Climbing up and down binding landscapes: a high-throughput study of mutational effects in homologous protein-protein complexes

Michael Heyne[1,2], Jason Shirian[1], Itay Cohen[2], Yoav Peleg[3], Evette S. Radisky[4], Niv Papo[2*], and Julia M. Shifman[1*]

[1]Department of Biological Chemistry, The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel

[2]Avram and Stella Goldstein-Goren Department of Biotechnology Engineering and the National Institute of Biotechnology in the Negev, Ben-Gurion University of the Negev, Beer-Sheva, Israel

[3]Life Sciences Core Facilities (LSCF) Structural Proteomics Unit (SPU), Weizmann Institute of Science, Rehovot, Israel

[4]Department of Cancer Biology, Mayo Clinic Comprehensive Cancer Center, Jacksonville, Florida, USA

* To whom correspondence should be addressed. E-mail: jshifman@mail.huji.ac.il; papo@bgu.ac.il

1

Abstract

Each protein-protein interaction (PPI) has evolved to possess binding affinity that is compatible with its cellular function. As such, cognate enzyme/inhibitor interactions frequently exhibit very high binding affinities, while structurally similar non-cognate PPIs possess substantially weaker binding affinities. To understand how slight differences in sequence and structure could lead to drastic changes in PPI binding free energy ($\Delta\Delta G_{bind}$), we study three homologous PPIs that span nine orders of magnitude in binding affinity and involve a serine protease interacting with an inhibitor BPTI. Using state-of-the-art methodology that combines protein randomization and affinity sorting coupled to next-generation sequencing and data normalization, we report quantitative binding landscapes consisting of $\Delta\Delta G_{bind}$ values for the three PPIs, gleaned from tens of thousands of single and double mutations in the BPTI binding interface. We demonstrate that the three homologous PPIs possess drastically different binding landscapes and lie at different points in respect to the landscape maximum. Furthermore, the three PPIs demonstrate distinct patterns of coupling energies between two simultaneous mutations that depend not only on positions involved but also on the nature of the mutation. Interestingly, we find that in all three PPIs positive epistasis is frequently observed at hot-spot positions where mutations lead to loss of high affinity, while conversely negative epistasis is observed at cold-spot positions, where mutations lead to affinity enhancement. The new insights on PPI evolution revealed in this study will be invaluable in understanding evolution of other biological complexes and can greatly facilitate design of novel high-affinity protein inhibitors.

Significance

Protein-protein interactions (PPIs) have evolved to display binding affinities that can support their function. As such, cognate and non-cognate PPIs could be highly similar structurally but exhibit huge differences in binding affinities. To understand this phenomenon, we studied the effect of tens of thousands of single and double mutations on binding affinity of three homologous protease-inhibitor complexes. We show that binding landscapes of the three complexes are strikingly different and depend on the PPI evolutionary optimality. We observe different patterns of couplings between mutations for the three PPIs with negative and positive epistasis appearing most frequently at hot-spot and cold-spot positions, respectively. The evolutionary trends observed here are likely to be universal to all biological complexes in the cell.

Introduction Protein function is determined by the protein amino acid sequence, which has undergone millions of years of evolution while subjected to various selection pressures. Native proteins have evolved not only to perform their main function but also to satisfy a number of criteria such as solubility (1), low propensity for aggregation, stability, resistance to stress conditions (2), etc. As a result, proteins usually function below their maximum capacity (3). Multiple experiments on enzymes and binding domains proved that protein fitness could be enhanced by several orders of magnitude by applying an appropriate pressure and selecting the fittest protein sequences (4-7).

Fitness landscapes explore the effects of all possible mutations on the ability of proteins to perform their main function. Such landscapes reveal how far a particular protein is from its functional maximum, what fraction of mutations leads up and down the "fitness hill", how large the mutational steps are and which residues are the most critical to protein function (8). Mapping of fitness landscapes is thus an attractive strategy for approaching various protein engineering projects with the goal to improve or modify protein function since the best mutations could be easily identified from the fitness landscape (9, 10). Development of new strategies for protein randomization and advances in Next Generation Sequencing (NGS) enabled several exciting studies that report fitness landscapes for a number of biological systems (1, 11-22). In these studies, the effects of mutations on enzyme catalysis, fluorescence, thermostability, and other functions have been reported, giving invaluable insights on how different biological functions have evolved.

One out of many important protein functions is binding between two or more protein partners. Binding is crucial in many cellular activities such as signal transduction, protein regulation, transcription/ translation and others. Mutations in protein-protein interactions (PPIs) frequently result in a change in free energy of binding ($\Delta\Delta G_{bind}$), sometimes weakening and sometimes stabilizing the interaction (23). A mutation resulting in substantial $\Delta\Delta G_{bind}$ in one PPI could translate into remodeling of the whole PPI network, frequently leading to dysregulation of

4

signal transduction pathways and disease (24, 25). Therefore, understanding how mutations in PPIs affect their binding affinity is of great importance to both basic biology and to biomedical sciences, where inhibition or activation of a particular PPI might help to treat the related disease.
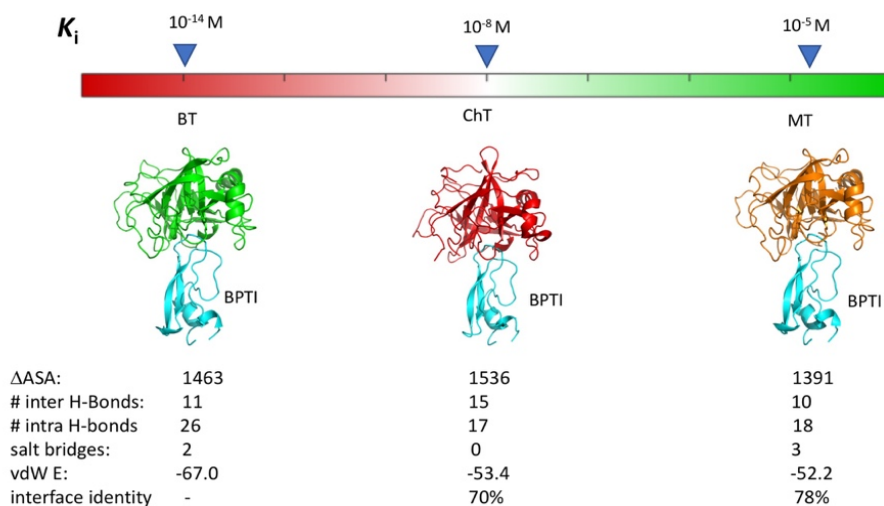
At the present moment, comprehensive binding landscapes have been mapped for only a handful of proteins(9, 26-30), while for most PPIs only a few $\Delta\Delta G_{bind}$ data points have been measured, most frequently corresponding to mutations to alanines (31-35). Comparison of the available sparse $\Delta\Delta G_{bind}$ data from different studies let us hypothesize that different classes of PPIs possess principally different binding landscapes and lie at different points relative to the binding landscape maximum, i. e. the amino acid sequence with the highest possible affinity. While in some PPIs, the majority of single mutations lead to large destabilization of the protein-protein complex (28, 36-38), in other PPIs frequent affinity-enhancing mutations are observed (27, 39). The magnitude of $\Delta\Delta G_{bind}$ due to mutation is likely to depend on the nature of the PPI under study as well as on the location of the mutation within the protein. It has been demonstrated that a few critical positions termed hot-spots of binding contribute the most significantly to the PPI binding energy with mutations at those positions usually leading to large drop in affinity (40-42). Cold-spot positions, on the other hand, present multiple possibilities for PPI affinity improvement (43).

To investigate the relationship between PPI structure/function and the corresponding binding free energy, we set to compare comprehensive binding landscapes of three homologous PPIs that span nine orders in magnitude in binding affinity (K_D). These complexes include Bovine Pancreatic Trypsin Inhibitor (BPTI) interacting with three serine proteases: Bovine trypsin (BT) ($K_D = 10^{-14}$ M) (38), Bovine α-Chymotrypsin (ChT) ($K_D = 10^{-8}$ M) (38) and Human Mesotrypsin (MT) ($K_D = 10^{-5}$ M) (44). In spite of such large differences in K_{DS}, the binding interfaces of the three complexes are highly similar, exhibiting nearly identical physicochemical properties (Figure 1).

In attempt to explain drastic differences in binding affinities of the three homologous PPIs, we explored $\Delta\Delta G_{bind}$ values between the three proteases and all single and double binding interface mutants of BPTI. To measure $\Delta\Delta G_{bind}$ values for tens of thousands of mutants in these three PPIs, we employed a recently developed by our group strategy that relies on yeast surface display (YSD) technology, NGS analysis and subsequent normalization of NGS data using a small dataset of $\Delta\Delta G_{bind}$ values measured of purified proteins (45). Applying such a protocol to the BPTI/BT

interaction, yielded a very high (R > 0.9) correlation between NGS-based and actual experimental values of $\Delta\Delta G_{bind}$ and allowed us to predict the effect of all single mutations in the BPTI/BT interaction on the PPI binding affinity (45). In the present study, we extended the above approach to quantify $\Delta\Delta G_{bind}$ values for single and double mutants of BPTI interacting with CT and MT and for double mutants of BPTI interacting with BT.

