivity landscapes of proteins through a combination of experimental multi-target selective library screening and in silico next-generation sequencing analysis. Employing the APPI/serine protease system as a model PPI, we show that our strategy can be used to map, in a rapid, single-step, cost-effective process, several crucial aspects of the selectivity landscape, including hot-spot residues, selectivity switch residues, and correlated-selectivity residues. The latter are of special importance, as characterizing correlated-selectivity mutations and analyzing their effects (both individually and combined) on target affinity and selectivity is challenging with currently available approaches.27,56.

Several previous studies have combined selective screening of a protein library and NGS analyses to map the binding landscape of various proteins, including influenza inhibitors (HB36.4, HB80.3)7, the human Yes Associated Protein 65 (hYAP65) WW domain19, and an anti-VEGF antibody.40 However, these approaches employed either libraries of clones with only single mutations or library screens that were performed against only a single target. Therefore, in these previous studies, it was difficult to identify mutations that change target selectivity or that work in concert to affect target affinity and selectivity in a correlated manner. Thus, a major advantage of our approach is its ability to identify correlated-selectivity mutations. For example, we found that the mutations T11V and G17R, when combined, yield a highly potent and selective inhibitor for KLK6, while combining the mutations T11S and G17R yield only a moderately potent and partially selective inhibitor for KLK6. These findings may suggest that a small and hydrophobic amino acid (e.g., Val in position 11) exerts a stronger effect on selectivity towards KLK6 than a small and polar amino acid (Ser in position 11).

Another advantage of our approach lies in using a pairwise selectivity screen, rather than the sequential affinity screen that is commonly used in other approaches40,43, to increase selectivity. This advantage is especially noticeable for the identification of clones that are selective but have distinct affinities toward both targets that are lower than that of the parent variant (in the current study, APPI-3M), as demonstrate in Supplementary Table 4. In addition, the pairwise screening approach can be easily scaled up for multiple target proteins per screen, such that a library can be screened against a target of interest versus a mixture of competitors. Such an approach is especially useful where there is a single primary target of interest, since it will be completed with only one sort, as demonstrate in Supplementary Table 6.

We chose the serine protease family as an ideal group of targets to demonstrate our strategy mainly because inhibiting the human serine proteases is of clinical value: both KLK6 and mesotrypsin are involved in cancer progression57–59, while anionic and cationic trypsins are involved in the etiology of pancreatitis60,61. However, the development of inhibitors capable of discriminating among trypsin-like proteases has been challenging. We and others have previously used X-ray crystallography to explore the structures of these proteases, in some cases identifying the distinguishing features that suggest the potential for developing highly selective inhibitors54,62,63. For example, several adaptive mutations have been used to show the active site of mesotrypsin for distinct substrate and inhibitor-binding selectivity54,63–65. Nevertheless, the development of truly selective inhibitors has yet to be achieved, and we anticipate that our novel approach, which is capable of rapidly and efficiently screening large libraries to comprehensively map selectivity, will enable the development of selective probes and therapeutic agents.

APPI has attracted our interest as a scaffold for engineering selective serine protease inhibitors due to the marked sequence diversity among Kunitz family members, which possess canonical binding loops that are highly tolerant to substitution or incorporation of additional amino acids66,67. Because the sequence of the canonical binding loop and neighboring residues largely determine the affinity and selectivity of the inhibitor to its targets53,68, using APPI as a scaffold offers a unique opportunity to optimize target affinity and selectivity without compromising
changes in the interactions between proteins. Currently available approaches can typically identify only greater described mutations that affect the binding affinity of a protein. For example, we identified mutations in APPI to serine proteases, as well as some novel mutations. For instance, residue R15K reduced the binding affinity of APPI to mesotrypsin, as all amino acid mutations in this residue (except A15C) reduced the binding affinity of APPI-3M to each of the four proteases. On the other hand, mutating the residue in position 11 switched the selectivity from anionic trypsin to mesotrypsin. Therefore, the difference between residues 13 and 11 is that the former facilitates a selectivity switch from three proteases to a specific protease, while the latter enables a selectivity switch from one protease to one other protease.

