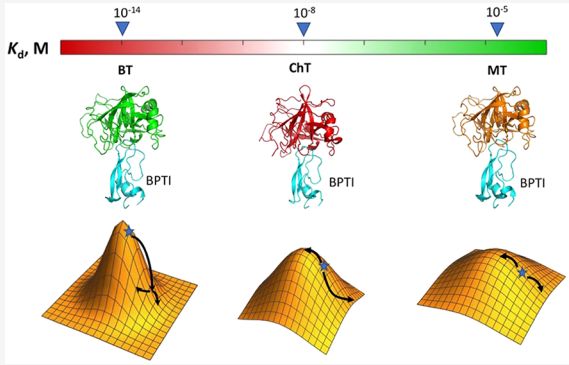
JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS
Article

Climbing Up and Down Binding Landscapes through Deep Mutational Scanning of Three Homologous Protein−Protein Complexes

Michael Heyne, Jason Shirian, Itay Cohen, Yoav Peleg, Evette S. Radisky, Niv Papo,* and Julia M. Shifman*

Cite This: J. Am. Chem. Soc. 2021, 143, 17261−17275
Read Online

ABSTRACT: Protein−protein interactions (PPIs) have evolved to display binding affinities that can support their function. As such, cognate and noncognate PPIs could be highly similar structurally but exhibit huge differences in binding affinities. To understand this phenomenon, we study three homologous protease−inhibitor PPIs that span 9 orders of magnitude in binding affinity. Using state-of-the-art methodology that combines protein randomization, affinity sorting, deep 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. 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 other biological complexes in the cell.



INTRODUCTION

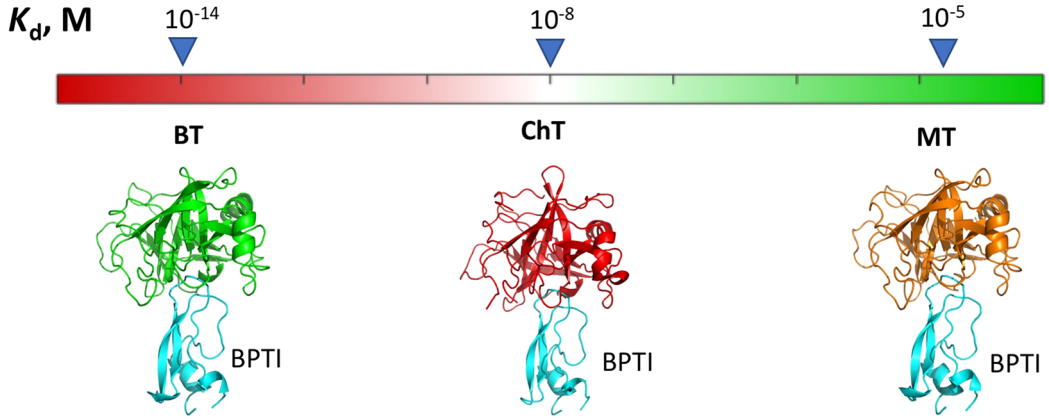
Protein function is determined by the protein amino acid sequence, which has undergone billions of years of evolution while subjected to various selection pressures. Native proteins have evolved not only to perform their main function but also to satisfy a number of criteria such as solubility,[1] low propensity for aggregation, stability, resistance to stress conditions,[2] etc. As a result of these opposing pressures[3,4] and mutation-selection balance,[5] proteins usually function below their maximum capacity.[6,7] 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.[8−11]

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.[12] 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.[13,14] 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.[1,15−27] 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.

Binding between two or more protein partners represents one of many important protein functions. 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.[28] 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 signal transduction pathways and disease.[29,30] Therefore, understanding how mutations in PPIs affect their binding affinity is of great importance to both basic

Received: August 17, 2021
Published: October 5, 2021

https://doi.org/10.1021/jacs.1c08707
J. Am. Chem. Soc. 2021, 143, 17261−17275



	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 three PPIs. Structures of the three complexes between BPTI and BT (PDB ID 3OTJ), ChT (PDB ID 1CBW), and MT (PDB ID 2R9P) are shown. Summarized are inhibition constants, change in solvent-accessible surface area upon complex formation in A^2 (ΔASA), number of intra- and inter-molecular hydrogen bonds, salt bridges, van der Waals (vdW) energy, and interface sequence identity of proteases relative to BT.

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,[13,31−36] while for most PPIs only a few $\Delta\Delta G_{bind}$ data points have been measured, most frequently involving mutations to alanines.[37−41] Comparison of the available sparse $\Delta\Delta G_{bind}$ data from different studies led us to 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,[33,42−44] in other PPIs frequent affinity-enhancing mutations are observed.[32,45] 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 a large reduction in affinity.[46−48] Cold-spot positions, on the other hand, present multiple possibilities for PPI affinity improvement.[49]

To investigate the basis for large differences in binding affinity between evolutionarily optimized cognate PPIs and nonoptimized noncognate PPIs, we compare comprehensive binding landscapes of three structurally similar PPIs that span 9 orders of magnitude in binding affinity (K_D) (Figure 1). We examine the interaction of bovine pancreatic trypsin inhibitor (BPTI) with its coevolved biological target bovine trypsin (BT) ($K_D = 10^{-14}$ M)[44] and with noncognate trypsin paralogs bovine α-chymotrypsin (ChT) ($K_D = 10^{-8}$ M)[44] and human mesotrypsin (MT) ($K_D = 10^{-5}$ M).[50] BPTI has coevolved with trypsin to protect the pancreas from premature trypsin activation and consequent autodigestion of the organ. BPTI is a compact 58-amino acid protein with three disulfide bonds; it binds to BT by inserting a binding loop into the trypsin active site in a substrate-like manner, with a Lys residue occupying the complementary trypsin specificity pocket.[51] The backbone of the BPTI binding loop is preconfigured in canonical conformation for lock-and-key recognition at the BT active site, forming main chain−main chain hydrogen bonds between inhibitor and enzyme that are structurally conserved across complexes with different proteases.[51,52] The structure of BPTI thus acts as a "molecular vise",[42] forcing the same Lys residue into the specificity pocket of most BT paralogs, irrespective of whether the residue forms favorable or locally deleterious interactions. Consequently, BPTI forms a structurally similar although weaker PPI with ChT, which possesses specificity for cleavage after large hydrophobic residues rather than Lys/Arg. BPTI likewise forms a similar PPI with MT, a trypsin paralog that is present only in hominids, having arisen from a relatively recent gene duplication.[53] MT retains the primary specificity of other trypsins for cleavage after Lys/Arg but has evolved unique resistance to inhibition by canonical trypsin inhibitors due to several mutations near the active site, resulting in much weaker affinity toward BPTI.[54−56] Consistent with the structural homology between these three PPIs, they exhibit binding interfaces of nearly identical physicochemical properties despite their large differences in K_D values (Figure 1).

In an attempt to better understand the drastic differences in binding affinities between these 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 strategy recently developed by our group that relies on protein randomization, yeast surface display (YSD) technology, NGS analysis, and a small experimental data set of

https://doi.org/10.1021/jacs.1c08707
J. Am. Chem. Soc. 2021, 143, 17261−17275

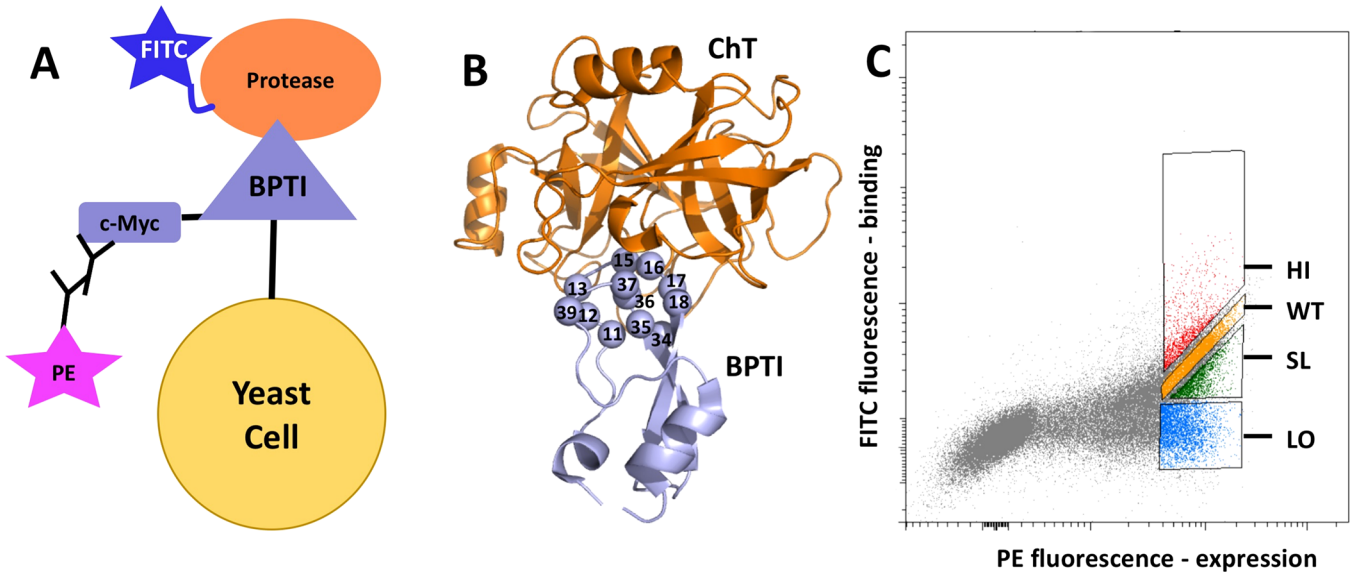


Figure 2. (A) Yeast surface display construct with BPTI displayed on the surface of yeast cells. C-Myc tag is used to monitor BPTI mutant expression using PE-labeled antibody. Proteases are labeled by FITC that is used to monitor binding between the two proteins. (B) Construction of the BPTI mutant library. Structure of the ChT/BPTI complex with ChT shown in orange and BPTI in violet. BPTI binding interface positions randomized to 20 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 are shown for the BPTI/ChT interaction, while similar data were obtained for the BPTI/MT and BPTI/BT interactions.

$\Delta\Delta G_{bind}$ values measured using purified proteins to generate $\Delta\Delta G_{bind}$ values for the remaining numerous mutants of the same protein complex.[57] We previously demonstrated that the above method produces a very high correlation ($R > 0.9$) between the NGS-based predictions of $\Delta\Delta G_{bind}$ values and the actual experimental values.[57] In the present study, we use this state-of-the-art methodology to construct and compare binding landscapes of the three structurally similar BPTI/protease complexes.