Our data demonstrate that the three homologous BPTI/protease complexes possess drastically different binding landscapes and lie at different points in respect to the binding landscape maximum. Additionally, these differences in landscape contour and placement underlie correspondingly different energetic consequences of mutation, including asymmetrical directionality and different tendencies toward positive or negative epistasis.


	BT	ChT	MT
ΔASA:	1463	1536	1391
# inter H-Bonds:	11	15	10
# intra H-bonds	26	17	18
salt bridges:	2	0	3
vdW E:	-67.0	-53.4	-52.2
interface identity	-	70%	78%

Figure 1: Comparison of K_Ds and structures for the 3 PPIs. Structures of the 3 complexes between BPTI and BT (PDB ID 3OTJ), ChT (PDB ID 1CBW) and MT (PDB ID 2R9P) are shown. Summarize are inhibition constants, change in solvent-accessible surface area upon complex formation (ΔASA), number of intra and inter hydrogen bonds, salt bridges, Van-der-Waals (vdW) energy of different protease/BPTI complexes and interface sequence identity of proteases relative to BT.

Results

To map binding landscapes of the three homologous BPTI/protease complexes, we first incorporated the wild-type BPTI (BPTI$_{WT}$) gene into the pCTCON vector, compatible with YSD technology. In such a construct, BPTI$_{WT}$ is expressed on the surface of a yeast cell with a C-terminal myc-tag for monitoring protein expression (Figure 2A). Binding of a protease to BPTI$_{WT}$

6

was accessed by monitoring fluorescence of the FITC fluorophore conjugated to the protease via neutravidin. The assessment of binding of BPTI$_{WT}$ to the three proteases by Fluorescently Activated Cell Sorting (FACS) showed a diagonal narrow distribution, demonstrating that BPTI$_{WT}$ is well expressed on the surface of yeast cells, is properly folded, and binds to each of the proteases (Figure S1).

We next generated a library of BPTI mutants that contained all single and double BPTI mutants at positions that comprise the direct binding interface with proteases in the BPTI/protease structures. We randomized twelve BPTI positions to twenty amino acids while leaving two cysteines that participate in a disulfide bond intact to preserve BPTI folding (Figure 2B). In addition, all possible combinations of double mutations encompassing these twelve positions were encoded in the library. The BPTI library, referred to as the naïve library, contained 228 single mutants and all possible pairs of such mutations, resulting in the total theoretical diversity of 26,400 BPTI sequences. The naïve library was transformed into yeast and sequenced by NGS. Sequencing results showed that all possible single mutations and 89% of all double mutations were covered in the naïve library (60% when a cutoff of 5 sequencing reads was applied).
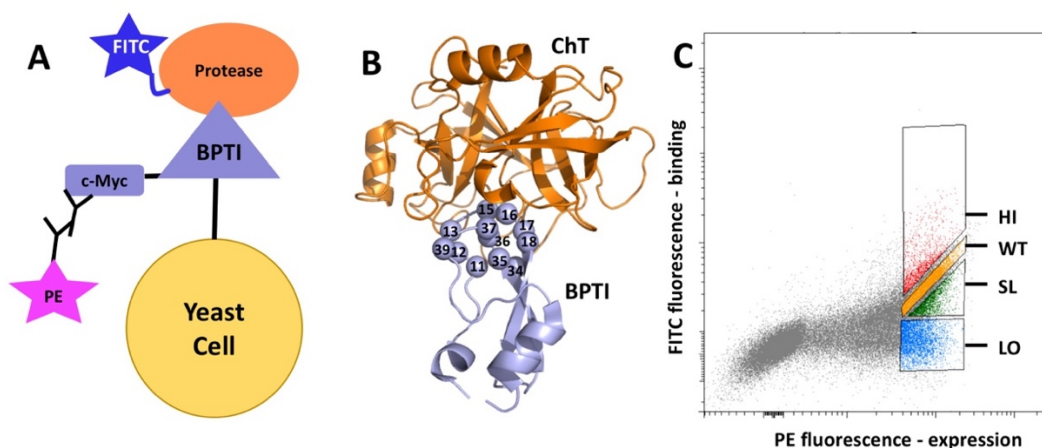
Figure 2: A) Yeast surface display construct with BPTI displayed on the surface of yeast cells. C-Myc tag is used to monitor BPTI mutant expression using PE-labeled antibody. Proteases are labeled by FITC that is used to monitor binding between the two proteins B) Construction of the BPTI mutant library. Structure of the ChT/BPTI complex with ChT shown in orange and BPTI in violet. BPTI binding interface positions randomized to twenty amino acids are shown as spheres (C) FACS data showing sorting of BPTI mutants binding to protease into four different populations. The uppermost HI gate contains BPTI mutants with affinity higher than that of BPTI$_{WT}$. The second uppermost gate, WT, contains BPTI mutants with an affinity similar to BPTI$_{WT}$. The third gate, SL, contains BPTI mutants with an affinity slightly lower than that of BPTI$_{WT}$ and the lowest gate, LO, contains BPTI mutants with an affinity much lower than that of BPTI$_{WT}$. The data is shown for the BPTI/ChT interaction, while similar data was obtained for the BPTI/MT and BPTI/BT interactions.

7

We next expressed the BPTI library on the yeast surface and measured expression and binding of the BPTI library to the three proteases using FACS (Figure 2C). Concentration of each protease was optimized to exhibit a considerable spread of the FACS binding signals from different BPTI mutants (Supplementary Figure S2). For each protease, we performed a sorting experiment and collected yeast cells with BPTI mutants belonging to four different affinity groups: higher than WT affinity (HI), WT-like affinity (WT), slightly lower than WT affinity (SL), and strongly lower than WT affinity (LO) (Figure 2C and Supplementary Figure S3-S5). The cells from each affinity gate were grown and sequenced with NGS, resulting in 300-900K reads per each population. For each BPTI mutant and each protease, we next calculated the enrichment value, which represents the ratio between the mutant's frequency in a particular affinity gate to its frequency in the naïve library. We thereby obtained heatmaps of the enrichment values for all positions as shown on Figure 3 for the CT/BPTI interaction.
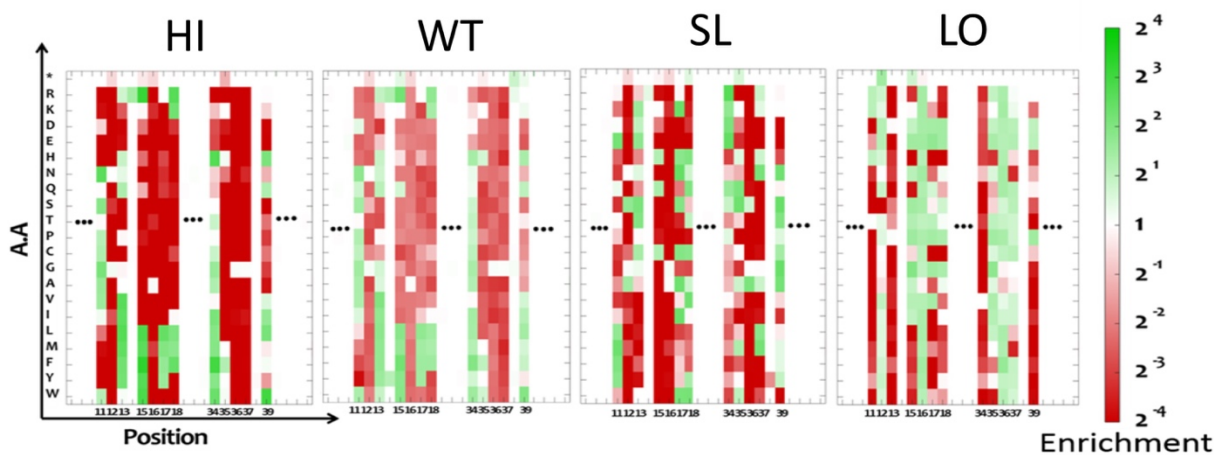
Figure 3: Heatmap showing the enrichment values for each single mutation in the ChT/BPTI complex in the four sorted affinity gates. The enrichment ratio varies from high (green) to low (red) as shown on the right axis. Similar maps were obtained for the other two PPIs.

While the enrichment maps give us qualitative measures of affinity changes due to various mutations, our goal was to construct and compare quantitative binding landscapes of the BPTI/protease interactions. We thus utilized the methodology developed in our recent paper that allows us to normalize the NGS-based enrichments using a small dataset of experimental $\Delta\Delta G_{bind}$ values measured by biophysical techniques on purified proteins (45). We first compiled such normalization datasets for the three complexes, collecting 34 and 33 $\Delta\Delta G_{bind}$ data points from literature for the ChT/BPTI and BT/BPTI interactions, respectively (38, 45-48). For the MT/BPTI interaction, where only a few $\Delta\Delta G_{bind}$ data points have been reported (44, 49), we produced the

8

normalization dataset by expressing and purifying 12 BPTI mutants and measuring their binding affinities to MT (Figure S6). The above datasets were used to obtain a normalization formula for each protease that converts the four enrichment values from the NGS data into the predicted $\Delta\Delta G_{bind}$ values. For all three enzymes, high correlation was found between the $\Delta\Delta G_{bind}$ values predicted from NGS and those experimentally determined using purified proteins (R= ~0.9; Figure S7).