### Table 4 The selectivity of APPI-3M variants (normalized to the unmodified APPI-3M) toward KLK6 versus the three other proteases

| Mutant           | vs. mesotrypsin | vs. anionic trypsin | vs. cationic trypsin | Calculated KLK6 total selectivity<sup>a</sup> | Expected KLK6 total selectivity<sup>b</sup> |
|------------------|----------------|--------------------|---------------------|-----------------------------------------------|-------------------------------------------|
| Unmodified APPI-3M | 1              | 1                  | 1                   | 1                                             | 1                                         |
| T11V/G17R        | 111.10         | 8.94               | 2.32                | 2304.13                                       | 242.45                                    |
| T11S/G17R        | 31.62          | 3.87               | 0.94                | 115.53                                        | 32.33                                     |
| T11S             | 2.14           | 0.30               | 0.12                | 0.08                                          | —                                         |
| T11V             | 0.64           | 1.58               | 0.60                | 0.60                                          | —                                         |
| G17R             | 32.26          | 7.42               | 1.69                | 404.09                                        | —                                         |

<sup>a</sup>Calculated selectivity = Calculated selectivity of A × calculated selectivity of B

<sup>b</sup>Expected selectivity of double-mutant AB = calculated selectivity of A × calculated selectivity of B

In addition, the affinity of the complexes between APPI and mesotrypsin, anionic trypsin, and cationic trypsin is similar, which facilitated the identification of cold spots, whereas the affinity of the APPI/KLK6 complex is three orders of magnitude lower than that of the other complexes, thus allowing us to identify selectivity-switch residues.

As a validation of the utility of our platform, we show that the results obtained using NGS of the selected APPI clones typically correlate well with the binding selectivity of the purified protein variants in solution (as measured by competitive inhibition studies), but at different scales (Supplementary Table 7). For example, the selectivity values of 13 combinations of enzyme–inhibitor variants (out of a total of 15 possible combinations examined), calculated using NGS, are well-correlated (whether the selectivity was improved or damaged) with those obtained in the enzymatic assay. Of note, in all 15 combinations, a clear correlation was found between the ranking of the selectivity values that were calculated by each method (ranking is according to the level of selectivity improvement within each method for each enzyme, with the greatest improvement ranked as one; see example in bold boxes in Supplementary Table 7).

As shown in Table 1, the NGS analysis predicted a selectivity increase of ~7 × 10<sup>2</sup>-fold from cationic trypsin to KLK6 for G17R compared with G17E, and of ~70 × 10<sup>3</sup>-fold from anionic trypsin to mesotrypsin for T11S compared with T11H; both these findings are in qualitative agreement with the increase in selectivity determined from the Ki values of the soluble proteins, namely, an increase of ~5.7-fold and ~1.6-fold, respectively (Table 2). However, no correlation was found between the magnitudes of the improvements, i.e., the 7 × 10<sup>3</sup>-fold improvement calculated by NGS was calculated as a ~5.7-fold improvement in the enzymatic assay, while the 70 × 10<sup>3</sup>-fold improvement calculated by NGS was calculated as only a ~1.6-fold improvement in the enzymatic assay. Therefore, the selectivity increase values that were calculated by the NGS cannot be directly compared with those of the competitive inhibition studies; rather, the values can be compared between experiments using each method, and not between the two methods. Nevertheless, the results shown in Tables 1 and 2 confirm that our approach can predict the positions that can change target selectivity, and that our approach is sufficiently sensitive to detect small affinity changes, whereas other currently available approaches can typically identify only greater changes in the interactions between proteins.

In further validation of our strategy, we identified most previously described mutations that affect the binding affinity and selectivity of APPI to serine proteases, as well as some novel mutations. For example, we identified residue 15 as a hot spot for all four human serine proteases, as all amino acid mutations in this residue (except R15K) reduced the binding affinity of APPI-3M to each of the four proteases (Fig. 2 and Supplementary Figure 3). Indeed, residue 15 had previously been identified as a hot spot in Kunitz-domain inhibitors in studies with BPTI<sup>70,71</sup>. In addition, our data identified, for the first time, to the best of our knowledge, that residue 13 is a selectivity cold-spot for mesotrypsin, as most of the mutations in this residue improved selectivity toward mesotrypsin versus all other proteases. On the other hand, mutating the residue in position 11 switched the selectivity from anionic trypsin to mesotrypsin. Therefore, the difference between residues 13 and 11 is that the former facilitates a selectivity switch from three proteases to a specific protease, while the latter enables a selectivity switch from one protease to one other protease.