Our data demonstrate that the three complexes possess drastically different binding landscapes and lie at different points with 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.

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. Using this construct, $BPTI_{WT}$ was expressed on the surface of a yeast cell with a C-terminal myc-tag for monitoring protein expression through binding of an antimyc antibody and a secondary antibody conjugated to phycoerythrin (PE) (Figure 2A). Binding of a protease to $BPTI_{WT}$ 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 fluorescence-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 in the Supporting Information).

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 12 BPTI positions to 20 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 12 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 were covered in the naïve library. Among double mutants, we saw 89% of all possible sequences; this percentage was reduced to 60% when a cutoff of five sequencing reads was applied.

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). The concentration of each protease was optimized to exhibit a considerable spread of the FACS binding signals from different BPTI mutants (Figure S2 in the Supporting Information). 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 Figures S3−S5 in the Supporting Information). The cells from each affinity gate were grown and sequenced with NGS, resulting in 300−900 K 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

https://doi.org/10.1021/jacs.1c08707
J. Am. Chem. Soc. 2021, 143, 17261−17275

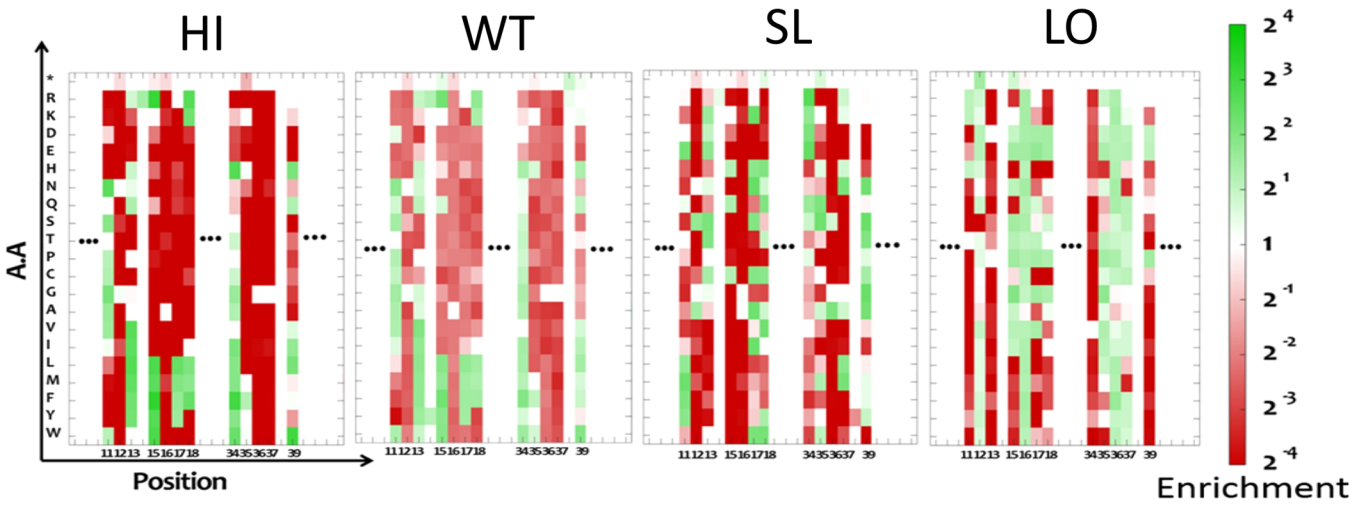


Figure 3. Heatmap showing the enrichment values for each single mutation in the ChT/BPTI complex in the four sorted affinity gates. The enrichment ratio varies from high (green) to low (red), as shown on the right axis. Similar maps were obtained for the other two PPIs.

heatmaps of the enrichment values for all positions as shown in Figure 3 for the CT/BPTI interaction.

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 data set of experimental $\Delta\Delta G_{bind}$ values measured by biophysical techniques on purified proteins.[57] We first compiled such normalization data sets for the three complexes, collecting 34 and 33 $\Delta\Delta G_{bind}$ data points from the literature for the ChT/BPTI and BT/BPTI interactions, respectively.[44,57−60] For the MT/BPTI interaction, where only a few $\Delta\Delta G_{bind}$ data points have been reported,[50,61] we produced the normalization data set by expressing and purifying 12 BPTI mutants and measuring their binding affinities to MT (Figure S6 in the Supporting Information). The above data sets 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 ($R \approx 0.9$; Figure S7 in the Supporting Information).

We next used these normalization formulas to predict $\Delta\Delta G_{bind}$ values for all single and double BPTI mutants that were detected by NGS for the three PPIs. While nearly all BPTI single mutants were 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. The relatively low coverage of double mutants is due to two main reasons. First, some BPTI mutants exhibited low folding stability, resulting in their low expression on yeast. Such destabilized mutants were not collected in our selection experiments for all three PPIs. Second, the number of NGS runs was not sufficient to detect all ∼26 000 double mutants in all affinity gates, thus resulting in some differences in invisible mutants for each protease. 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 gate. 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 (see Methods for details). 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 the ChT/BPTI interaction (47%), and 13 354 for the MT/BPTI interaction (51%). We thus constructed full single mutant binding landscapes (Figure 4) 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 three PPIs, we summarized our results in a histogram that includes $\Delta\Delta G_{bind}$ values from all 228 single mutations for each PPI (Figure 5A−C). While all 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 5A−C). On average, a single mutation destabilized a BT/BPTI interaction by 4.5 kcal/mol, a ChT/BPTI interaction by 1.6 kcal/mol, and an MT/BPTI interaction by 0.82 kcal/mol. 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 were absolute hot-spots in the BT/BPTI interaction, exhibiting only positive $\Delta\Delta G_{bind}$ values (Figure 4). In contrast, only four absolute hot-spots were 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

https://doi.org/10.1021/jacs.1c08707
J. Am. Chem. Soc. 2021, 143, 17261−17275

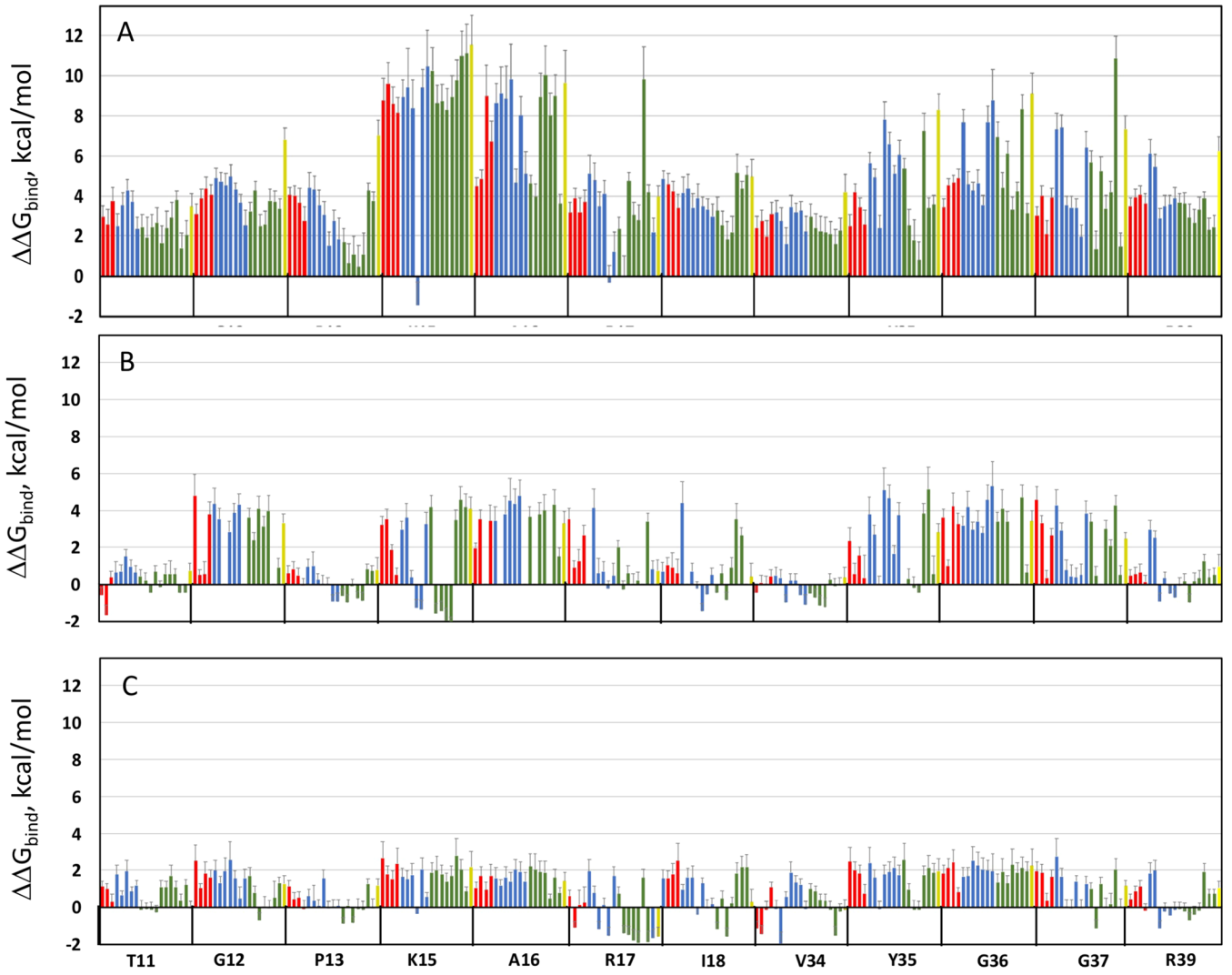


Figure 4. Changes in $\Delta\Delta G_{bind}$ for all single mutants of BPTI interacting with BT (A), ChT (B), and MT (C). Each bar represents a mutation to one amino acid including hydrophobic amino acids (green), polar amino acids (red), charged amino acids (blue), and Cys (yellow). The *x*-axis shows the WT residue followed by position. Error bars represent the 95% CI. Figure (A) in reprinted without changes from Heyne, M.; Papo, N.; Shifman, J. M. Generating quantitative binding landscapes through fractional binding selections combined with deep sequencing and data normalization. *Nat. Commun.* **2020**, *11* (1), 1−7, under the terms of the CC BY license (Creative Commons Attribution 4.0 International License), Copyright © 2020, The Author(s).