We next used the acquired normalization formulas to predict $\Delta\Delta G_{bind}$ values for all single and double BPTI mutants that have been detected by NGS for the three PPIs. While nearly all BPTI single mutants have been sequenced in all four affinity gates for the three proteases, the double mutants were covered less extensively in the NGS results with only 576, 3393, and 636 double mutants appearing in all four affinity gates for ChT, BT and MT, respectively. To increase the coverage of $\Delta\Delta G_{bind}$ predictions for the double mutants and to complete the predictions for single mutants, we examined whether normalization formulas could be obtained from subsets of three, two, and one affinity gates. While all subsets of gates were examined, only those subsets that produced high correlation with experimental data on pure proteins were selected for the final predictions. For each $\Delta\Delta G_{bind}$ prediction, we estimated the uncertainty in $\Delta\Delta G_{bind}$ predictions using the bootstrapping of the NGS data as described in detail in our previous work (45). Overall, we were able to make reliable predictions for 13,113 double mutants for the BT/BPTI interaction (50% of all binding interface double mutations), 12,537 for ChT/BPTI interaction (47%), and 4317 for MT/BPTI interaction (16 %). We thus constructed full single-mutant binding landscapes and partial-double mutant binding landscapes for BPTI interacting with the three homologous proteases with highly divergent K_{DS}.
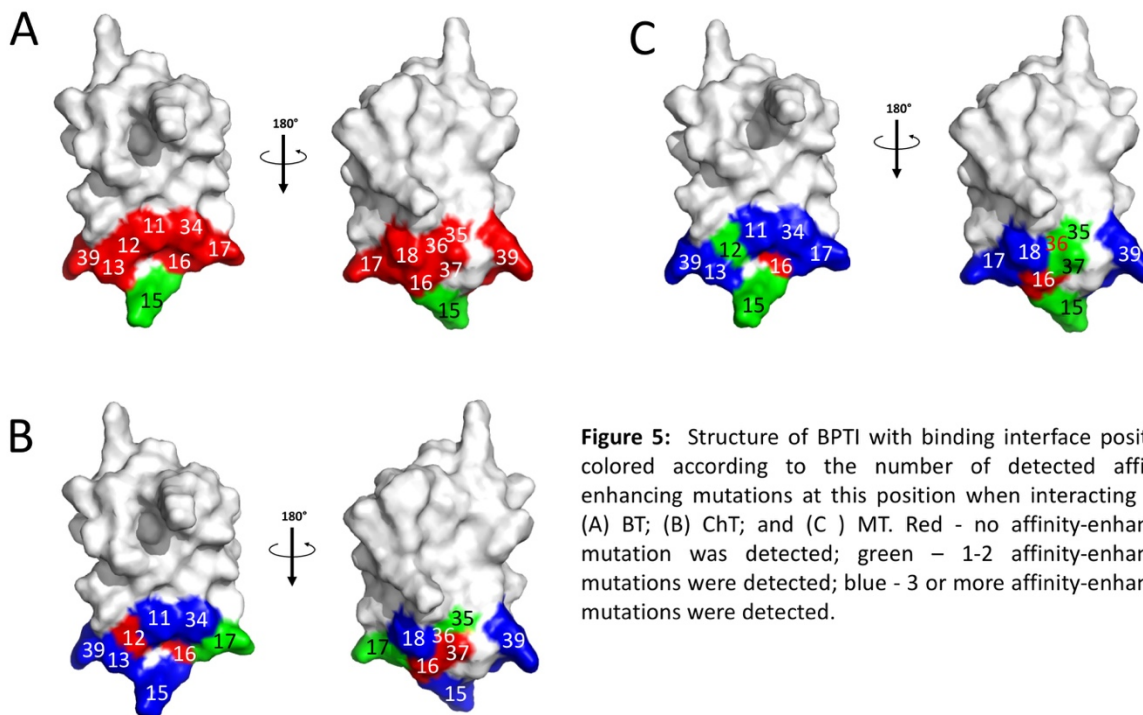
Analysis of the single-mutant binding landscapes. To compare how single mutations affect free energy of binding in the 3 PPIs, we summarized our results in a histogram that includes $\Delta\Delta G_{bind}$ values from all 228 single mutations for each PPI (Figure 4A-C). While all the three histograms show predominance of destabilizing mutations ($\Delta\Delta G_{bind}> 0$), the magnitude of destabilization due to single mutations differs substantially among the three PPIs. For the high-affinity BT/BPTI complex, very high ~12 kcal/mol destabilizations were observed due to some single mutations, medium destabilizations (up to 6 kcal/mol) were observed in the ChT/BPTI complex and small destabilizations (up to ~3 kcal/mol) were observed for the low-affinity MT/BPTI complex (Figure 4 and Figure S8). On average, a single mutation destabilized BT/BPTI interaction by 4.5 kcal/mol,

9

ChT/BPTI interaction by 1.6 kcal/mol and MT/BPTI interaction by 0.82 kcal/mol. On the contrary,
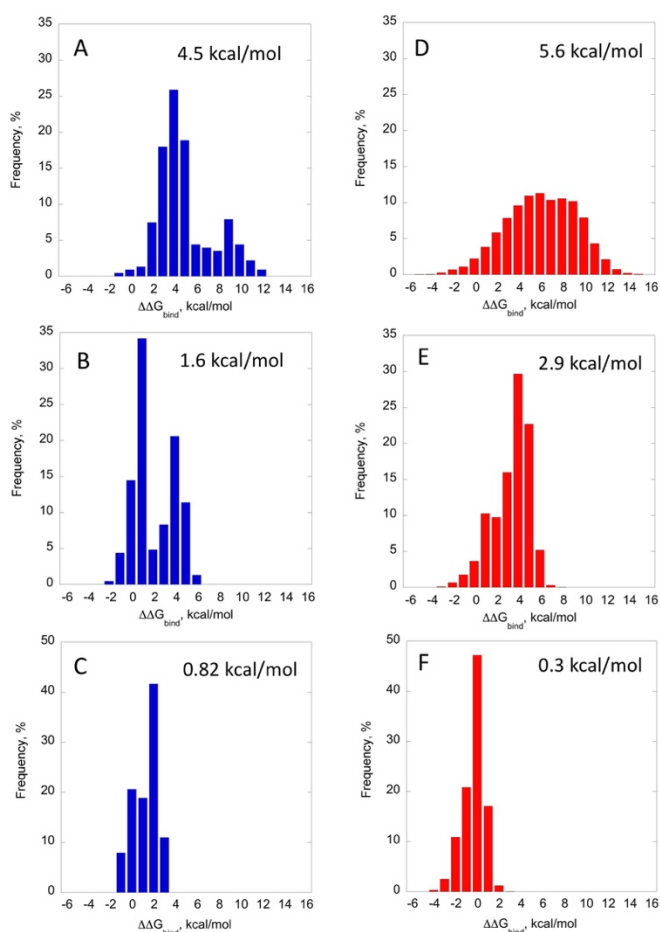
Figure 4: Histograms of $\Delta\Delta G_{bind}$ for single and double BPTI mutants. Single BPTI mutants interacting with (A) BT; (B) ChT; (C) MT. Double BPTI mutants interacting with (D) BT; (E) ChT; (F) MT. Mean value for $\Delta\Delta G_{bind}$ for each histogram is displayed on top of each graph. While all 228 single mutants are incorporated into the single-mutant histograms for all proteases; only ~50% of double mutants are summarized for BT and ChT and only 16% for MT. The data is available in the Source_data_file.

affinity-enhancing mutations appeared more frequently in the low-affinity MT/BPTI complex (50 mutations or 22%), less frequently in the medium-affinity ChT/BPTI complex (37 mutations or 16%), and only once (<1%) in the high-affinity BT/BPTI complex. Per-position analysis of $\Delta\Delta G_{bind}$ values revealed that all but one position on BPTI are absolute hot spots in the BT/BPTI interaction, exhibiting only positive $\Delta\Delta G_{bind}$ values. In contrast, only four absolute hot-spots are present in the ChT/BPTI interaction (positions 12, 16, 36, 37) and only two in the MT/BPTI interaction (position 16 and 36). The spatial distribution of cold-spot and hot-spot positions showed different patterns among the three complexes (Figure 5). Interestingly, position 15, which

10

Figure 5: Structure of BPTI with binding interface positions colored according to the number of detected affinity-enhancing mutations at this position when interacting with (A) BT; (B) ChT; and (C) MT. Red - no affinity-enhancing mutation was detected; green – 1-2 affinity-enhancing mutations were detected; blue - 3 or more affinity-enhancing mutations were detected.

is central to the binding interface, is a cold spot in the ChT/BPTI interaction with all hydrophobic amino acids leading to improved affinity. The same position contains highly destabilizing mutations in both the BT/BPTI and the MT/BPTI complexes, except for one mutation K15R that leads to affinity improvement in both complexes. These differences in the position 15 preferences are in complete agreement with previous studies on purified proteins for the BT/BPTI and ChT/BPTI complexes (47). Additionally, amino acid preferences at position 15 of BPTI discovered here reflect the preferences for substrates that these enzymes cleave (Lys and Arg for trypsins and hydrophobic amino acids for chymotrypsins), indicating that these enzymes have evolved to possess optimal binding pockets for these amino acids.