The use of NGS covered the entire library and provided a comprehensive map of the binding interface. However, generating the library by using a combination of site-specific saturation mutagenesis on the APPI loop, and random mutations also on other parts of the gene, limited our ability to analyze residues that are distant from the interaction site. We attribute this limitation to technical aspects of our library design, as the random mutations generated by using the error-prone PCR were represented to a lower extent than mutations generated by using site-saturation libraries. Nevertheless, the residues that we found to improve the selectivity of APPI toward the four serine proteases can provide an explanation for the basis for target selectivity of inhibitors toward serine proteases. These selectivity-improving mutations can also be beneficial for designing targeted therapeutics for cancer and other diseases, as they can potentially inhibit the desired serine protease in a selective manner, so as to minimize toxic effects. This study also serves as an example for the general utility of our new platform, as many PPI mediators and disease targets belong to large families of related proteins, making target selectivity a highly desirable but challenging goal in drug development. Thus, we our approach for simply and efficiently mapping PPI selectivity landscapes offers great promise for designing novel target-selective therapeutics.

### Methods

#### YSD and flow cytometry cell sorting

The yeast-displayed APPI-3M library was constructed as described in Supplementary Methods. To display the APPI-3M library on the surface of the yeast, the library was grown in an SDCAA selective medium (2% dextrose, 0.67% Difco yeast nitrogen base, 0.5% Bacto casamino acids, 0.52% Na<sub>2</sub>HPO<sub>4</sub>, and 0.856% NaH<sub>2</sub>PO<sub>4</sub>-H<sub>2</sub>O) and induced for expression with a galactose medium (as for SDCAA, but with galactose 2%, instead of dextrose) according to an established protocol<sup>72</sup>. Inactive forms of mesotrypsin, anionic trypsin, and cationic trypsin containing the mutation S195A were used as a precaution against enzymatic cleavage during the experiments<sup>31</sup>. The four serine proteases were labeled with Alexa Fluor dyes (Invitrogen, Carlsbad, CA) and used to detect binding. For pairwise selectivity screen, ~ 1 × 10<sup>6</sup> yeast cells were incubated with different Alexa Fluor-labeled serine proteases in a binding buffer (100 mM Tris, pH = 8.0, 1 mM CaCl<sub>2</sub>, 1% BSA) for 1.5 h at room temperature. Then, the cells were washed with the binding buffer and sorted for the high-selective variant by conducting several independent sorts, using FACSAria [the Ilse Katz Institute for Nanoscale Science and Technology, Ben-Gurion University of the Negev (BGU), Israel]. The complexes included the following pairs and
concentrations: mesotrypsin/KLK6 [25 nM/7 nM], mesotrypsin/cationic trypsin [25 nM/8 nM], mesotrypsin/anionic trypsin [25 nM/12.5 nM], anionic trypsin/KLK6 [250 nM/7 nM], cationic trypsin/KLK6 [100 nM/7 nM], and anionic trypsin/cationic trypsin [12.5 nM/100 nM]. APPI-3M variants that showed a high binding affinity (top 5% of the entire population) toward one serine protease in the pair and a low binding affinity toward the other were selected. Dual-color flow cytometry (BD Accuri C6, Piscataway, NJ) was used to test the selective binding of each sorted library to one serine protease in the pair in the presence of the other.

Quality filtration and integration of sequences. Sequencing data from each library were treated identically and evaluated in triplicates, and Spearman’s rank correlation coefficient was calculated to be above 95%. An average Illumina quality score was calculated for each read in a given set of paired-end reads, and read pairs in which either read had an average Illumina score lower than 20 (i.e., less than 99% accuracy) were discarded. The remaining read pairs were merged into a single sequence by fast structure of the mesotrypsin/BPTI complex is available in the PDB under the accession code 1nzc (2012). The structural and energetic basis for high selectivity in a high-affinity protein-protein interaction. Proc. Natl Acad. Sci. USA 107, 10080–10085 (2010).

Karanicolas, J. et al. A de novo protein binding pair by computational design and directed evolution. Mol. Cell 42, 250–260 (2011).

Shirinian, J., Shariabi, O. & Shifman, J. M. Cold spots in protein binding. Trends Biochem. Sci. 41, 739–745 (2016).

Sio, C. F., Otten, L. G., Cool, R. H. & Quax, W. J. Analysis of a substrate specificitv switch residue of cephalosporin acylase. Biochem. Biophys. Res. Commun. 312, 755–760 (2003).

Gobel, U., Sande, C., Schödel, R. & Valencia, A. Correlated mutations and residue contacts in proteins. Proteins 18, 309–317 (1994).

Pazos, F., Helmer-Citterich, M., Audzio, G. & Valencia, A. Correlated mutations contain information about protein-protein interaction. J. Mol. Biol. 271, 511–523 (1997).

Fowler, D. M. et al. High-resolution mapping of protein sequence-function relationships. Nat. Methods 7, 741–746 (2010).