among the three complexes (Figure 6). Position 15 is nearly a hot-spot in both the BT/BPTI and the MT/BPTI complexes, with all mutations leading to high destabilization of the two complexes except for the K15R mutation, which leads to affinity improvement. However, the same position is a cold-spot in the ChT/BPTI complex, where all hydrophobic amino acids lead to affinity improvement. 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.[59] Additionally, the amino acid preferences of BPTI at position 15 observed 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 $\Delta\Delta G_{bind}$ values for ∼50% of all double mutations, for which $\Delta\Delta G_{bind}$ predictions were available (Figure 5D−F). 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 0.63 kcal/mol, showing the same tendency of increased destabilization due to double mutation with increasing affinity of the PPI as observed for single mutants. When comparing an average effect from a double and a single mutation, BT/BPTI and ChT/BPTI exhibit a higher mean $\Delta\Delta G_{bind}$ value for a double mutation, consistent with the interpretation that a majority of single mutations are destabilizing in these two PPIs. For the low-affinity MT/BPTI complex, the double mutant average is slightly lower compared to that of the single mutant average, consistent with the fact that in this PPI many affinity-enhancing mutations have been detected. All three double mutation distributions exhibit a lower mean $\Delta\Delta G_{bind}$ value in comparison to what would be predicted from additivity of all single mutations (Figure 5D−F). These results could be

https://doi.org/10.1021/jacs.1c08707
J. Am. Chem. Soc. 2021, 143, 17261−17275

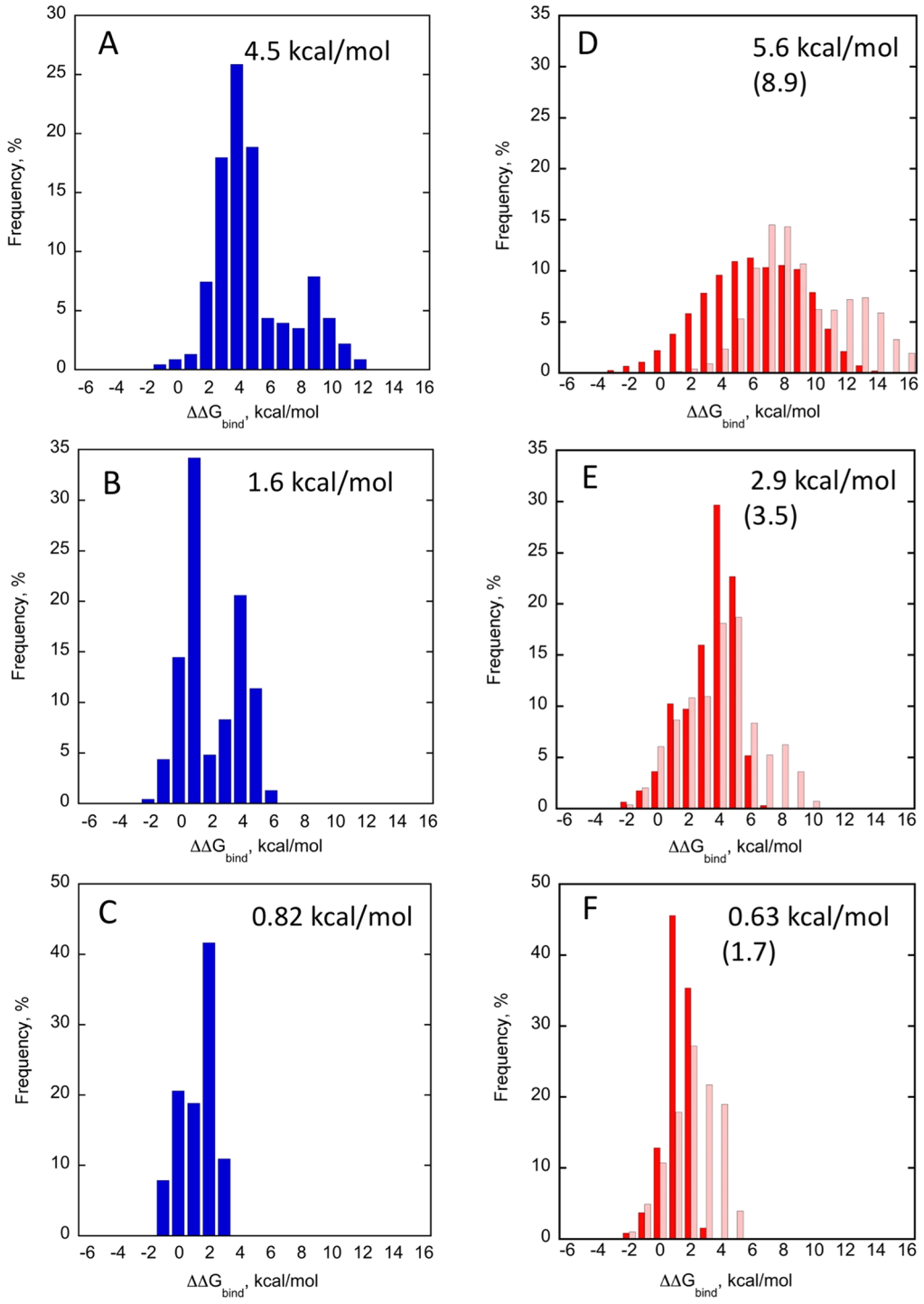


Figure 5. Histograms of $\Delta\Delta G_{bind}$ for single and double BPTI mutants. Single BPTI mutants interacting with (A) BT; (B) ChT; and (C) MT. Double BPTI mutants interacting with (D) BT; (E) ChT; and (F) MT. Dark red represents the actual measured values, while light red represents the values that would result from all single mutations being additive. Mean value for $\Delta\Delta G_{bind}$ for each histogram is displayed on top of each graph; in parentheses is the value that would result from all double mutations being additive. While all 228 single mutants are incorporated into the single mutant histograms for all proteases, only ~50% of double mutants are summarized. The data are available in the Source Data file in the Supporting Information.

explained by the fact that ~50% of the BPTI double mutations are absent from our analysis. Among the invisible mutations, the majority result in high BPTI destabilization and are likely to exhibit high positive $\Delta\Delta G_{bind}$ upon binding to proteases. Such mutations, if included in the analysis, would shift the double mutant distribution to higher mean $\Delta\Delta G_{bind}$ values for all three PPIs. The minority of the double mutants are missing due to their low coverage in the NGS data. These mutants, however, are randomly distributed in the $\Delta\Delta G_{bind}$ histogram and are not expected to change the mean.

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

https://doi.org/10.1021/jacs.1c08707
J. Am. Chem. Soc. 2021, 143, 17261−17275

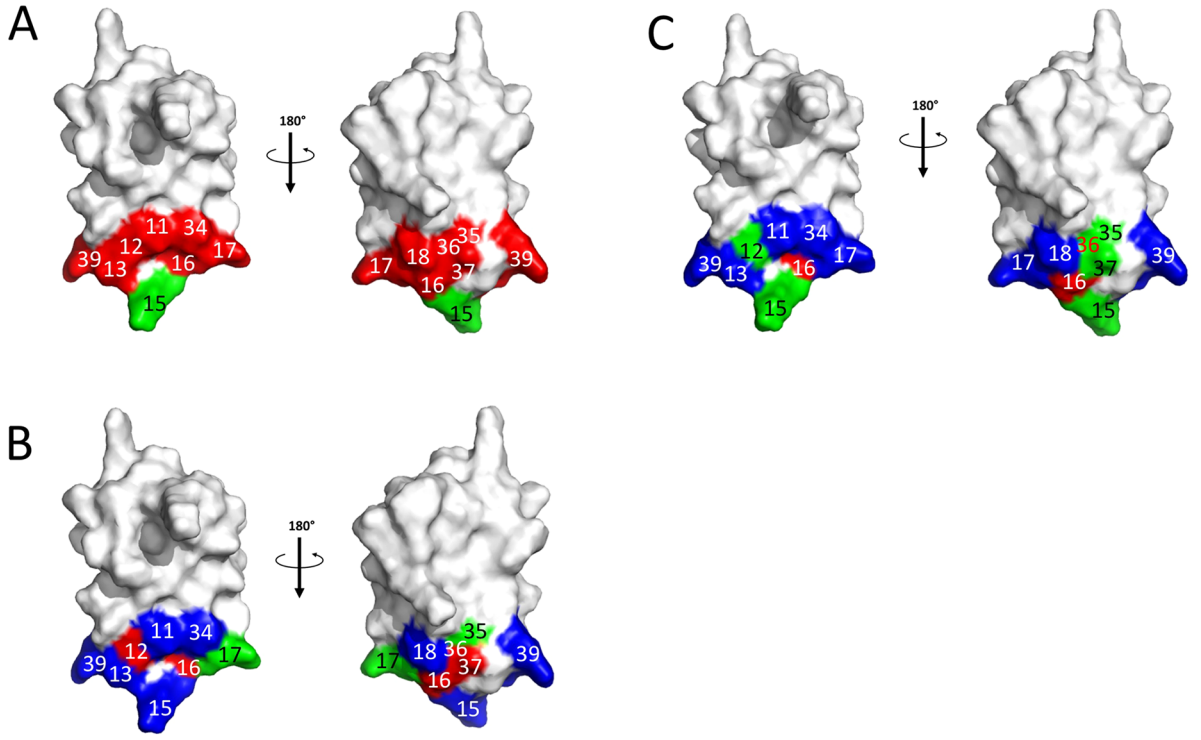


Figure 6. Structure of BPTI with binding interface positions colored according to the number of detected affinity-enhancing mutations at this position when interacting with (A) BT; (B) ChT; and (C) MT. Red: no affinity-enhancing mutation was detected; green: 1 or 2 affinity-enhancing mutations were detected; blue: 3 or more affinity-enhancing mutations were detected.

representative single BPTI mutants in the cognate BPTI/BT complex: BPTI_K15R, which 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, which 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 (Figure S8 in the Supporting Information). K15R, which improves the affinity 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 distribution that contains both stabilizing and destabilizing steps with magnitudes ranging from −6 to +8 kcal/mol. The highly destabilizing mutation K15A exhibits a distribution 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 because 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 nonadditivity of some of the single mutations in BPTI.

