Generating quantitative binding landscapes through fractional binding selections, deep sequencing and data normalization

Michael Heyne[1,2], Niv Papo[2*], Julia Shifman[1*]

[1]Department of Biological Chemistry, Hebrew University of Jerusalem, Givat Ram Campus, 91906 Jerusalem, Israel

[2]Avram and Stella Goldstein-Goren Department of Biotechnology Engineering and the National Institute of Biotechnology, Ben-Gurion University of the Negev, P.O.B. 653, 8410501 Beer-Sheva, Israel

[*]Corresponding Authors: jshifman@mail.huji.ac.il; papo@bgu.ac.il.

1

Abstract: Quantifying the effects of various mutations on binding free energy is crucial for understanding the evolution of protein-protein interactions and would greatly facilitate protein engineering studies. Yet, measuring changes in binding free energy ($\Delta\Delta G_{bind}$) remains a tedious task that requires expression of each mutant, its purification, and affinity measurements. We developed a new approach that allows us to quantify $\Delta\Delta G_{bind}$ for thousands of protein mutants in one experiment. Our protocol combines protein randomization, Yeast Surface Display technology, Next Generation Sequencing, and a few experimental $\Delta\Delta G_{bind}$ data points on purified proteins to generate $\Delta\Delta G_{bind}$ values for the remaining numerous mutants of the same protein complex. Using this methodology, we comprehensively map the single-mutant binding landscape of one of the highest-affinity interaction between BPTI and Bovine Trypsin. We show that $\Delta\Delta G_{bind}$ for this interaction could be quantified with high accuracy over the range of 12 kcal/mol displayed by various BPTI single mutants.

Introduction

Protein-protein interactions (PPIs) control virtually all processes in the cell. Mutations at PPI binding interfaces frequently affect free energy of binding ($\Delta\Delta G_{bind}$), sometimes abrogating and sometimes stabilizing the interaction. This change in binding affinity of one PPI could translate into remodeling of the whole PPI network, frequently leading to dysregulation of signal transduction pathways and disease[1,2]. Therefore, understanding how specific mutations in protein complexes affect their binding affinity is extremely important to both basic biology and to biomedical sciences, where inhibition of a particular PPI might help to treat the related disease.

In the recent years, many groups reported computational methods for predicting $\Delta\Delta G_{bind}$ from structure and/or sequence [3-9]. While achieving good predictions on average, these methods frequently give large errors in particular cases, revealing that our comprehension of the precise molecular forces that govern binding affinity in PPIs remains incomplete[10]. Our knowledge in this area could be greatly expanded by acquiring large sets of data for $\Delta\Delta G_{bind}$ values in various PPIs, facilitating progress in computational modeling. Yet, experiments that determine $\Delta\Delta G_{bind}$ remain laborious since they involve DNA manipulation, protein expression and purification in different organisms and binding affinity measurements using different techniques. Thus, experimental data describing mutational effects on binding affinity for each particular PPI remain sparse and sometimes inconsistent between different reported experiments[11]. Furthermore, the majority of mutations reported in the literature are mutations to alanine[11-14]. Such mutations are of limited use for studies of protein evolution and protein engineering since they most frequently lead to complex destabilization and are rarely observed in nature.

A much more attractive and informative approach would be to explore all possible mutational effects for a particular PPI in a single experiment, thus generating a comprehensive binding landscape for this PPI[15,16]. Such binding landscapes could be used to define evolutionary paths accessible to a particular PPI, to characterize energetic contribution of each position, and to locate frequently sought affinity- and specificity-enhancing mutations[4,17]. First efforts in this direction utilized phage display technology that allows to select binders from a large combinatorial library of protein mutants[16,18,19]. Through several rounds of selection, protein mutants compatible with binding to a particular target are selected. Subsequent sequencing of

multiple selected clones allows us to calculate the frequency of each amino acid at each position, providing information on binding hot-spots[20] and cold-spots[21]. Further studies in this direction utilized Yeast Surface Display (YSD)[22] for selecting protein binders coupled to Next-Generation sequencing (NGS) to produce binding landscapes for various PPIs[23]. While YSD enables fast sorting using Fluorescently Activated Cell Sorting (FACS), NGS permits more accurate calculation of amino acid frequencies for each of the detected mutants. The ratio between the amino acid frequency in the selected pool of binders and the same frequency in the initial naïve library, referred to as the enrichment value, is calculated for each amino acid at each of the explored position. The enrichment values are then plotted to produce PPI binding landscapes. In such an approach, Whitehead *et al.* mapped the full single-mutant binding landscape of a *de novo* designed inhibitor interacting with hemagglutinin and used the enrichment information to design affinity-matured inhibitors[24].