Analysis of the double-mutant binding landscapes. We next compared the double-mutant binding landscapes for the three PPIs. We first plotted the histograms of $\Delta\Delta G_{bind}$ for the double mutations, for which $\Delta\Delta G_{bind}$ predictions were available (Figure 4D-E). Our results show that on average a double mutation destabilizes the high-affinity BT/BPTI complex by 5.9 kcal/mol, the medium affinity ChT/BPTI complex by 2.9 kcal/mol and the low-affinity MT/BPTI complex by 0.3 kcal/mol, showing the same tendency of increased destabilization due to double mutation with increasing affinity of the PPI as was observed for single mutants. When comparing an average effect from a double and a single mutation, BT/BPTI and ChT/BPTI exhibited higher $\Delta\Delta G_{bind}$

11

value for a double mutation, consistent with a majority of single mutations being destabilizing in these two PPIs. For the low-affinity MT/BPTI complex, the double mutant average was slightly lower compared to that of the single mutant average. This could be due to the fact that a large fraction of single mutants for this PPI leads to affinity improvement or due to the relatively small coverage of double mutants for this PPI in our study (only 16% of double mutations had $\Delta\Delta G_{bind}$ predicted).

Using the extensive $\Delta\Delta G_{bind}$ data for double mutations, we further explored how a single mutational step from a WT sequence alters the distribution of $\Delta\Delta G_{bind}$ values for the second mutation. For this analysis, we selected three representative single BPTI mutants in the highest-affinity BPTI/BT complex: BPTI_K15R that shows slight improvement in affinity compared to $BPTI_{WT}$ ($\Delta\Delta G_{bind}$= -1.4 kcal/mol), BPTI_A16S whose affinity to BT is considerably weaker in comparison to $BPTI_{WT}$ ($\Delta\Delta G_{bind}$= +4.5 kcal/mol) and BPTI_K15A that shows dramatically reduced affinity in comparison to $BPTI_{WT}$ ($\Delta\Delta G_{bind}$= +11.1 kcal/mol). We next compared the $\Delta\Delta G_{bind}$ distributions for single mutations taken on the background of each of the three specified first mutations. While only partial single-mutant landscapes could be constructed for these three BPTI mutants interacting with BT (as we have the data for ~50% of the double mutants), for the detected mutants we observe significant differences in the binding landscapes of the three BPTI mutants with BT ion (Figure S9). K15R that improves the fitness of the BT/BPTI interaction produces a histogram with mostly destabilizing mutations going as far as +12 kcal/mol, yet some affinity-enhancing mutations are also observed. The medium-destabilizing mutation A16S results in a landscape that contains both stabilizing and destabilizing steps with magnitudes ranging from -6 to +8 kcal/mol. The highly destabilizing mutation K15A exhibits a landscape that mostly contains stabilizing mutations with the highest stabilization of -6 kcal/mol. Note that for the K15A mutant we do not observe a mutational step that would reach the affinity of the WT BT/BPTI complex. This is likely due to the fact that position 15 is the most important energetically for the BPTI/protease interaction, thus destroying the favorable interaction at this position could not be fully compensated by any other mutation on BPTI. Our results hence indicate that with every mutational step taken from the WT BPTI sequence the binding landscape would be changed depending on the first mutation; this change is a result of non-additivity of some of single mutations in BPTI.

12

Next, using the extensive quantitative data on $\Delta\Delta G_{bind}$ for single and double mutants, we investigated the extent of coupling between various point mutations in BPTI when it interacts with the three proteases. We have classified mutations into three classes: additive, exhibiting positive and negative epistasis according to the magnitude of the coupling energy ΔGi upon two mutations X and Y:

$$\Delta G_i \ = \Delta\Delta G_{bind}^X + \Delta\Delta G_{bind}^Y - \Delta\Delta G_{bind}^{XY} \tag{1}$$

Here, $\Delta\Delta G_{bind}^X$ and $\Delta\Delta G_{bind}^Y$ represent the change of the binding free energy of the single mutants X and Y, respectively, $\Delta\Delta G_{bind}^{XY}$ represents the change of binding free energy of the double mutant containing mutations X and Y. Two mutations were defined as additive if ΔG_i is zero within the uncertainty of our $\Delta\Delta G_{bind}$ predictions (see Methods for details). Positive epistasis was defined when $\Delta G_i > 0$, i .e the double mutation exhibits higher fitness compared to what is expected from additivity of two single mutations. Negative epistasis was defined when $\Delta G_i < 0$, i .e the double mutation exhibits lower fitness compared to what is expected from additivity of the two single mutations.

Coupling energy analysis shows that in the BT/BPTI interaction, 59% of the detected mutations are additive, 40% of mutations show positive epistasis, and only ~1% of mutations show negative epistasis. In the ChT/BPTI interaction, 74% of mutations are additive, 18% of mutations show positive epistasis and 8% of mutations show negative epistasis. Finally, among the detected mutations in the MT/BPTI interaction we observe 70% of mutations are additive, 26% exhibit positive epistasis and 4% - negative epistasis. Note that the smaller fraction of mutations with negative vs. positive epistasis observed for all complexes could be in part due to the fact that we do not see highly destabilizing double mutations in our experiment and such mutations are likely to be among those exhibiting negative epistasis. We further constructed the per-position correlation matrices displaying coupling energy between all detected mutations in the three PPIs (Figure 6). Figure 6 shows that the sign of the epistasis depends not only on a pair of positions but also on the mutation type. Yet, certain preference for either negative or positive epistasis frequently dominates coupling at certain position pairs as some squares are predominantly red or blue on Figure 6. To analyze which positions exhibit higher degree of coupling, we averaged ΔGi values over all detected mutations at each pair of positions (Figure 7). Figure 7 shows that different proteases

Figure 6: The matrix of ΔGi for double mutations in the three PPIs : (A) BT/BPTI; (B) ChT/BPTI and (C) MT/BPTI. On the left and on the top are BPTI binding interface positions randomized to 20 amino acids. Each square is a 20x20 matrix containing ΔGi values for coupling between a particular mutation at one position to another particular mutation at another position in the same order. Color coding shows the degree of cooperativity: Green: additivity; red - Positive Epistasis; Blue -Negative Epistasis; white - no data is available.

Figure 7: Coupling energies *ΔGi* averaged over all detected mutations at a pair of BPTI positions: (A) BT/BPTI interaction; (B) ChT/BPTI interaction; (C) MT/BPTI interaction. On the top and on the left BPTI positions are shown. The values are color coded from high positive epistasis (dark red) to additive (green) and to negative epistasis (dark blue).
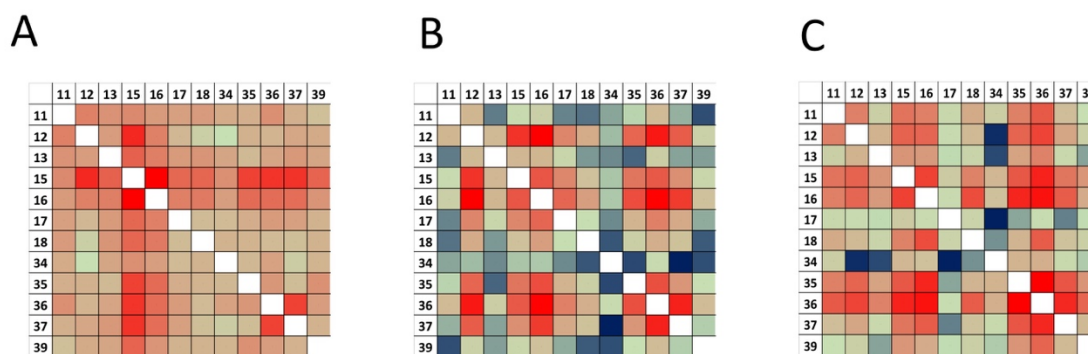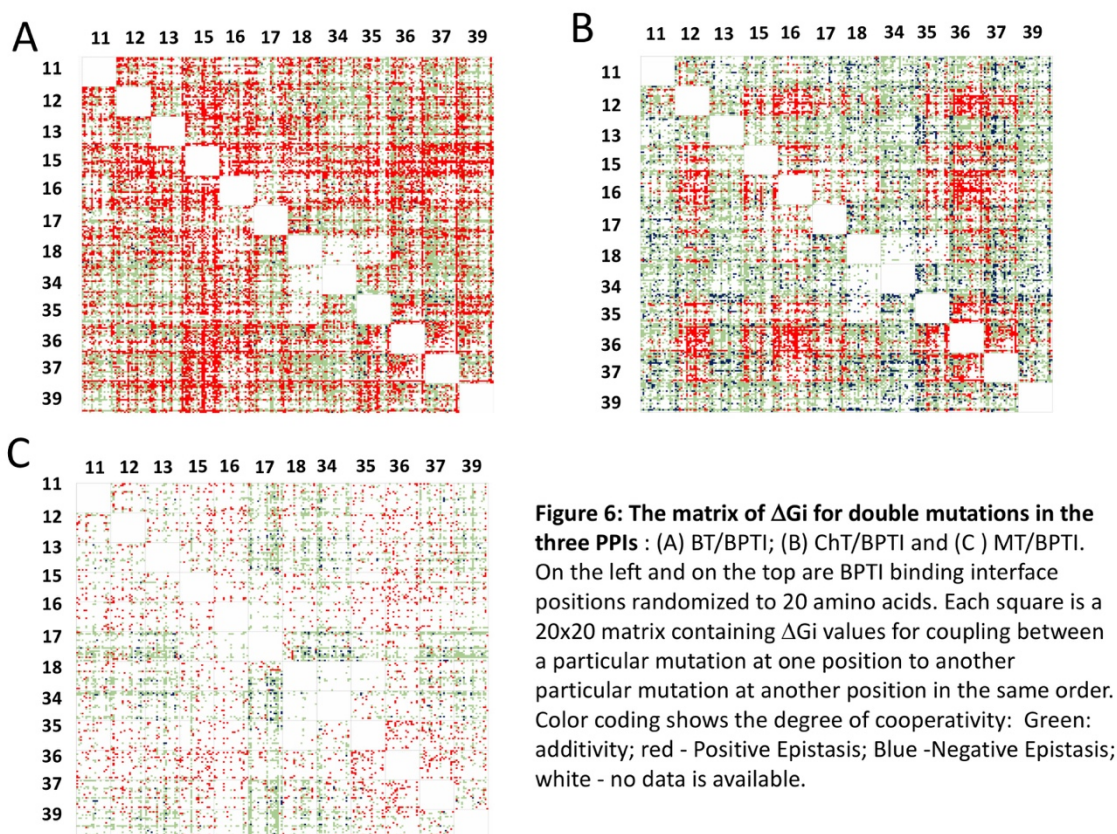
exhibit different patterns of coupling between pairs of mutations. Interestingly, positions where highest destabilization is observed for single mutations show high degree of positive epistasis with all other positions (such as for example positions 15 and 16 in the BPTI/BT complex, position 12 in the ChT/BPTI complex and positions 35, 36 in the MT/BPTI complex). On the other hand, cold