We next explored the robustness and evolvability of the BPTI sequence toward its main function, high-affinity binding to BT. We assumed that BPTI mutants that either stabilize the BPTI/BT complex or destabilize it by a small amount (1 kcal/mol or less) would be functional in the cellular environment. With this definition, only 2% of all single BPTI mutations were functional. The number of functional mutations was increased to ∼16% among the detected double mutations. However, among double mutations that involve the most energetically important position 15, only 5% would support high-affinity binding to BT. Both of the numbers for double mutations are likely an overestimation due to the absence of highly destabilized BPTI mutants from our data. We further compared the effect of the same BPTI mutations on its binding affinity to the three proteases by computing the correlation between $\Delta\Delta G_{bind}$ values for the same mutations between pairs of PPIs. Our analysis shows that the same single mutations frequently result in similar binding affinity changes for all three PPIs (Figure S9A in the Supporting Information; R values of ∼0.55 for all PPI pairs), consistent with high sequence homology of the studied proteases. As expected, this correlation goes down to 0.3−0.4 when all single and double mutations are considered (Figure S9B−D in the Supporting Information). Further analysis shows that the largest majority of mutations in BPTI leads to a simultaneous decrease in affinity to pairs of proteases (91%, 81%, and 80% for BT/ChT, BT/MT, and ChT/MT pairs, respectively). Mutations that decrease affinity to the cognate protease BT but increase affinity to the homologous protease are observed with intermediate frequency (4.8% and 14.5% for BT/ChT and BT/MT, respectively). Mutations that increase affinity to the

https://doi.org/10.1021/jacs.1c08707
J. Am. Chem. Soc. 2021, 143, 17261−17275

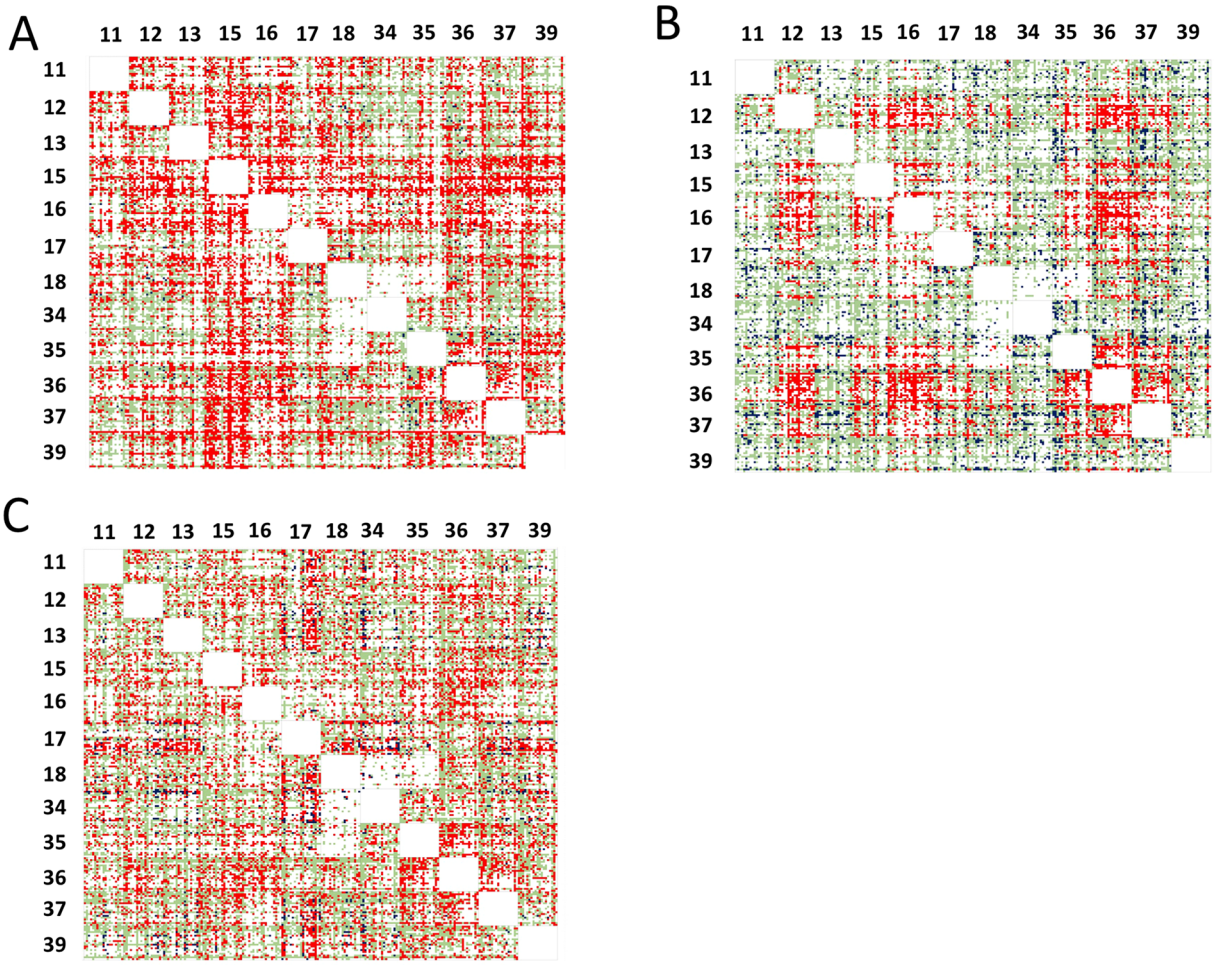


Figure 7. The matrix of ΔGi for double mutations in the three PPIs: (A) BT/BPTI; (B) ChT/BPTI; and (C) MT/BPTI. On the left and on the top are BPTI binding interface positions randomized to 20 amino acids. Each square is a 20 × 20 matrix containing ΔGi values for coupling between a particular mutation 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 are available.

cognate protease BT but decrease affinity to the homologous protease are rather rare (2.9% and 2.7% for BT/ChT and BT/MT, respectively). Very rare yet detectable are BPTI mutations that simultaneously increase affinity for two proteases (1%, 1.6%, and 2% for BT/ChT, BT/MT, and ChT/MT, respectively). These results agree with previous studies on enzymes where the native activity was found to be less robust to mutations than the promiscuous activity.[62] In addition, our results demonstrate the possibility of designing BPTI mutants that lead to an increase in binding specificity toward one particular protease.

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 and exhibiting positive and negative epistasis according to the magnitude of the coupling energy ΔG_i upon two mutations X and Y:

$$\Delta G_i = \Delta\Delta G^{X}_{bind} + \Delta\Delta G^{Y}_{bind} - \Delta\Delta G^{XY}_{bind} \quad (1)$$

Here, $\Delta\Delta G^{X}_{bind}$ and $\Delta\Delta G^{Y}_{bind}$ represent the change of the binding free energy of the single mutants X and Y, respectively, $\Delta\Delta G^{XY}_{bind}$ represents the change of binding free energy of the double mutant containing mutations X and Y. Negative epistasis was defined when $\Delta G_i < 0$ within the uncertainties of the $\Delta\Delta G_{bind}$ predictions for the double mutant and the two corresponding single mutants (eq 8), i.e., the double mutation exhibited lower affinity compared to what is expected from additivity of the two single mutations. Positive epistasis was defined when $\Delta G_i > 0$, within the uncertainties of the $\Delta\Delta G_{bind}$ predictions (eq 9), i.e., the double mutation exhibited higher affinity compared to what is expected from additivity of two single mutations. Two mutations were defined as additive if the mutations did not fall into either positive or negative epistasis groups, that is, $\Delta G_i = 0$ within the uncertainties of the $\Delta\Delta G_{bind}$ predictions. In such an analysis, we are likely overestimating the number of additive mutations as mutations with small epistasis and large uncertainties would be assigned into the "additive" group.

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 54% of mutations are additive, 42% exhibit positive epistasis, and 4% show negative epistasis. Note that among the invisible double mutations, the majority significantly destabi-

https://doi.org/10.1021/jacs.1c08707
J. Am. Chem. Soc. 2021, 143, 17261−17275

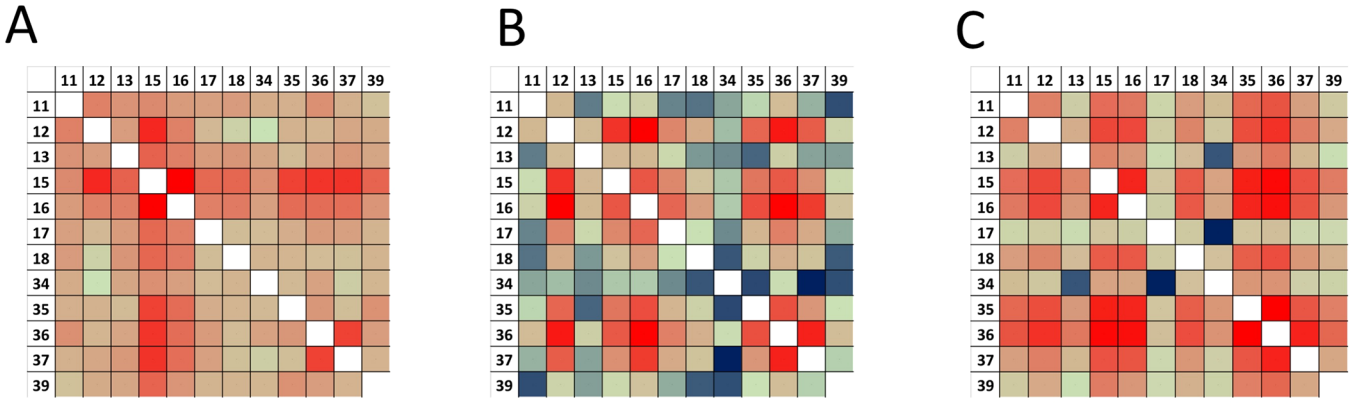


Figure 8. Coupling energies ΔGi averaged over all detected mutations at a pair of BPTI positions: (A) BT/BPTI interaction; (B) ChT/BPTI interaction; (C) MT/BPTI interaction. On the top and on the left BPTI positions are shown. The values are color coded from high positive epistasis (dark red) to additive (green) and to negative epistasis (dark blue).