In spite of great promise of this approach, further studies on different biological systems revealed its potential limitations. While affinity enhancing mutations could be readily identified by this methodology, relatively low correlation (R value of 0.5) between the NGS-derived enrichment values and experimental $\Delta\Delta G_{bind}$ values for purified proteins was observed[17]. Additional studies showed that $\Delta\Delta G_{bind}$ could be inferred from the NGS-based enrichment values only in the narrow range of energies from -0.8 to +0.5 kcal/mol[25,26], preventing construction of quantitative binding landscapes for all of the explored mutations with broader range of target affinities. In addition, the above methodology set a requirement on the concentration of the target protein in the selection experiment; the concentration should be similar to the interaction K_d, thus limiting the application of the approach to only subset of all PPIs with medium affinities. For high-affinity PPIs (Kd < 10^{-10} M), this condition would imply the usage of very low target protein concentration, resulting in undetectable fluorescence signal by FACS. For low-affinity PPIs (Kd > 10^{-5} M), high concentrations of protein would be necessary, making such experiments impractical.

We introduce a novel approach that allows us to overcome the abovementioned limitations and to generate quantitative binding landscapes for any PPI, regardless of their K_d value. Here, we demonstrate the applicability of our approach to a particularly difficult target, a

complex between Bovine Trypsin (BT) and its inhibitor BPTI that possesses ultra-high affinity of 10^{-14} M. We show that through our high-throughput NGS-based approach, we can obtain $\Delta\Delta G_{bind}$ values for all BPTI binding interface mutants that correlate extremely well with experimental results on purified proteins over the range of more than 12 kcal/mol free energy changes. Our method allowed us to comprehensively map the binding landscape for this ultra-high affinity interaction, which would be impossible using any alternative technique.

Results and Discussion

To demonstrate how our approach could be used to produce quantitative binding landscapes, we first prepared the BPTI/BT complex for YSD experiments. For this purpose, the BPTI$_{WT}$ gene was incorporated into the pCTCON vector, that facilitates BPTI expression on the surface of yeast cells with a C-terminal myc-tag (cMyc) for monitoring protein expression (**Figure 1A**). Binding of BT to BPTI mutants was accessed by monitoring fluorescence of the FITC fluorophore conjugated to a biotinylated BT via NeutrAvidin. The assessment of binding of BPTI$_{WT}$ to BT by FACS showed a diagonal narrow distribution, demonstrating that BPTI is well expressed on the surface of yeast cells, is properly folded, and binds to BT (**Figure S1**).
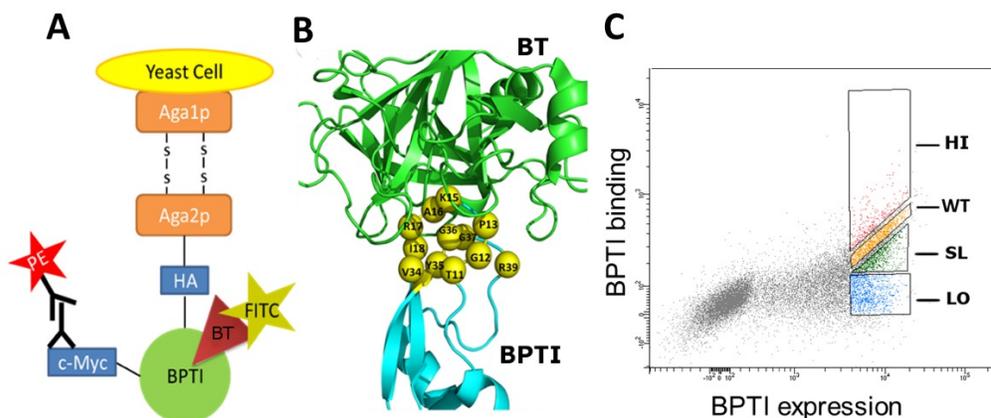

Figure 1: A) Yeast surface display construct with BPTI displayed on the surface of yeast cells B) Construction of the BPTI single mutant library. Structure of the BT/BPTI complex as shown from PDB 3OTJ. BT is colored in green, BPTI is colored in cyan and the BPTI binding interface positions are shown as spheres, labelled and colored in yellow. (C) FACS data showing sorting of four different populations of the BT/BPTI complex. The PE signal monitoring BPTI expression is shown on the x-axis and the FITC signal monitoring binding between fluorescently-labelled BT and BPTI variants expressed on the yeast surface is shown on the y-axis. The uppermost sorting gate HI represents all mutants with affinity higher than WT. The second uppermost gate WP represents all mutants with an affinity similar to BPTI$_{WT}$. The third gate SL represents all mutants with an affinity slightly lower than WT and the undermost gate LO represents all mutants with an affinity much lower than WT. BT concentration was 5 nM in this experiment.