14

spot positions tend to exhibit negative epistasis with many other positions in the protein (see for example, positions 13, 18, 34, 39 in the ChT/BPTI complex and 34 in the MT/BPTI complex). The average coupling energies are larger for the highest affinity BT/BPTI complex, medium for the CT/BPTI complex and the lowest for the MT/BPTI complex (Figure S10). We further tested whether the degree of coupling between the two mutations depends on the distance between the mutated positions (Figure S11). Our results show that mutations at two closely-located positions could exhibit various degrees of coupling from high to low. As the distance is increased between positions, the average coupling between the two mutations at these positions decreases (Figure S11). A similar trend has been observed for all three PPIs and is in agreement with previous studies in various biological systems (50).

Discussion

In this study we measured quantitative effects of tens of thousands of single and double mutational steps in three homologous enzyme-inhibitor complexes. While the complexes are similar in their sequences and structures, they differ greatly in binding affinities that range from ultra-high to low. We find that the binding landscape of each PPI depends strongly on the interaction K_D. In particular, the ultra-high affinity BT/BPTI complex is highly evolutionary optimized. Accordingly, the sequence of WT BPTI lies nearly at the maximum of the binding landscape, with only one mutation leading to significant affinity improvement. The landscape also exhibits a steep gradient, with a majority of single mutations leading to large steps down the hill with a maximum drop of ~12 kcal/mol and an average drop of 4.5 kcal/mol (45) (Figure 8). Such high destabilizations from single mutations are extremely rare. For example, the SKEMPI database (51) that reports 5079 single mutant binding affinity changes in various PPIs contains only 16 single mutations (0.3%) with $\Delta\Delta G_{bind}$ values greater than 8 kcal/mol, and all of them belong to high-affinity complexes (with a K_D of 10^{-12} M or better). The medium-affinity ChT/BPTI complex shows lower degree of evolutionary optimality, with a larger fraction of mutations leading to affinity improvement (16%) and a maximum improvement of 2.6 kcal/mol. Yet even in this complex, single mutational steps could lead to high complex destabilization of up to 6 kcal/mol. Thus, the landscape of the ChT/BPTI complex exhibits a medium gradient and the WT sequence lies about 2/3 up the landscape hill (Figure 8). One might expect that the low-affinity MT/BPTI complex would exhibit $\Delta\Delta G_{bind}$ distribution that is the opposite of that observed for the BPTI/BT

15

complex, with high number of mutations that lead to very large improvement in binding affinity. Yet, this is not what we observe in the present study. The MT/BPTI complex indeed exhibits the highest fraction of mutations leading to affinity improvement among the three complexes (22%) but the largest improvement due to a single mutation does not exceed 1.9 kcal/mol, smaller than what is observed for the ChT/BPTI complex. Yet, the reduction in binding affinity due to single mutations is also the smallest for the MT/BPTI complex not exceeding 3 kcal/mol. Thus, we conclude that the difference between the MT/BPTI and BT/BPTI complexes is not only the location of their sequences relative to the maximum of the binding landscape, but the landscapes themselves show different gradients, high for the high-affinity complex and low for the low-affinity complex (Figure 8).
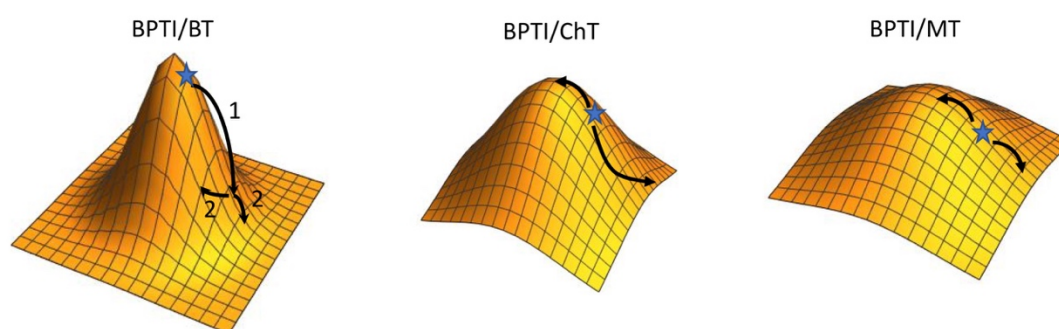
Figure 8: Schematic illustration of the single-mutant binding landscapes for the three studied PPIs. The maximum on the surface corresponds to the highest possible binding affinity. A star indicates the position of the WT BPTI sequence in respect to the maximum. Arrows illustrate how single mutations lead to affinity changes in the three complexes.

The binding landscape characteristics of the three studied PPIs have been dictated by their evolutionary history. BPTI is a cognate inhibitor of BT, thus both proteins have coevolved in one organism to optimize their affinity for each other. ChT and MT bind to BPTI only due to homology to BT. ChT has other cognate inhibitors with which it interacts with much higher affinity, such as turkey ovomucoid third domain (OMTKY3) (K_D of 1.9×10^{-11}) (52). MT, by contrast, appears to have evolved for widespread natural resistance to the tight-binding mechanism of the proteinaceous canonical trypsin inhibitors such as BPTI (53),(54). Furthermore, by binding to these inhibitors orders of magnitude more weakly than other trypsins, mesotrypsin has evolved the capability to cleave many proteinaceous trypsin inhibitors as substrates (55-59). This functional role of the MT/BPTI interaction agrees with our finding that no single or double mutation on the BPTI side could convert this complex into a high-affinity PPI.

16

Analysis of cold spots   In this study we identified a number of cold spot positions in ChT/BPTI and MT/BPTI complexes at which multiple mutations lead to affinity improvement. In our previous study we identified two scenarios of how cold spots occur either through removal of unfavorable interaction with the partner protein or through introduction of a new favorable interaction, where no interaction exists (43). Cold spots discovered here could be easily fit into these two scenarios. We observe the first scenario occurring in the MT/BPTI complex at position 17, where an Arg on BPTI is found in close proximity to an Arg 193 on MT. Substituting Arg 17 with a small and/or hydrophobic amino acid results in affinity improvement. On the contrary, Arg 17 of BPTI is found in a largely hydrophobic environment in the BPTI/ChT complex; its replacement with a hydrophobic Met and Leu results in slightly negative $\Delta\Delta G_{bind}$ values.

We observe the second scenario for cold spot formation at position 34 in the ChT/BPTI and MT/BPTI complexes. V34 at this position does not form any interactions with these proteases. Its replacement with larger hydrophobic amino acids that bury additional surface area increases affinity to ChT. Its replacement with polar or negatively charged residues improves affinity to MT by likely forming new hydrogen bonds to Tyr 151 and/or Gln 192 on MT.

Epistasis in protease/BPTI complexes

Using the data for tens of thousands of double mutants we were able to analyze how two mutations are coupled in the three protease/BPTI complexes. Our data shows that in the high-affinity BT/BPTI complex a large proportion of double mutations results in positive epistasis and only a minority of mutations produces negative epistasis. More equal distribution of mutations with positive and negative epistasis is observed for the ChT/BPTI complex. Note, that we only observe ~50% of all double mutations for these two PPIs; the remaining 50% that are invisible in our experiment are likely to be highly destabilizing, which would put them in the category of either additive or showing negative epistasis. The abundance of positive epistasis in the BT/BPTI interaction could be explained from the perspective of binding landscape theory (Figure 8). Due to the steepness of the gradient in the area of the wild-type BPTI sequence, the first mutation in this PPI leads to a large step down the hill into the area of low-gradient. Second mutation from this point could lead up or down, but the change would be relatively small, resulting in positive epistasis, i.e., better $\Delta\Delta G_{bind}$ compared to what would be predicted from additivity of the two highly destabilizing mutations. Positive epistasis particularly predominates at positions where the

17

largest affinity drops are recorded (such as at positions 15 and 16 for the BT/BPTI complex or position 12 for the ChT/BPTI complex), where the gradient is steep.