Araya, C. L. & Fowler, D. M. Deep mutational scanning: assessing protein function on a massive scale. Trends Biotechnol. 29, 435–442 (2011).

Hietpas, R. T., Jensen, J. D. & Bolon, D. N. Experimental illumination of a fitness landscape. Proc. Natl Acad. Sci. USA 108, 7896–7901 (2011).

Siloto, R. M. & Randall, J. Weseleale. Site saturation mutagenesis: methods and applications in protein engineering. Biocatal. Agric. Biotechnol. 1, 181–189 (2012).

Ashkenazi, A. et al. Mapping the CD4 binding site for human immunodeficiency virus by alanine-scanning mutagenesis. Proc. Natl Acad. Sci. USA 87, 7150–7154 (1990).

Cunningham, B. C. & Wells, J. A. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. Science 244, 1081–1085 (1989).

Hietpas, R. T., Bank, C., Jensen, J. D. & Bolon, D. N. A. Shifting fitness landscapes in response to altered environments. Evol.; Int. J. Evol. 67, 3512–3522 (2013).

Kortemme, T., Kim, D. E. & Baker, D. Computational alanine scanning of protein-protein interfaces. Sci. STKE 2004, pl2 (2004).

Kristensen, C. et al. Alanine scanning mutagenesis of insulin. J. Biol. Chem. 272, 12978–12983 (1997).

Weiss, G. A., Watanabe, C. K., Zhong, A., Goddard, A. & Sidhu, S. S. Rapid mapping of protein functional epitopes by combinatorial alanine scanning. Proc. Natl Acad. Sci. USA 97, 8950–8954 (2000).

Xu, P. et al. Design of specific serine protease inhibitors based on a versatile peptide scaffold: conversion of a urokinase inhibitor to a plasma kallikrein inhibitor. J. Med. Chem. 58, 8868–8876 (2015).

Boder, E. T. & Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. Nat. Biotechnol. 15, 533–537 (1997).

Cortese, R. et al. Epitope discovery using peptide libraries displayed on phage. Trends Biotechnol. 12, 262–267 (1994).

Fack, F. et al. Epitope mapping by phage display: random versus gene-fragment libraries. J. Immunol. Methods 206, 43–52 (1997).

Gai, S. A. & Wittrup, K. D. Yeast surface display for protein engineering and characterization.Curr. Opin. Struct. Biol. 17, 467–473 (2007).

Pal, G., Kouadio, J. L., Artsis, D. R., Kossiakoff, A. A. & Sidhu, S. S. Comprehensive and quantitative mapping of energy landscapes for protein-protein interactions by rapid combinatorial scanning. J. Biol. Chem. 281, 22378–22385 (2006).

Rabinovich, E. et al. Identifying residues that determine SCF protein-anchoring interactions through a combination of experimental and in silico analyses. J. Mol. Biol. 429, 97–114 (2014).

Rosenfeld, L., Heyne, M., Shifman, J. M. & Papo, N. Protein engineering by combined computational and in vitro evolution approaches. Trends Biochem. Sci. 41, 421–433 (2016).

Rosenfeld, L. et al. Combinatorial and computational approaches to identify interactions of macrophage colony-stimulating factor (M-CSF) and its receptor c-FMS. J. Biol. Chem. 290, 26180–26193 (2015).
38. Cakar, Z. P., Turanli-Yildiz, B., Alikin, C. & Yilmaz, U. Evolutionary engineering of Saccharomyces cerevisiae for improved industrially important proteins. Fem. Res. Rev. 12, 171–182 (2012).

39. Fowler, D. M., Araya, C. L., Gerard, W. & Fields, S. Enrich: software for analysis of protein function by enrichment and depletion of variants. Bioinformatics 27, 3430–3431 (2011).

40. Koenig, P. et al. Deep sequencing-guided design of a high affinity dual specificity antibody to target two angiogenic factors in neovascular age-related macular degeneration. J. Biol. Chem. 290, 21773–21786 (2015).

41. Cohen-Khait, R. & Schreiber, G. Low-stringency selection of TEM1 for BLIP shows interface plasticity and selection for faster binders. Proc. Natl Acad. Sci. USA 113, 14982–14987 (2016).

42. Mendes, K. R. et al. High-throughput identification of DNA-encoded IgG ligands that distinguish active and latent Mycobacterium tuberculosis infections. Acsc. Chem. Biol. 12, 234–243 (2017).

43. Jardine, G. et al. HIV-1 broadly neutralizing antibody precursor B cells revealed by germline-targeting immunogen. Science 351, 1458–1463 (2016).