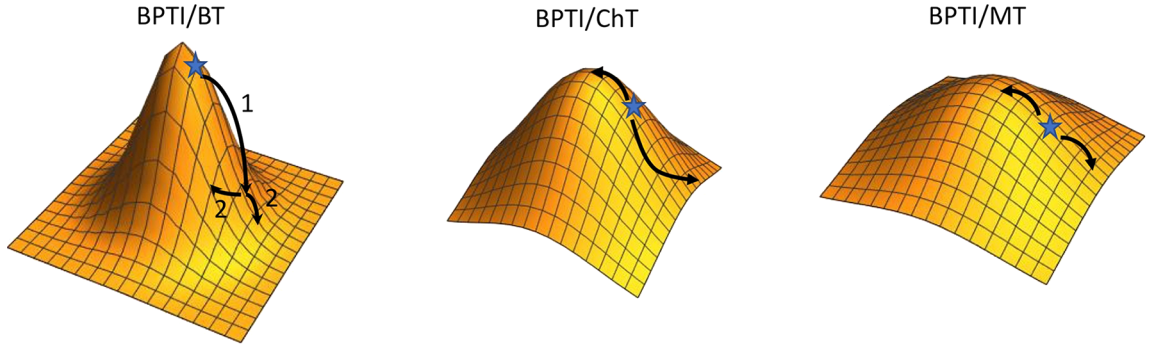


Figure 9. Schematic illustration of the single mutant binding landscapes for the three studied PPIs. The maximum on the surface corresponds to the highest possible binding affinity. A star indicates the position of the WT BPTI sequence with respect to the maximum. Arrows illustrate how single mutations lead to affinity changes in the three complexes.

lize BPTI folding. Such destabilizing mutations are likely to exhibit negative epistasis; thus, their absence from our data is consistent with a higher percentage of positive vs negative epistasis for all three complexes. We further constructed the per-position correlation matrices displaying coupling energy between all detected mutations in the three PPIs (Figure 7). Figure 7 shows that the sign of ΔG_i depends not only on a pair of positions but also on the mutation type, thus demonstrating multiple examples of sign epistasis.[63] Yet, a certain preference for either negative or positive epistasis frequently dominates coupling at certain position pairs, as some squares are predominantly red or blue in Figure 7. To analyze which positions exhibit a higher degree of coupling, we averaged ΔG_i values over all detected mutations at each pair of positions (Figure 8). Figure 8 shows that different proteases exhibit different patterns of coupling between pairs of mutations. Interestingly, positions where highest destabilization is observed for single mutations show a high degree of positive epistasis with all other positions (for example, positions 15 and 16 in the BPTI/BT complex, positions 12, 16, and 36 in the ChT/BPTI complex, and positions 15, 16, 35, and 36 in the MT/BPTI complex). On the other hand, cold-spot positions tend to exhibit negative epistasis with many other positions in the protein (see, for example, positions 13, 18, 34, and 39 in the ChT/BPTI complex and position 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, in agreement with the overall dynamic range of $\Delta\Delta G_{bind}$ values exhibited by the three complexes (Figure S10 in the Supporting Information). We further analyzed whether nonadditive mutations of the same $\Delta\Delta G_{bind}$ sign exhibited antagonistic or synergetic epistasis; that is, the effect of combination of the two mutations resulted in smaller (antagonistic) or larger (synergetic) change compared to additivity. We found that antagonistic epistasis predominated in all three PPIs with 95%, 78%, and 99% of mutations in the BT/BPTI, ChT/BPTI, and MT/BPTI complexes, respectively. When combining a stabilizing mutation with a destabilizing mutation, affinity improvement was recorded for 92%, 55%, and 23% of nonadditive double mutations for the BT/BPTI, MT/BPTI, and ChT/BPTI complexes, respectively.

We further tested whether the degree of coupling between the two mutations depends on the distance between the mutated positions (Figure S11 in the Supporting Information). Our results show that mutations at two closely located positions could exhibit various degrees of coupling from high to low. As the distance between positions increases, the average coupling between the two mutations decreases (Figure S11 in the Supporting Information). A similar trend has been observed for all three PPIs and is in agreement with previous studies in various biological systems.[64]

DISCUSSION

Link between PPI Binding Landscape and Its Function. In this study we measured quantitative effects of

https://doi.org/10.1021/jacs.1c08707
J. Am. Chem. Soc. 2021, 143, 17261−17275

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 ultrahigh to low. We find that the binding landscape of each PPI depends strongly on the interaction K_D. In particular, the ultrahigh affinity BT/BPTI complex is highly optimized in its sequence. Accordingly, the sequence of WT BPTI lies nearly at the maximum of the single mutant binding landscape (Figure 9), 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 (Figures 4A and 5A). Such high destabilizations from single mutations are extremely rare. For example, the SKEMPI database[65] 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. Functionally, this result means that BPTI cannot accept mutations at key positions without losing its main function, i.e., high-affinity binding to BT. The medium-affinity ChT/BPTI complex shows a lower degree of optimality, with a larger fraction of single 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 single mutant landscape of the ChT/BPTI complex exhibits a medium gradient and the WT sequence lies about two-thirds (6 /(6 + 2.6)) up the landscape hill (Figure 9). One might expect that the low-affinity MT/BPTI complex would exhibit a $\Delta\Delta G_{bind}$ distribution that is the opposite of that observed for the BPTI/BT complex, with a high number of mutations that lead to a very large improvement in binding affinity. Yet, this is not what we observed 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 single mutant landscapes themselves show different gradients, high for the high-affinity complex and low for the low-affinity complex (Figure 9).

The binding landscape characteristics of the three studied PPIs have been dictated by their evolutionary history. BPTI has coevolved with BT to optimize their affinity for each other, as a mechanism to protect the pancreas. Activation of trypsin is normally catalyzed by enteropeptidase in the duodenum and serves as a master regulatory step in digestive enzyme activation, since trypsin is the common activator of all other pancreatic zymogens.[66] Premature autoactivation of trypsin in the pancreas can cause a runaway activation cascade, leading to tissue damage, inflammation, and pancreatitis.[66] Thus, BPTI and other potent pancreatic trypsin inhibitors have evolved to regulate this key digestive enzyme. Humans do not possess an exact ortholog of BPTI, but the same biological protective function is filled by SPINK1,[66] another canonical trypsin inhibitor that has coevolved with human trypsin.

By contrast, ChT and MT are not susceptible to autoactivation and lack specific endogenous coevolved inhibitors in the mammalian pancreas. They bind to BPTI due to their sequence and structural homology to BT and due to the preconfiguration of the BPTI binding loop in canonical conformation with complementarity to conserved features of the serine protease active site.[51,52] These nonoptimized PPIs notably show weaker affinity than that exhibited by BPTI/BT. The interaction with MT is particularly weak, despite the similar specificity of BT and MT for cleavage after Lys/Arg, due to the evolution of MT for widespread natural resistance to the canonical trypsin inhibitors such as BPTI.[67,68] Furthermore, by binding to these inhibitors orders of magnitude more weakly than other trypsins, MT has evolved the capability to cleave many proteinaceous trypsin inhibitors as substrates.[54,56,69−71] The influence of this evolutionary adaption on the MT/BPTI interaction is consistent with our finding that no single or double mutation on the BPTI side could convert this complex into a high-affinity PPI. We anticipate that BPTI may yet be converted to a high-affinity inhibitor of MT but in this case requiring a greater number of combined mutations. Consistent with such a possibility, APPI, a human BPTI paralog with 44% sequence identity, binds to MT ∼100-fold more tightly ($K_D \approx 10^{-7}$ M),[50] while triple and quadruple mutants of APPI identified by directed evolution can bind to MT more than 1000-fold yet more tightly ($K_D < 10^{-10}$ M).[72,73]

Comparison of Cold-Spot and Hot-Spot Locations in the Three PPIs. Cold-spots are positions in proteins that are occupied by nonoptimal amino acids; at such positions, multiple mutations lead to functional improvement. In this study we analyzed cold-spot and hot-spot locations on BPTI and found several notable differences between the three studied PPIs. The most prominent difference is observed at position 15, which is nearly a hot-spot in the BT/BPTI and the MT/MPTI complexes but is a cold-spot in the ChT/BPTI complex. This central to the binding interface position is occupied by Lys in WT BPTI, which fits perfectly in the BT specificity pocket, where it participates in a network of hydrogen bond interactions with several polar residues on BT (Figure S12 in the Supporting Information). The K15 binding pocket is completely conserved in the MT/BPTI complex. In agreement with structural identity of the MT binding pocket, all mutations but mutation to Arg at position 15 result in destabilization of the MT/BPTI complex, yet mutational changes are smaller in comparison to those exhibited by the BT/BPTI interaction. On the contrary, in the ChT/BPTI interaction, mutations to all hydrophobic amino acids at position 15 lead to improved affinity. Here, slight differences in the ChT specificity pocket prevent the formation of the hydrogen bond network with K15 of BPTI; hence K15 exhibits a different, less buried conformation, avoiding unfavorable charge burial (Figure S12 in the Supporting Information).

Additional differences in cold-spot and hot-spot locations are observed at the periphery of the BPTI/protease binding interface, where they could be explained from the structural perspective. Affinity-enhancing mutations at cold-spots could occur either through removal of an unfavorable interaction with the partner protein or through introduction of a new favorable interaction, where no interaction exists.[49] We observed the first scenario occurring in the MT/BPTI complex at position 17, where an Arg on BPTI is found in close proximity to Arg 193 on MT, resulting in unfavorable charge

https://doi.org/10.1021/jacs.1c08707
J. Am. Chem. Soc. 2021, 143, 17261−17275

repulsion (Figure S13 in the Supporting Information). As our experimental data show, substituting Arg 17 with a small and/or hydrophobic amino acid results in affinity improvement (Figure 4C). On the contrary, Arg 17 is found in a largely hydrophobic environment in the ChT/BPTI complex; its replacement with a hydrophobic Met and Leu results in slightly negative $\Delta\Delta G_{bind}$ values (Figure 4B). A similar scenario is observed at position 39, which is a cold-spot in the MT/BPTI and ChT/BPTI complexes but a hot-spot in the BT/BPTI complex. Here, Arg 39 on WT BPTI is found in close proximity to Lys 175 on MT and ChT; Arg 39 substitution by a noncharged residue, such as Trp, results in affinity improvement in both complexes (Figure S14 in the Supporting Information). In the BT/BPTI complex, position 175 is occupied by Gln, which forms a hydrogen bond with Arg 39 on BPTI, explaining the hot-spot nature of this position in the BT/BPTI complex.

We observe the second scenario for cold-spot formation at position 34 in the ChT/BPTI and MT/BPTI complexes. Val 34 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 show that in all three BT/BPTI complexes a large proportion of double mutations results in positive epistasis and only a minority of mutations produces negative epistasis. We also observe the predominance of antagonistic epistasis vs synergetic epistasis. This absence of symmetry is at least partially due to the absence of highly destabilizing double mutations from our data. The abundance of positive epistasis in the BT/BPTI interaction could be explained from the perspective of binding landscape theory (Figure 9). 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. A 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 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 nonmutated positions. Yet, if any of the hot-spot residues is mutated (such as K15 in the BPTI/BT and BPTI/MT complex), 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 toward 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 or antagonistic 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 a benchmark for testing new computational methodology and for the design of new high-affinity protease inhibitors. Using the obtained data, we demonstrate striking differences between the single mutant 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.