Positive epistasis could be also explained from the structural perspective. Highly optimized PPIs usually retain their original binding conformation upon introduction of a single mutation due to the abundance of favorable interactions generated at non-mutated positions. Yet, if any of the hot-spot residues is mutated, substantially weakening the interaction, then the impact of a second deleterious mutation may be mitigated by an increase in flexibility at the interface, enabling the protein to adopt alternative conformations that introduce new favorable intermolecular contacts and enhance affinity. If the same mutation would occur on the background of the wild-type residue in the hot-spot position, the new conformation would not be accessible and the affinity enhancement would not be achieved. Thus, such a double mutant would possess better $\Delta\Delta G_{bind}$ compared to the sum of single mutants, exhibiting positive epistasis.

On the contrary, negative epistasis is more frequent for medium and low-affinity PPIs and appears mostly when one mutation is performed at a cold-spot position. If the first mutation improves binding affinity and thus makes a step up the binding landscape towards the maximum, the second mutation would be made from the point of steeper gradient and is likely to make a large step down, thus resulting in negative epistasis. Structurally that means that if at one cold spot position a new favorable interaction was created, this interaction might lock the PPI into a new slightly different conformation. Another conformation might be acquired upon introduction of a different affinity-enhancing mutation. But the two favorable conformations could not be achieved simultaneously, resulting in worse $\Delta\Delta G_{bind}$ for a double mutation compared to the sum of two single mutations (negative epistasis).

In summary, in this study we report $\Delta\Delta G_{bind}$ values for tens of thousands of single and double mutations in three protease/BPTI complexes with similar structures but highly variable binding affinities, thus generating an unprecedented amount of mutational data that could be used as benchmark for testing new computational methodology and for design of new high-affinity protease inhibitors. Using the obtained data, we demonstrate striking differences between the binding landscapes of the three PPIs that could be explained by the level of the PPI evolutionary optimality. Furthermore, we study how two single mutations in these PPIs couple to each other and demonstrate that the coupling energy depends not only on positions of mutations but also on the identities of the mutated amino acid. Furthermore, we observe that mutations at hot-spot

positions generally exhibit positive epistasis with other mutations while mutations at cold-spot positions generally exhibit negative epistasis and explain this phenomenon from the perspective of binding landscape theory. Our powerful experimental methodology could be used to access the binding landscapes in many additional PPIs with different structures, functions, and affinities and to probe whether the reported evolutionary trends hold in other biological systems.

Author contributions

J. M. S. and N. P. designed the research. M. H., J. S. and I. C. performed the research. M. H. and J. M. S. performed data analysis. Y. P. provided help with combinatorial library construction and transformation. E. S. R. provided protease samples. J. M. S. wrote the manuscript with assistance from N. P. and E. S. R.

Methods

BPTI library construction. Twelve positions on BPTI that lie in the binding interface with BT (PDB ID 3OTJ) were subject to randomization: T11, G12, P13, K15, A16, R17, I18, V34, Y35, G36, G37 and R39. A BPTI library was constructed that randomized two positions at a time with an NNS codon (where N = A/C/G/T DNA base, S = C/G DNA base); encoding all amino acids at the randomized positions, including the WT amino acid. The library was divided into 66 sub-libraries that each incorporates all possible pairs of the twelve randomized positions. TPCR protocol(60) was used to create each library using two primers that either combined two mutations in one primer or divided them into two primers depending on their proximity to each other (Supplementary Note 1 in Supplementary information). These primers were used in a PCR

19

together with BPTI$_{WT}$ plasmid to incorporate these mutations at the specific positions in BPTI and to amplify the whole plasmid. Agarose gel analysis was used to confirm the success of each TPCR reaction. The TPCR- amplified plasmid DNA was treated with DpnI (New England Biolabs, Ipswich, MA) to remove any parental plasmid used as a template to construct the library, cleaned up with magnetic beads (AMPure XP, Beckman coulter, Brea, CA), transferred into *E. coli* and selected colonies were sequenced to confirm the successful generation and transformation of the BPTI library. The vectors containing the BPTI library were extracted using QIAprep Spin miniprep (Qiagen, Hilden, Germany) and all the sub-libraries were pooled together and balanced by their DNA amount to use the same amount of DNA from each sub-library (~3.6μg). Then, the pooled library was transferred into *S. cerevisiae* using 20 transformations resulting in 60,000 – 70,000 colonies for the complete library as estimated by plating 1/20 amount of the library sample and counting the colonies after transformation on a SDCAA plate.

NGS analysis. The paired-end reads from the NGS experiments were merged(61) and their quality scores were calculated in the FastQC tool (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). In the Matlab script, the sequences were aligned, and sequences containing extra mutations at non-randomized positions were filtered out. The number of each remaining BPTI mutant i was counted in the sorted and the naïve populations and its frequency f^i in the libraries was calculated (**Eq. 2**). Using the frequency of the mutant in one of the sorted populations and the naïve population, the enrichment e^i of each BPTI mutant was calculated (**Eq**. **3**).

$$f^i = \frac{count^i}{count_{total}} \tag{2}$$

$$e^i = \ln\left(\frac{(f^i)_{sorted}}{(f^i)_{naïve}}\right) \tag{3}$$

To estimate the uncertainty in BPTI mutant frequencies we applied a bootstrapping method to the NGS data for all sorted gates and the naïve library as described in(62). Briefly, the original NGS data was used to randomly draw sequences to obtain a resampling data set of the same size and to calculate the frequency of each BPTI mutant in each population. The resampling process was repeated 1000 times and the average frequency and the standard deviation was calculated from 1000 resampling data sets for each BPTI mutant in each sorting gate and in the naïve library. The

error was propagated into equations 2 and 3 to calculate the error in enrichment values as described in(45).

Predicting ΔΔG$_{bind}$ values from NGS data. All available experimental data on ΔΔG$_{bind}$ for the BPTI/protease complexes was used to obtain the best normalization formulas for each complex for converting enrichment values from the four sorted populations into ΔΔG$_{bind}$ values. To this end, we used a linear regression model function in Mathematica (Wolfram Research) with 5 parameters (Y= aX_1+bX_2+cX_3+dX_4+f) if all four enrichment values were available in our NGS data for this particular mutation. The parameters a, b, c, d, f were optimized using the experimental data set as values of Y and the set of X_1, X_2, X_3, X_4 enrichment values. The obtained normalization formula (different for each protease) was used to calculate ΔΔG$_{bind}$ values for all the remaining single and double BPTI mutants that had four enrichment values recorded in the NGS experiment. To make ΔΔG$_{bind}$ prediction for mutants where fewer than 4 enrichment values were available, we repeated the normalization procedure using different subsets of enrichment values (such as X_1 and X_4; X_1, X_2, X_3; etc.). Accordingly, we varied the number of parameters in the normalization formula. We checked whether high correlation with the experimental data set of ΔΔG$_{bind}$ values could be obtained using this particular subset of variables as predictors. If a correlation of R>0.80 was obtained between the predicted and the experimental ΔΔG$_{bind}$ values, the set of gates was selected as good for making predictions. Additional cross check for validity of predictions from this subset of gates was performed by comparing ΔΔG$_{bind}$ predictions for all single mutants based on all four gates and based on the selected subset of gates and confirming high correlation between the two predicted ΔΔG$_{bind}$ values over all single mutations. For each of the mutant, we used the available enrichment values to make separate predictions from all possible "good" subset of gates. First all predictions were recorded for mutants where enrichment values were available for all 4 gates. For mutants where predictions were available for only gates X1, X2, and X4, predictions were made based on these 3 gates providing that this set of gates was defined as good. For mutants where predictions were available in only gates X1 and X4 predictions were made based on these two gates if this set of gates was defined as good for predictions. For each prediction from each subset of gates, the uncertainty of the prediction was calculated using the procedure outlined above. Finally, for each mutation, ΔΔG$_{bind}$ prediction was selected from all the predictions according to the subset of gates where the highest correlation with experimental data was observed.

The dataset of final $\Delta\Delta G_{bind}$ predictions for all single and double mutants for the three PPIs could be found in the Dataset S1 file.

Production of BPTI variants.

The BPTI$_{WT}$ sequence was cloned into a pPIC9K vector (Invitrogen, Carlsbad, CA), TPCR was used for site-directed mutagenesis to create the sequences of all variants, transformed, expressed in *P. pastoris* (*GS115* strain; obtained from Invitrogen, Carlsbad, CA) and purified by nickel affinity chromatography, followed by size-exclusion chromatography, as described in a previous work[85]. The correct DNA sequence of each produced protein was confirmed by extracting the plasmidic DNA from *P. pastoris* after protein purification by nickel chromatography, amplifying the BPTI gene and sequencing it. Protein purity was validated by SDS-PAGE on a 20% polyacrylamide gel, and the mass was determined with a MALDI-TOF REFLEX-IV (Bruker) mass spectrometer (IKI, BGU; data not shown). Purification yields for all BPTI variants were 2-15 mg per liter of medium. The concentration of purified BPTI variants was determined by an activity assay.