44. Wang, X. et al. Fine epitope mapping of two antibodies neutralizing the bovine pancreatic trypsin inhibitor domain is a highly specific substrate of mesotrypsin. J. Biol. Chem. 285, 1939–1949 (2010).

45. Boder, E. T., Midelfort, K. S. & Wittrup, K. D. Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. Proc. Natl Acad. Sci. USA 97, 10701–10705 (2000).

46. Graff, C. P., Chester, K., Regent, R. & Wittrup, K. D. Directed evolution of an anti-carcinoembryonic antigen scFv with a 4-day monovalent dissociation half-time at 37 degrees C. Protein Eng. Des. Sel. 17, 293–304 (2004).

47. Kiek, M. C. et al. Selection of functional T cell receptor mutants from a yeast surface-display library. Proc. Natl Acad. Sci. U. S. A. 96, 5651–5656 (1999).

48. Krowarsch, D., Cierpicki, T., Jelen, F. & Otlewski, J. Canonical protein substrate speciﬁcity of mesotrypsin inhibitors of serine proteases. Cell. Mol. Life Sci. -CMLS 60, 2427–2444 (2003).

49.ifold, A., Gendreizig, S., Tafelmeyer, P. & Johnsson, K. Changing the substrate speciﬁcity of cytochrome c peroxidase using directed evolution. Biochem. Biophys. Res. Commun. 286, 126–132 (2001).

50. Chao, G., Cochran, J. R. & Wittrup, K. D. Analysis of serine proteinase-inhibitor interaction by alanine shaving. Protein Sci. 11, 806–819 (2002).

51. Kim, Y. S., Bhandari, R., Cochran, J. R., Kuriyan, J. & Wittrup, K. D. Directed evolution of an antibody to target two angiogenic factors in neovascular age-related macular degeneration. J. Biol. Chem. 270, 25411–25417 (1995).

52. Cohen, I. et al. Combinatorial protein engineering of proteolytically resistant human kallikrein 6 reveals that a trypsin-like kallikrein is expressed in the central nervous system. J. Biol. Chem. 287, 24562–24570 (2002).

53. Kukor, Z., Toth, M. & Sahin-Toth, M. Human anionic trypsinogen: properties of autocatalytic activation and degradation and implications in pancreatic diseases. Euro. J. Bioch. 12, 171–182 (2003).

54. Berry, M. J. et al. Crystal structure and biochemical characterization of human kallikrein 6 reveals that a trypsin-like kallikrein is expressed in the central nervous system. J. Biol. Chem. 277, 24562–24570 (2002).

55. Katona, G., Berglund, G. I., Hajdu, J., Graf, L. & Szilagyi, L. Crystal structure reveals basis for the inhibitor resistance of human brain trypsin. J. Mol. Biol. 315, 1209–1218 (2002).

56. Alley, A. P. et al. Mesotrypsin has evolved four unique residues to cleave trypsin inhibitors as substrates. J. Biol. Chem. 290, 21523–21535 (2015).

57. Salameh, M. A., Soares, A. S., Alloy, A. & Radisky, E. S. Presence versus absence of hydrogen bond donor Tyr-39 influences interactions of cationic trypsin and mesotrypsin with protein protease inhibitors. Protein Sci. 21, 1103–1112 (2012).

60. Lopez-Onin, C. & Matrisian, L. M. Emerging roles of proteases in tumour suppression. Nat. Rev. Cancer 7, 800–808 (2007).

61. Kukor, Z., Toth, M. & Sahin-Toth, M. Human anionic trypsinogen: properties of autocatalytic activation and degradation and implications in pancreatic diseases. Euro. J. Bioch. 12, 171–182 (2003).

62. Bergmann, M. et al. Crystal structure and biochemical characterization of human kallikrein 6 reveals that a trypsin-like kallikrein is expressed in the central nervous system. J. Biol. Chem. 277, 24562–24570 (2002).

63. Katona, G., Berglund, G. I., Hajdu, J., Graf, L. & Szilagyi, L. Crystal structure reveals basis for the inhibitor resistance of human brain trypsin. J. Mol. Biol. 315, 1209–1218 (2002).

64. Alley, A. P. et al. Mesotrypsin has evolved four unique residues to cleave trypsin inhibitors as substrates. J. Biol. Chem. 290, 21523–21535 (2015).

65. Salameh, M. A., Soares, A. S., Alloy, A. & Radisky, E. S. Presence versus absence of hydrogen bond donor Tyr-39 influences interactions of cationic trypsin and mesotrypsin with protein protease inhibitors. Protein Sci. 21, 1103–1112 (2012).