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 sublibraries that each incorporates all possible pairs of the 12 randomized positions. The TPCR protocol[74] 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 the Supporting Information). These primers were used in a PCR together with the $BPTI_{WT}$ 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, USA) to remove any parental plasmid used as a template to construct the library, cleaned up with magnetic beads (AMPure XP, Beckman Coulter, Brea, CA, USA), and 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

https://doi.org/10.1021/jacs.1c08707
J. Am. Chem. Soc. 2021, 143, 17261−17275

QIAprep Spin miniprep (Qiagen, Hilden, Germany), and all the sublibraries were pooled together and balanced by their DNA amount to use the same amount of DNA from each sublibrary (~3.6 μg). 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 the amount of the library sample and counting the colonies after transformation on a SDCAA plate.

YSD Sorting Experiments. Yeast cells displaying the BPTI library or the $BPTI_{WT}$ with a cMyc-tag at the C-terminus on the YSD were grown in SDCAA selective medium and induced for BPTI protein expression with a galactose-containing SGCAA medium as previously described.[72] BPTI expression and binding to individual proteases were detected by incubating approximately 1×10^6 yeast cells with a 1:50 dilution of mouse anti-cMyc antibody (9E10, Abcam, Cambridge, UK) in 1× phosphate-buffered saline (PBS) supplemented with 1% bovine serum albumin (BSA, Thermo Fisher Scientific, Waltham, MA, USA) for 1 h at room temperature, washed with ice-cold 1× PBS, and then incubated with different concentrations of biotinylated BT (biotin and biotinylation protocol from Thermo Fisher Scientific) in 1× PBS with 1% BSA for 1 h at room temperature. Thereafter, cells were washed with ice-cold 1× PBS, followed by incubation with a 1:50 dilution of phycoerythrin (PE)-conjugated anti-mouse secondary antibody (Sigma-Aldrich, St. Louis, MO, USA) and NeutrAvidin (Thermo Fisher Scientific) conjugated with FITC in 1× PBS with 1% BSA for 20 min on ice. Finally, the cells were washed with ice-cold PBS, and the fluorescence intensity was analyzed by dual-color flow cytometry (Accuri C6, BD Biosciences). The yeast cells were next sorted into four populations by FACSAria (BD Biosciences, San Jose, CA, USA) including HI, WT, SL, and LO populations. Sorted cells were then grown in a selective medium, and the plasmidic DNA was extracted for each of the sorted population and the naïve library and submitted to NGS by MiSeq, Illumina (service provided by Hylabs, Rechovot, IL, USA).

NGS Analysis. The paired-end reads from the NGS experiments were merged,[75] 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 nonrandomized 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).

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

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

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 ref 76. Briefly, the original NGS data were 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 were calculated from 1000 resampling data sets for each BPTI mutant in each sorting gate and in the naïve library. The error was propagated into eqs 2 and 3 to calculate the error in enrichment values:

$$e^{i,j} = \frac{(f^{i,j})_{\text{naive}}}{(f^{i,j})_{\text{sorted}}} \times \sqrt{\left[\partial((f^{i,j})_{\text{sorted}})\frac{1}{(f^{i,j})_{\text{naive}}}\right]^2 + \left[-\partial((f^{i,j})_{\text{naive}})\frac{(f^{i,j})_{\text{sorted}}}{(f^{i,j})_{\text{naïve}}^2}\right]^2} \tag{4}$$

The error values in enrichment values were propagated to calculate the errors in $\Delta\Delta G_{\text{bind}}$ predictions.

Predicting $\Delta\Delta G_{\text{bind}}$ Values from NGS Data. All available experimental data on $\Delta\Delta G_{\text{bind}}$ for the BPTI/protease complexes were used to obtain the best normalization formulas for each complex for converting enrichment values from the four sorted populations into $\Delta\Delta G_{\text{bind}}$ values. To this end, we used a linear regression model function in Mathematica (Wolfram Research) with five 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, and f were optimized using the experimental data set as values of Y and the set of X_1, X_2, X_3, and X_4 enrichment values. The obtained normalization formula (different for each protease) was used to calculate $\Delta\Delta G_{\text{bind}}$ values for all the remaining single and double BPTI mutants that had four enrichment values recorded in the NGS experiment.

The standard deviation of $\Delta\Delta G_{\text{bind}}$ predictions for each BPTI mutant was calculated according to the formula

$$\sigma = \sqrt{(a\partial X_1)^2 + (b\partial X_2)^2 + (c\partial X_3)^2 + (d\partial X_4)^2 + (X_1\partial a)^2 + (X_2\partial b)^2 + (X_3\partial c)^2 + (X_4\partial d)^2 + \partial f^2} \tag{5}$$

Where a, b, c, d are the coefficients in front of X_1, X_2, X_3, and X_4, respectively; ∂X_1, ∂X_2, ∂X_3, and ∂X_4 are the standard deviations on these variables obtained from the bootstrapping analysis of the NGS data, and ∂a, ∂b, ∂c, ∂d, and ∂f are the standard deviations of these coefficients obtained from the leave-one-out analysis. To make $\Delta\Delta G_{\text{bind}}$ prediction for mutants where fewer than four 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 $\Delta\Delta G_{\text{bind}}$ 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 $\Delta\Delta G_{\text{bind}}$ 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 $\Delta\Delta G_{\text{bind}}$ 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 $\Delta\Delta G_{\text{bind}}$ values over all single mutations. For each of the mutants, we used the available enrichment values to make separate predictions from all possible "good" subsets of gates. First all predictions were recorded for mutants where enrichment values were available for all four gates. For mutants where predictions were available for only gates X_1, X_2, and X_4, predictions were made based on these three gates providing that this set of gates was defined as good. For mutants where predictions were available in only gates X_1 and X_4 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 by propagating an error from the enrichment values (see ref 57 for details). Finally, for each mutation, $\Delta\Delta G_{\text{bind}}$ prediction was selected from all the predictions according to the subset of gates where the highest correlation with experimental data was observed. The data set of final $\Delta\Delta G_{\text{bind}}$ predictions for all single and double mutants for the three PPIs could be found in the Source Data file.

Production of BPTI Variants. The $BPTI_{WT}$ sequence was cloned into a pPIC9K vector (Invitrogen, Carlsbad, CA, USA), and TPCR was used for site-directed mutagenesis to create the sequences of all variants, transformed, expressed in *P. pastoris* (*GS115* strain; obtained from Invitrogen) and purified by nickel affinity chromatography, followed by size-exclusion chromatography, as described in a previous work.[72] 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

https://doi.org/10.1021/jacs.1c08707
J. Am. Chem. Soc. 2021, 143, 17261−17275

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.[72] Here, 304 μL of BPTI (four different concentrations ranging from 5.2 to 52.6 μM) was mixed with 8 μL of the substrate Z-GPR-pNA (Sigma-Aldrich) (five different concentrations ranging from 0.4 to 10 mM). The mix was incubated for 3 min. Then, the reaction was initialized by adding 8 μL of MT (10 nM) and the absorbance of the samples was measured at a wavelength of 410 nm for 5 min. A negative control was added replacing BPTI with 304 μL of buffer (10 mM Tris, pH 8, 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 6, as described previously.[72]

$$v = \frac{K_{cat}[E]_0[S]}{K_m\left(1 + \frac{[I]}{K_i}\right) + [S]} \tag{6}$$

The change in experimental binding energy $\Delta\Delta G_{bind}$ was calculated from eq 7 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.

$$\Delta\Delta G_{bind} = RT \ln \frac{K_i(\text{mutant})}{K_i(\text{WT})} \tag{7}$$

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,

$$\Delta G_i = \Delta\Delta G^X_{bind} + \Delta\Delta G^Y_{bind} + \sqrt{\delta\Delta\Delta G^{X\ 2}_{bind} + \delta\Delta\Delta G^{Y\ 2}_{bind}} - \Delta\Delta G^{XY}_{bind} + \delta\Delta\Delta G^{XY}_{bind} < 0 \tag{8}$$

Here, $\delta\Delta\Delta G^X_{bind}$, $\delta\Delta\Delta G^Y_{bind}$, and $\delta\Delta\Delta G^{XY}_{bind}$ 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,

$$\Delta G_i = \Delta\Delta G^X_{bind} + \Delta\Delta G^Y_{bind} - \sqrt{\delta\Delta\Delta G^{X\ 2}_{bind} + \delta\Delta\Delta G^{Y\ 2}_{bind}} - \Delta\Delta G^{XY}_{bind} - \delta\Delta G^{XY}_{bind} > 0 \tag{9}$$

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

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.1c08707.

Source Data file: data for all $\Delta\Delta G_{bind}$ predictions for single and double BPTI mutants interacting with the three proteases (XLSX)

FACS data for experimental optimization and YSD selection experiments, MT/BPTI mutant binding affinity data, NGS normalization data for all PPIs, analysis of the binding landscapes for some single BPTI mutants, pairwise correlations between $\Delta\Delta G_{bind}$ values, absolute values of coupling energies, dependence of coupling on the Ca−Ca distance, structural modeling figures (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Julia M. Shifman − *Department of Biological Chemistry, The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem 9190401, Israel;* orcid.org/0000-0001-6678-6497; Email: jshifman@mail.huji.ac.il

Niv Papo − *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 8410501, Israel;* orcid.org/0000-0002-7056-2418; Email: papo@bgu.ac.il

Authors

Michael Heyne − *Department of Biological Chemistry, The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem 9190401, Israel; 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 8410501, Israel*

Jason Shirian − *Department of Biological Chemistry, The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem 9190401, Israel*

Itay Cohen − *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 8410501, Israel*

Yoav Peleg − *Life Sciences Core Facilities (LSCF) Structural Proteomics Unit (SPU), Weizmann Institute of Science, Rehovot 7610001, Israel*

Evette S. Radisky − *Department of Cancer Biology, Mayo Clinic Comprehensive Cancer Center, Jacksonville, Florida 32224, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.1c08707

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank U. Hadad (BGU) for help with FACS experiments. This work was supported by the Israel Science Foundation (ISF) grant 1873/15 and 2335/20 (J.M.S.), by the European Research Council (ERC) grant 336041, and by the ISF grant 1615/19 (N.P.). N.P. and E.S.R. acknowledge support from the US-Israel Binational Science Foundation (BSF). J.S.M. acknowledges support from the US-Israel Binational Science Foundation (BSF) and the Israel Cancer Research Foundation (ICRF). E.S.R. acknowledges support from U.S. National Institutes of Health grants R01 CA154387 and R01 GM132100.