For MT, values of the inhibition constant (K_i) were determined using a general enzyme activity essay for PPIs characterized by medium to low affinity[62]. Here, 304 µL BPTI (4 different concentrations ranging from 5.2 to 52.6 µM) was mixed with 8 µL of the substrate Z-GPR-pNA (Sigma-Aldrich, St. Louis, MO) (5 different concentrations ranging from 0.4 to 10 mM). The mix was incubated for three minutes. Then, the reaction was initialized by adding 8 µL MT (10nM) and the absorbance of the samples was measured at a wavelength of 410 nm for 5 minutes. A negative control was added replacing BPTI with 304 µL buffer (10 mM Tris, pH8, 1 mM CaCl$_2$). The range of concentrations of BPTI was adapted when the determined K_i was not in the range of these BPTI concentrations.

For MT, the K_i could be determined from **Eq. 4**, as described previously (63).

$$v = \frac{K_{cat}[E]_0[S]}{K_m\left(1+\frac{[I]}{K_i}\right)+[S]} \tag{4}$$

The change in experimental binding energy $\Delta\Delta G_{bind}$ was calculated from **Eq. 5** using the K_i of the WT and the mutant, the temperature T at which the affinity was measured and the ideal gas constant R.

$$\Delta\Delta G_{bind} = R * T \ln\frac{K_i(mutant)}{K_i(WT)} \tag{5}$$

Analysis of additivity and cooperativity For each double mutation with available $\Delta\Delta G_{bind}$ prediction, we calculated the interaction energy between the two single mutations according to eq. (1).

The mutation was defined as exhibiting negative epistasis if ΔG_i was negative within the uncertainty of the predictions, that is:

$$\Delta G_i = \Delta\Delta G_{bind}^X + \Delta\Delta G_{bind}^Y + \sqrt{\delta\Delta\Delta G_{bind}^X{}^2 + \delta\Delta\Delta G_{bind}^Y{}^2} - \Delta\Delta G_{bind}^{XY} - \delta\Delta\Delta G_{bind}^{XY} < 0 \quad (6).$$

Here, $\delta\Delta\Delta G_{bind}^X$, $\delta\Delta\Delta G_{bind}^Y$ and $\delta\Delta\Delta G_{bind}^{XY}$ are uncertainties in prediction of $\Delta\Delta G_{bind}$ for mutation X, Y, and XY, respectively.

The mutation was defined as exhibiting positive epistasis if ΔG_i was positive within the uncertainty of the predictions, that is:

$$\Delta G_i = \Delta\Delta G_{bind}^X + \Delta\Delta G_{bind}^Y - \sqrt{\delta\Delta\Delta G_{bind}^X{}^2 + \delta\Delta\Delta G_{bind}^Y{}^2} - \Delta\Delta G_{bind}^{XY} + \delta\Delta\Delta G_{bind}^{XY} > 0 \quad (7).$$

If the value of ΔG_i fell between the values of positive or negative epistasis the mutation was defined as additive.

Competing interests

The authors declare no competing interests.

References

1. Klesmith JR, Bacik JP, Wrenbeck EE, Michalczyk R, & Whitehead TA (2017) Trade-offs between enzyme fitness and solubility illuminated by deep mutational scanning. *Proc Natl Acad Sci U S A* 114(9):2265-2270.
2. Mavor D, Barlow K, Thompson S, Barad BA, Bonny AR, Cario CL, Gaskins G, Liu Z, Deming L, Axen SD, Caceres E, Chen W, Cuesta A, Gate RE, Green EM, Hulce KR, Ji W, Kenner LR, Mensa B, Morinishi LS, Moss SM, Mravic M, Muir RK, Niekamp S, Nnadi CI, Palovcak E, Poss EM, Ross TD, Salcedo EC, See SK, Subramaniam M, Wong AW, Li J, Thorn KS, Conchuir SO, Roscoe BP, Chow ED, DeRisi JL, Kortemme T, Bolon DN, & Fraser JS (2016) Determination of ubiquitin fitness landscapes under different chemical stresses in a classroom setting. *Elife* 5.
3. Harms MJ & Thornton JW (2013) Evolutionary biochemistry: revealing the historical and physical causes of protein properties. *Nat Rev Genet* 14(8):559-571.
4. Cho S, Swaminathan CP, Yang J, Kerzic MC, Guan R, Kieke MC, Kranz DM, Mariuzza RA, & Sundberg EJ (2005) Structural basis of affinity maturation and intramolecular cooperativity in a protein-protein interaction. *Structure* 13(12):1775-1787.
5. Bloom JD, Romero PA, Lu Z, & Arnold FH (2007) Neutral genetic drift can alter promiscuous protein functions, potentially aiding functional evolution. *Biol Direct* 2:17.

6.   Ernst A, Avvakumov G, Tong J, Fan Y, Zhao Y, Alberts P, Persaud A, Walker JR, Neculai AM, Neculai D, Vorobyov A, Garg P, Beatty L, Chan PK, Juang YC, Landry MC, Yeh C, Zeqiraj E, Karamboulas K, Allali-Hassani A, Vedadi M, Tyers M, Moffat J, Sicheri F, Pelletier L, Durocher D, Raught B, Rotin D, Yang J, Moran MF, Dhe-Paganon S, & Sidhu SS (2013) A strategy for modulation of enzymes in the ubiquitin system. *Science* 339(6119):590-595.

7.   Goldsmith M & Tawfik DS (2012) Directed enzyme evolution: beyond the low-hanging fruit. *Curr Opin Struct Biol* 22(4):406-412.

8.   de Visser J, Elena SF, Fragata I, & Matuszewski S (2018) The utility of fitness landscapes and big data for predicting evolution. *Heredity (Edinb)* 121(5):401-405.

9.   Whitehead TA, Chevalier A, Song Y, Dreyfus C, Fleishman SJ, De Mattos C, Myers CA, Kamisetty H, Blair P, Wilson IA, & Baker D (2012) Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat Biotechnol* 30(6):543-548.

10.   Elazar A, Weinstein J, Biran I, Fridman Y, Bibi E, & Fleishman SJ (2016) Mutational scanning reveals the determinants of protein insertion and association energetics in the plasma membrane. *Elife* 5.

11.   Sarkisyan KS, Bolotin DA, Meer MV, Usmanova DR, Mishin AS, Sharonov GV, Ivankov DN, Bozhanova NG, Baranov MS, Soylemez O, Bogatyreva NS, Vlasov PK, Egorov ES, Logacheva MD, Kondrashov AS, Chudakov DM, Putintseva EV, Mamedov IZ, Tawfik DS, Lukyanov KA, & Kondrashov FA (2016) Local fitness landscape of the green fluorescent protein. *Nature* 533(7603):397-401.

12.   Hietpas RT, Jensen JD, & Bolon DN (2011) Experimental illumination of a fitness landscape. *Proc Natl Acad Sci U S A* 108(19):7896-7901.

13.   Canale AS, Cote-Hammarlof PA, Flynn JM, & Bolon DN (2018) Evolutionary mechanisms studied through protein fitness landscapes. *Curr Opin Struct Biol* 48:141-148.

14.   Firnberg E, Labonte JW, Gray JJ, & Ostermeier M (2014) A comprehensive, high-resolution map of a gene's fitness landscape. *Mol Biol Evol* 31(6):1581-1592.

15.   Stiffler MA, Hekstra DR, & Ranganathan R (2015) Evolvability as a function of purifying selection in TEM-1 beta-lactamase. *Cell* 160(5):882-892.

16.   Wrenbeck EE, Azouz LR, & Whitehead TA (2017) Single-mutation fitness landscapes for an enzyme on multiple substrates reveal specificity is globally encoded. *Nat Commun* 8:15695.

17.   Wrenbeck EE, Faber MS, & Whitehead TA (2017) Deep sequencing methods for protein engineering and design. *Curr Opin Struct Biol* 45:36-44.

18.   Melnikov A, Rogov P, Wang L, Gnirke A, & Mikkelsen TS (2014) Comprehensive mutational scanning of a kinase in vivo reveals substrate-dependent fitness landscapes. *Nucleic Acids Res* 42(14):e112.

19.   Fowler DM, Stephany JJ, & Fields S (2014) Measuring the activity of protein variants on a large scale using deep mutational scanning. *Nat Protoc* 9(9):2267-2284.

20.   Fowler DM, Araya CL, Fleishman SJ, Kellogg EH, Stephany JJ, Baker D, & Fields S (2010) High-resolution mapping of protein sequence-function relationships. *Nat Methods* 7(9):741-746.

21. Kowalsky CA, Klesmith JR, Stapleton JA, Kelly V, Reichkitzer N, & Whitehead TA (2015) High-resolution sequence-function mapping of full-length proteins. *PLoS One* 10(3):e0118193.

22. Yang G, Anderson DW, Baier F, Dohmen E, Hong N, Carr PD, Kamerlin SCL, Jackson CJ, Bornberg-Bauer E, & Tokuriki N (2019) Higher-order epistasis shapes the fitness landscape of a xenobiotic-degrading enzyme. *Nat Chem Biol* 15(11):1120-1128.

23. Kumar H & Shifman JM (2020) Predicting mutational effects. *In Protein interactions: computational methods, analysis and applications*, ed Gromiha MM (World Scientific Publishing Singapore).