■ REFERENCES

(1) Klesmith, J. R.; Bacik, J. P.; Wrenbeck, E. E.; Michalczyk, R.; Whitehead, T. A. Trade-offs between enzyme fitness and solubility illuminated by deep mutational scanning. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (9), 2265−2270.

(2) Mavor, D.; Barlow, K.; Thompson, S.; Barad, B. A.; Bonny, A. R.; Cario, C. L.; Gaskins, G.; Liu, Z.; Deming, L.; Axen, S. D.; Caceres, E.; Chen, W.; Cuesta, A.; Gate, R. E.; Green, E. M.; Hulce, K. R.; Ji, W.; Kenner, L. R.; Mensa, B.; Morinishi, L. S.; Moss, S. M.; Mravic, M.; Muir, R. K.; Niekamp, S.; Nnadi, C. I.; Palovcak, E.; Poss, E. M.;

Ross, T. D.; Salcedo, E. C.; See, S. K.; Subramaniam, M.; Wong, A. W.; Li, J.; Thorn, K. S.; Conchuir, S. O.; Roscoe, B. P.; Chow, E. D.; DeRisi, J. L.; Kortemme, T.; Bolon, D. N.; Fraser, J. S. Determination of ubiquitin fitness landscapes under different chemical stresses in a classroom setting. *eLife* **2016**, *5*, e15802.
(3) Fromer, M.; Shifman, J. M. Tradeoff between stability and multispecificity in the design of promiscuous proteins. *PLoS Comput. Biol.* **2009**, *5* (12), e1000627.
(4) Tokuriki, N.; Stricher, F.; Serrano, L.; Tawfik, D. S. How protein stability and new functions trade off. *PLoS Comput. Biol.* **2008**, *4* (2), e1000002.
(5) Taverna, D. M.; Goldstein, R. A. Why are proteins marginally stable? *Proteins: Struct., Funct., Genet.* **2002**, *46* (1), 105−9.
(6) Harms, M. J.; Thornton, J. W. Evolutionary biochemistry: revealing the historical and physical causes of protein properties. *Nat. Rev. Genet.* **2013**, *14* (8), 559−71.
(7) Bar-Even, A.; Noor, E.; Savir, Y.; Liebermeister, W.; Davidi, D.; Tawfik, D. S.; Milo, R. The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* **2011**, *50* (21), 4402−10.
(8) Cho, S.; Swaminathan, C. P.; Yang, J.; Kerzic, M. C.; Guan, R.; Kieke, M. C.; Kranz, D. M.; Mariuzza, R. A.; Sundberg, E. J. Structural basis of affinity maturation and intramolecular cooperativity in a protein-protein interaction. *Structure* **2005**, *13* (12), 1775−87.
(9) Bloom, J. D.; Romero, P. A.; Lu, Z.; Arnold, F. H. Neutral genetic drift can alter promiscuous protein functions, potentially aiding functional evolution. *Biol. Direct* **2007**, *2*, 17.
(10) Ernst, A.; Avvakumov, G.; Tong, J.; Fan, Y.; Zhao, Y.; Alberts, P.; Persaud, A.; Walker, J. R.; Neculai, A. M.; Neculai, D.; Vorobyov, A.; Garg, P.; Beatty, L.; Chan, P. K.; Juang, Y. C.; Landry, M. C.; 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, M. F.; Dhe-Paganon, S.; Sidhu, S. S. A strategy for modulation of enzymes in the ubiquitin system. *Science* **2013**, *339* (6119), 590−5.
(11) Goldsmith, M.; Tawfik, D. S. Directed enzyme evolution: beyond the low-hanging fruit. *Curr. Opin. Struct. Biol.* **2012**, *22* (4), 406−12.
(12) de Visser, J.; Elena, S. F.; Fragata, I.; Matuszewski, S. The utility of fitness landscapes and big data for predicting evolution. *Heredity* **2018**, *121* (5), 401−405.
(13) Whitehead, T. A.; Chevalier, A.; Song, Y.; Dreyfus, C.; Fleishman, S. J.; De Mattos, C.; Myers, C. A.; Kamisetty, H.; Blair, P.; Wilson, I. A.; Baker, D. Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat. Biotechnol.* **2012**, *30* (6), 543−8.
(14) Elazar, A.; Weinstein, J.; Biran, I.; Fridman, Y.; Bibi, E.; Fleishman, S. J. Mutational scanning reveals the determinants of protein insertion and association energetics in the plasma membrane. *eLife* **2016**, *5*, e12125.
(15) Sarkisyan, K. S.; Bolotin, D. A.; Meer, M. V.; Usmanova, D. R.; Mishin, A. S.; Sharonov, G. V.; Ivankov, D. N.; Bozhanova, N. G.; Baranov, M. S.; Soylemez, O.; Bogatyreva, N. S.; Vlasov, P. K.; Egorov, E. S.; Logacheva, M. D.; Kondrashov, A. S.; Chudakov, D. M.; Putintseva, E. V.; Mamedov, I. Z.; Tawfik, D. S.; Lukyanov, K. A.; Kondrashov, F. A. Local fitness landscape of the green fluorescent protein. *Nature* **2016**, *533* (7603), 397−401.
(16) Hietpas, R. T.; Jensen, J. D.; Bolon, D. N. Experimental illumination of a fitness landscape. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (19), 7896−901.
(17) Canale, A. S.; Cote-Hammarlof, P. A.; Flynn, J. M.; Bolon, D. N. Evolutionary mechanisms studied through protein fitness landscapes. *Curr. Opin. Struct. Biol.* **2018**, *48*, 141−148.
(18) Firnberg, E.; Labonte, J. W.; Gray, J. J.; Ostermeier, M. A comprehensive, high-resolution map of a gene's fitness landscape. *Mol. Biol. Evol.* **2014**, *31* (6), 1581−92.
(19) Stiffler, M. A.; Hekstra, D. R.; Ranganathan, R. Evolvability as a function of purifying selection in TEM-1 beta-lactamase. *Cell* **2015**, *160* (5), 882−892.
(20) Wrenbeck, E. E.; Azouz, L. R.; Whitehead, T. A. Single-mutation fitness landscapes for an enzyme on multiple substrates reveal specificity is globally encoded. *Nat. Commun.* **2017**, *8*, 15695.
(21) Wrenbeck, E. E.; Faber, M. S.; Whitehead, T. A. Deep sequencing methods for protein engineering and design. *Curr. Opin. Struct. Biol.* **2017**, *45*, 36−44.
(22) Melnikov, A.; Rogov, P.; Wang, L.; Gnirke, A.; Mikkelsen, T. S. Comprehensive mutational scanning of a kinase in vivo reveals substrate-dependent fitness landscapes. *Nucleic Acids Res.* **2014**, *42* (14), e112.
(23) Fowler, D. M.; Stephany, J. J.; Fields, S. Measuring the activity of protein variants on a large scale using deep mutational scanning. *Nat. Protoc.* **2014**, *9* (9), 2267−84.
(24) Fowler, D. M.; Araya, C. L.; Fleishman, S. J.; Kellogg, E. H.; Stephany, J. J.; Baker, D.; Fields, S. High-resolution mapping of protein sequence-function relationships. *Nat. Methods* **2010**, *7* (9), 741−6.
(25) Kowalsky, C. A.; Klesmith, J. R.; Stapleton, J. A.; Kelly, V.; Reichkitzer, N.; Whitehead, T. A. High-resolution sequence-function mapping of full-length proteins. *PLoS One* **2015**, *10* (3), e0118193.
(26) Yang, G.; Anderson, D. W.; Baier, F.; Dohmen, E.; Hong, N.; Carr, P. D.; Kamerlin, S. C. L.; Jackson, C. J.; Bornberg-Bauer, E.; Tokuriki, N. Higher-order epistasis shapes the fitness landscape of a xenobiotic-degrading enzyme. *Nat. Chem. Biol.* **2019**, *15* (11), 1120−1128.
(27) Chen, J. Z.; Fowler, D. M.; Tokuriki, N. Comprehensive exploration of the translocation, stability and substrate recognition requirements in VIM-2 lactamase. *eLife* **2020**, *9*, e56707.
(28) Kumar, H.; Shifman, J. M. Predicting mutational effects. In *In Protein Interactions: Computational Methods, Analysis and Applications*; Gromiha, M. M., Ed.; World Scientific Publishing: Singapore, 2020.
(29) Jubb, H. C.; Pandurangan, A. P.; Turner, M. A.; Ochoa-Montano, B.; Blundell, T. L.; Ascher, D. B. Mutations at protein-protein interfaces: Small changes over big surfaces have large impacts on human health. *Prog. Biophys. Mol. Biol.* **2017**, *128*, 3−13.
(30) Gonzalez, M. W.; Kann, M. G. Chapter 4: Protein interactions and disease. *PLoS Comput. Biol.* **2012**, *8* (12), e1002819.
(31) Pal, G.; Kouadio, J. L.; Artis, 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.* **2006**, *281* (31), 22378−85.
(32) Leung, I.; Dekel, A.; Shifman, J. M.; Sidhu, S. S. Saturation scanning of ubiquitin variants reveals a common hot spot for binding to USP2 and USP21. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (31), 8705−10.
(33) Aizner, Y.; Sharabi, O.; Shirian, J.; Dakwar, G.; Abvraham, O.; Risman, M.; Shifman, J. M. Mapping the binding landscape of a picomolar protein-protein complex through computation and experiment. *Structure* **2014**, *22*, 1−10.
(34) Kowalsky, C. A.; Whitehead, T. A. Determination of binding affinity upon mutation for type I dockerin-cohesin complexes from Clostridium thermocellum and Clostridium cellulolyticum using deep sequencing. *Proteins: Struct., Funct., Genet.* **2016**, *84* (12), 1914−1928.
(35) McLaughlin, R. N., Jr.; Poelwijk, F. J.; Raman, A.; Gosal, W. S.; Ranganathan, R. The spatial architecture of protein function and adaptation. *Nature* **2012**, *491* (7422), 138−42.
(36) Podgornaia, A. I.; Laub, M. T. Protein evolution. Pervasive degeneracy and epistasis in a protein-protein interface. *Science* **2015**, *347* (6222), 673−7.
(37) Moal, I. H.; Fernandez-Recio, J. SKEMPI: a Structural Kinetic and Energetic database of Mutant Protein Interactions and its use in empirical models. *Bioinformatics* **2012**, *28* (20), 2600−7.
(38) Ashkenazi, A.; Presta, L. G.; Marsters, S. A.; Camerato, T. R.; Rosenthal, K. A.; Fendly, B. M.; Capon, D. J. Mapping the CD4 binding site for human immunodeficiency virus by alanine-scanning mutagenesis. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87* (18), 7150−4.
(39) Cunningham, B. C.; Wells, J. A. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science* **1989**, *244* (4908), 1081−5.

https://doi.org/10.1021/jacs.1c08707
J. Am. Chem. Soc. 2021, 143, 17261−17275

(40) 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. U. S. A.* **2000**, *97* (16), 8950−4.
(41) Wang, C. Y.; Chang, P. M.; Ary, M. L.; Allen, B. D.; Chica, R. A.; Mayo, S. L.; Olafson, B. D. ProtaBank: A repository for protein design and engineering data. *Protein Sci.* **2018**, *27* (6), 1113−1124.
(42) Lu, S. M.; Lu, W.; Qasim, M. A.; Anderson, S.; Apostol, I.; Ardelt, W.; Bigler, T.; Chiang, Y. W.; Cook, J.; James, M. N.; Kato, I.; Kelly, C.; Kohr, W.; Komiyama, T.; Lin, T. Y.; Ogawa, M.; Otlewski, J.; Park, S. J.; Qasim, S.; Ranjbar, M.; Tashiro, M.; Warne, N.; Whatley, H.; Wieczorek, A.; Wieczorek, M.; Wilusz, T.; Wynn, R.; Zhang, W.; Laskowski, M., Jr. Predicting the reactivity of proteins from their sequence alone: Kazal family of protein inhibitors of serine proteinases. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (4), 1410−5.
(43) Meenan, N. A.; Sharma, A.; Fleishman, S. J.; Macdonald, C. J.; Morel, B.; Boetzel, R.; Moore, G. R.; Baker, D.; Kleanthous, C. The structural and energetic basis for high selectivity in a high-affinity protein-protein interaction. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (22), 10080−5.
(44) Castro, M. J.; Anderson, S. 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* **1996**, *35* (35), 11435−46.
(45) Sharabi, O.; Shirian, J.; Grossman, M.; Sagi, I.; Shifman, J. M. Affinity- and specificity-enhancing mutations are frequent in mulstispecific interaction between MMP14 and its inhibitor TIMP2. *PLoS One* **2014**, *9* (4), e93712.
(46) DeLano, W. L. Unraveling hot spots in binding interfaces: progress and challenges. *Curr. Opin. Struct. Biol.* **2002**, *12* (1), 14−20.
(47) Clackson, T.; Wells, J. A. A hot spot of binding energy in a hormone-receptor interface. *Science* **1995**, *267* (5196), 383−6.
(48) Bogan, A. A.; Thorn, K. S. Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **1998**, *280* (1), 1−9.
(49) Shirian, J.; Sharabi, O.; Shifman, J. M. Cold-spots in Protein Binding. *Trends Biochem. Sci.* **2016**, *41* (9), 739−45.
(50) Salameh, M. A.; Soares, A. S.; Navaneetham, D.; Sinha, D.; Walsh, P. N.; Radisky, E. S. Determinants of affinity and proteolytic stability in interactions of Kunitz family protease inhibitors with mesotrypsin. *J. Biol. Chem.* **2010**, *285* (47), 36884−96.
(51) Ascenzi, P.; Bocedi, A.; Bolognesi, M.; Spallarossa, A.; Coletta, M.; De Cristofaro, R.; Menegatti, E. The bovine basic pancreatic trypsin inhibitor (Kunitz inhibitor): a milestone protein. *Curr. Protein Pept. Sci.* **2003**, *4* (3), 231−51.
(52) Laskowski, M.; Qasim, M. A. What can the structures of enzyme-inhibitor complexes tell us about the structures of enzyme substrate complexes? *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **2000**, *1477* (1−2), 324−37.
(53) Rowen, L.; Williams, E.; Glusman, G.; Linardopoulou, E.; Friedman, C.; Ahearn, M. E.; Seto, J.; Boysen, C.; Qin, S.; Wang, K.; Kaur, A.; Bloom, S.; Hood, L.; Trask, B. J. Interchromosomal segmental duplications explain the unusual structure of PRSS3, the gene for an inhibitor-resistant trypsinogen. *Mol. Biol. Evol.* **2005**, *22* (8), 1712−20.
(54) Salameh, M. A.; Soares, A. S.; Hockla, A.; Radisky, E. S. Structural basis for accelerated cleavage of bovine pancreatic trypsin inhibitor (BPTI) by human mesotrypsin. *J. Biol. Chem.* **2008**, *283* (7), 4115−23.
(55) 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.* **2012**, *21* (8), 1103−12.
(56) Alloy, A. P.; Kayode, O.; Wang, R.; Hockla, A.; Soares, A. S.; Radisky, E. S. Mesotrypsin Has Evolved Four Unique Residues to Cleave Trypsin Inhibitors as Substrates. *J. Biol. Chem.* **2015**, *290* (35), 21523−35.
(57) Heyne, M.; Papo, N.; Shifman, J. M. Generating quantitative binding landscapes through fractional binding selections combined with deep sequencing and data normalization. *Nat. Commun.* **2020**, *11* (1), 1−7.
(58) Krowarsch, D.; Dadlez, M.; Buczek, O.; Krokoszynska, I.; Smalas, A. O.; Otlewski, J. Interscaffolding additivity: binding of P1 variants of bovine pancreatic trypsin inhibitor to four serine proteases. *J. Mol. Biol.* **1999**, *289* (1), 175−86.
(59) Otlewski, J.; Jaskolski, M.; Buczek, O.; Cierpicki, T.; Czapinska, H.; Krowarsch, D.; Smalas, A. O.; Stachowiak, D.; Szpineta, A.; Dadlez, M. Structure-function relationship of serine protease-protein inhibitor interaction. *Acta Biochim Pol* **2001**, *48* (2), 419−28.
(60) Beeser, S. A.; Goldenberg, D. P.; Oas, T. G. Enhanced protein flexibility caused by a destabilizing amino acid replacement in BPTI. *J. Mol. Biol.* **1997**, *269* (1), 154−64.
(61) Salameh, M. A.; Soares, A. S.; Hockla, A.; Radisky, D. C.; Radisky, E. S. The P(2)′ residue is a key determinant of mesotrypsin specificity: engineering a high-affinity inhibitor with anticancer activity. *Biochem. J.* **2011**, *440* (1), 95−105.
(62) Kaltenbach, M.; Emond, S.; Hollfelder, F.; Tokuriki, N. Functional Trade-Offs in Promiscuous Enzymes Cannot Be Explained by Intrinsic Mutational Robustness of the Native Activity. *PLoS Genet.* **2016**, *12* (10), e1006305.
(63) Weinreich, D. M.; Watson, R. A.; Chao, L. Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* **2005**, *59* (6), 1165−74.
(64) Jemimah, S.; Gromiha, M. M. Exploring additivity effects of double mutations on the binding affinity of protein-protein complexes. *Proteins: Struct., Funct., Genet.* **2018**, *86* (5), 536−547.
(65) Jankauskaite, J.; Jimenez-Garcia, B.; Dapkunas, J.; Fernandez-Recio, J.; Moal, I. H. SKEMPI 2.0: an updated benchmark of changes in protein-protein binding energy, kinetics and thermodynamics upon mutation. *Bioinformatics* **2019**, *35* (3), 462−469.
(66) Mayerle, J.; Sendler, M.; Hegyi, E.; Beyer, G.; Lerch, M. M.; Sahin-Toth, M. Genetics, Cell Biology, and Pathophysiology of Pancreatitis. *Gastroenterology* **2019**, *156* (7), 1951−1968.
(67) Nyaruhucha, C. N.; Kito, M.; Fukuoka, S. I. Identification and expression of the cDNA-encoding human mesotrypsin(ogen), an isoform of trypsin with inhibitor resistance. *J. Biol. Chem.* **1997**, *272* (16), 10573−8.
(68) 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.* **2002**, *315* (5), 1209−18.
(69) Salameh, M. A.; Robinson, J. L.; Navaneetham, D.; Sinha, D.; Madden, B. J.; Walsh, P. N.; Radisky, E. S. The amyloid precursor protein/protease nexin 2 Kunitz inhibitor domain is a highly specific substrate of mesotrypsin. *J. Biol. Chem.* **2010**, *285* (3), 1939−49.
(70) Szmola, R.; Kukor, Z.; Sahin-Toth, M. Human mesotrypsin is a unique digestive protease specialized for the degradation of trypsin inhibitors. *J. Biol. Chem.* **2003**, *278* (49), 48580−9.
(71) Pendlebury, D.; Wang, R.; Henin, R. D.; Hockla, A.; Soares, A. S.; Madden, B. J.; Kazanov, M. D.; Radisky, E. S. Sequence and conformational specificity in substrate recognition: several human Kunitz protease inhibitor domains are specific substrates of mesotrypsin. *J. Biol. Chem.* **2014**, *289* (47), 32783−97.
(72) Cohen, I.; Kayode, O.; Hockla, A.; Sankaran, B.; Radisky, D. C.; Radisky, E. S.; Papo, N. Combinatorial protein engineering of proteolytically resistant mesotrypsin inhibitors as candidates for cancer therapy. *Biochem. J.* **2016**, *473* (10), 1329−41.
(73) Cohen, I.; Naftaly, S.; Ben-Zeev, E.; Hockla, A.; Radisky, E. S.; Papo, N. Pre-equilibrium competitive library screening for tuning inhibitor association rate and specificity toward serine proteases. *Biochem. J.* **2018**, *475* (7), 1335−1352.
(74) Erijman, A.; Dantes, A.; Bernheim, R.; Shifman, J. M.; Peleg, Y. Transfer-PCR (TPCR): a highway for DNA cloning and protein engineering. *J. Struct. Biol.* **2011**, *175* (2), 171−7.
(75) Magoc, T.; Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **2011**, *27* (21), 2957−63.
(76) Kulesa, A.; Krzywinski, M.; Blainey, P.; Altman, N. Sampling distributions and the bootstrap. *Nat. Methods* **2015**, *12* (6), 477−8.

https://doi.org/10.1021/jacs.1c08707
J. Am. Chem. Soc. 2021, 143, 17261−17275