24. Jubb HC, Pandurangan AP, Turner MA, Ochoa-Montano B, Blundell TL, & Ascher DB (2017) Mutations at protein-protein interfaces: Small changes over big surfaces have large impacts on human health. *Prog Biophys Mol Biol* 128:3-13.

25. Gonzalez MW & Kann MG (2012) Chapter 4: Protein interactions and disease. *PLoS Comput Biol* 8(12):e1002819.

26. Pal G, Kouadio JL, Artis DR, Kossiakoff AA, & Sidhu SS (2006) Comprehensive and quantitative mapping of energy landscapes for protein-protein interactions by rapid combinatorial scanning. *J Biol Chem* 281(31):22378-22385.

27. Leung I, Dekel A, Shifman JM, & Sidhu SS (2016) Saturation scanning of ubiquitin variants reveals a common hot spot for binding to USP2 and USP21. *Proc Natl Acad Sci U S A* 113(31):8705-8710.

28. Aizner Y, Sharabi O, Shirian J, Dakwar G, Abvraham O, Risman M, & Shifman JM (2014) Mapping the binding landscape of a picomolar protein-protein complex through computation and experiment. *Structure* 22:1-10.

29. Kowalsky CA & Whitehead TA (2016) Determination of binding affinity upon mutation for type I dockerin-cohesin complexes from Clostridium thermocellum and Clostridium cellulolyticum using deep sequencing. *Proteins* 84(12):1914-1928.

30. McLaughlin RN, Jr., Poelwijk FJ, Raman A, Gosal WS, & Ranganathan R (2012) The spatial architecture of protein function and adaptation. *Nature* 491(7422):138-142.

31. Moal IH & Fernandez-Recio J (2012) SKEMPI: a Structural Kinetic and Energetic database of Mutant Protein Interactions and its use in empirical models. *Bioinformatics* 28(20):2600-2607.

32. Ashkenazi A, Presta LG, Marsters SA, Camerato TR, Rosenthal KA, Fendly BM, & Capon DJ (1990) Mapping the CD4 binding site for human immunodeficiency virus by alanine-scanning mutagenesis. *Proc Natl Acad Sci U S A* 87(18):7150-7154.

33. Cunningham BC & Wells JA (1989) High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science* 244(4908):1081-1085.

34. Weiss GA, Watanabe CK, Zhong A, Goddard A, & Sidhu SS (2000) Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc Natl Acad Sci U S A* 97(16):8950-8954.

35. Wang CY, Chang PM, Ary ML, Allen BD, Chica RA, Mayo SL, & Olafson BD (2018) ProtaBank: A repository for protein design and engineering data. *Protein Sci* 27(6):1113-1124.

36. Lu SM, Lu W, Qasim MA, Anderson S, Apostol I, Ardelt W, Bigler T, Chiang YW, Cook J, James MN, Kato I, Kelly C, Kohr W, Komiyama T, Lin TY, Ogawa M, Otlewski J, Park SJ, Qasim S, Ranjbar M, Tashiro M, Warne N, Whatley H, Wieczorek A, Wieczorek M, Wilusz T, Wynn R, Zhang W, & Laskowski M, Jr. (2001) 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(4):1410-1415.

37. Meenan NA, Sharma A, Fleishman SJ, Macdonald CJ, Morel B, Boetzel R, Moore GR, Baker D, & Kleanthous C (2010) The structural and energetic basis for high selectivity in a high-affinity protein-protein interaction. *Proc Natl Acad Sci U S A* 107(22):10080-10085.

38. Castro MJ & Anderson S (1996) Alanine point-mutations in the reactive region of bovine pancreatic trypsin inhibitor: effects on the kinetics and thermodynamics of binding to beta-trypsin and alpha-chymotrypsin. *Biochemistry* 35(35):11435-11446.

39. Sharabi O, Shirian J, Grossman M, Sagi I, & Shifman JM (2014) Affinity- and specificity-enhancing mutations are frequent in mulstispecific interaction between MMP14 and its inhibitor TIMP2. *PloS ONE* 9(4):e93712.

40. DeLano WL (2002) Unraveling hot spots in binding interfaces: progress and challenges. *Curr Opin Struct Biol* 12(1):14-20.

41. Clackson T & Wells JA (1995) A hot spot of binding energy in a hormone-receptor interface. *Science* 267(5196):383-386.

42. Bogan AA & Thorn KS (1998) Anatomy of hot spots in protein interfaces. *J Mol Biol* 280(1):1-9.

43. Shirian J, Sharabi O, & Shifman JM (2016) Cold-spots in Protein Binding. *Trends Biochem Sci* 41(9):739-745.

44. Salameh MA, Soares AS, Navaneetham D, Sinha D, Walsh PN, & Radisky ES (2010) Determinants of affinity and proteolytic stability in interactions of Kunitz family protease inhibitors with mesotrypsin. *J Biol Chem* 285(47):36884-36896.

45. Heyne M, Papo N, & Shifman JM (2020) Generating quantitative binding landscapes through fractional binding selections combined with deep sequencing and data normalization. *Nature Communications* (11(1)):1-7.

46. Krowarsch D, Dadlez M, Buczek O, Krokoszynska I, Smalas AO, & Otlewski J (1999) Interscaffolding additivity: binding of P1 variants of bovine pancreatic trypsin inhibitor to four serine proteases. *J Mol Biol* 289(1):175-186.

47. Otlewski J, Jaskolski M, Buczek O, Cierpicki T, Czapinska H, Krowarsch D, Smalas AO, Stachowiak D, Szpineta A, & Dadlez M (2001) Structure-function relationship of serine protease-protein inhibitor interaction. *Acta Biochim Pol* 48(2):419-428.

48. Beeser SA, Goldenberg DP, & Oas TG (1997) Enhanced protein flexibility caused by a destabilizing amino acid replacement in BPTI. *J Mol Biol* 269(1):154-164.

49. Salameh MA, Soares AS, Hockla A, Radisky DC, & Radisky ES (2011) The P(2)' residue is a key determinant of mesotrypsin specificity: engineering a high-affinity inhibitor with anticancer activity. *Biochem J* 440(1):95-105.

50. Jemimah S & Gromiha MM (2018) Exploring additivity effects of double mutations on the binding affinity of protein-protein complexes. *Proteins* 86(5):536-547.

51. Jankauskaite J, Jimenez-Garcia B, Dapkunas J, Fernandez-Recio J, & Moal IH (2019) SKEMPI 2.0: an updated benchmark of changes in protein-protein binding energy, kinetics and thermodynamics upon mutation. *Bioinformatics* 35(3):462-469.

52. Lu W, Apostol I, Qasim MA, Warne N, Wynn R, Zhang WL, Anderson S, Chiang YW, Ogin E, Rothberg I, Ryan K, & Laskowski M, Jr. (1997) Binding of amino acid side-chains to S1 cavities of serine proteinases. *J Mol Biol* 266(2):441-461.

53. Nyaruhucha CN, Kito M, & Fukuoka SI (1997) Identification and expression of the cDNA-encoding human mesotrypsin(ogen), an isoform of trypsin with inhibitor resistance. *J Biol Chem* 272(16):10573-10578.
54. Katona G, Berglund GI, Hajdu J, Graf L, & Szilagyi L (2002) Crystal structure reveals basis for the inhibitor resistance of human brain trypsin. *J Mol Biol* 315(5):1209-1218.
55. Salameh MA, Soares AS, Hockla A, & Radisky ES (2008) Structural basis for accelerated cleavage of bovine pancreatic trypsin inhibitor (BPTI) by human mesotrypsin. *J Biol Chem* 283(7):4115-4123.
56. Salameh MA, Robinson JL, Navaneetham D, Sinha D, Madden BJ, Walsh PN, & Radisky ES (2010) The amyloid precursor protein/protease nexin 2 Kunitz inhibitor domain is a highly specific substrate of mesotrypsin. *J Biol Chem* 285(3):1939-1949.
57. Szmola R, Kukor Z, & Sahin-Toth M (2003) Human mesotrypsin is a unique digestive protease specialized for the degradation of trypsin inhibitors. *J Biol Chem* 278(49):48580-48589.
58. Pendlebury D, Wang R, Henin RD, Hockla A, Soares AS, Madden BJ, Kazanov MD, & Radisky ES (2014) Sequence and conformational specificity in substrate recognition: several human Kunitz protease inhibitor domains are specific substrates of mesotrypsin. *J Biol Chem* 289(47):32783-32797.
59. Alloy AP, Kayode O, Wang R, Hockla A, Soares AS, & Radisky ES (2015) Mesotrypsin Has Evolved Four Unique Residues to Cleave Trypsin Inhibitors as Substrates. *J Biol Chem* 290(35):21523-21535.
60. Erijman A, Dantes A, Bernheim R, Shifman JM, & Peleg Y (2011) Transfer-PCR (TPCR): a highway for DNA cloning and protein engineering. *J Struct Biol* 175(2):171-177.
61. Magoc T & Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27(21):2957-2963.
62. Kulesa A, Krzywinski M, Blainey P, & Altman N (2015) Sampling distributions and the bootstrap. *Nat Methods* 12(6):477-478.
63. Cohen I, Kayode O, Hockla A, Sankaran B, Radisky DC, Radisky ES, & Papo N (2016) Combinatorial protein engineering of proteolytically resistant mesotrypsin inhibitors as candidates for cancer therapy. *Biochem J* 473(10):1329-1341.