Next, a combinatorial library was generated containing all single BPTI mutants at positions that are in the direct binding interface with BT. Thus, twelve BPTI positions were randomized to all twenty amino acids with an NNS codon, leaving intact two cysteines (C14 and C38) that form a disulfide bond and thus are crucial for BPTI folding (**Figure 1B).** The library of 228 (19 x 12) BPTI single mutants was constructed using the TPCR protocol[27]. The BPTI mutant library was expressed on the surface of yeast cells and incubated with a fluorescently labeled BT at concentration of 5 nM. This concentration of BT, although five orders of magnitude higher than the K_d of BT/BPTI interaction, was chosen since it was the minimum concentration of BT that resulted in a considerable spread of FACS binding signals from different BPTI mutants to BT (**Figure 1C**).

Improving accuracy and extending prediction range by collecting more data   One of the limitations of previous approaches for binding landscape generation was that $\Delta\Delta G_{bind}$ showed linear dependence on the NGS-enrichment value only in the narrow range of $\Delta\Delta G_{bind}$ values close to zero[26]. The methodology could not previously discriminate between different highly destabilizing mutations since such mutations were characterized with the same enrichment values. The same was true for mutations that showed high stabilization of the complex. To overcome this limitation and to increase the range of sensitivity for $\Delta\Delta G_{bind}$ predictions, we introduced our first innovation and used multiple affinity gates from which the mutants were collected during the YSD selection experiment. The multiple gates would allow us to collect information for each mutant several times, and each particular mutant would be enriched in at least one affinity gate. In this particular work, we used four affinity gates for mutant collection: higher than WT affinity (HI), WT-like affinity (WT), slightly lower than WT affinity (SL), and strongly lower than WT affinity (LO) (**Figure 1C**). The WT affinity gate was set according to the FACS signal produced by BPTI$_{WT}$ binding to BT at same conditions (**Figure 1S**). The cells from each gate were then grown, analyzed for binding to BT (**Figure S4**) and sequenced with NGS, resulting in 300-900K reads per each population. In addition, the naïve pre-sorted library of BPTI mutants was sequenced.

We further assessed the quality of the NGS data using synonymous mutations as a test. Since some errors in the data could come from errors in the NGS process, especially for sequences

detected with low frequency, we tested different cut-off values below which the data on the BPTI mutant would be discarded. Using different cut-off values, we calculated deviations in enrichment values for synonymous mutations expressing the same BPTI variant. Our data shows that at the cutoff value of 100 sequences per BPTI mutant, deviations in enrichment values were negligible (< 0.001) (**Figure S2**). Using this threshold, we were able to detect all 228 BPTI single mutants present in the naïve library. No threshold was applied to the sorted populations, since in such population the low number of sequences was caused by the depletion of that mutant from the population.
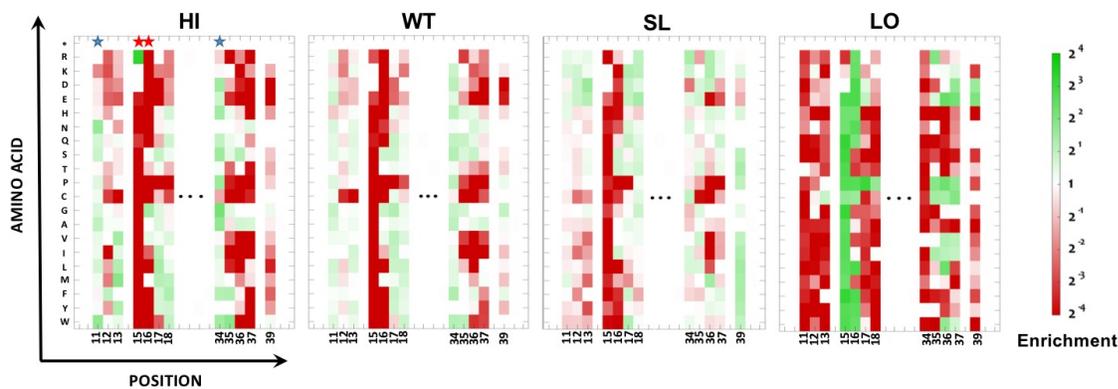
Figure 2: Experimental Binding landscapes for the BT/BPTI complex. These heat maps show the enrichment of each variant in each of the four sorted affinity gates relative to the pre-sort. The enrichment ratio varies from high (green) to low (red) as shown on the right axis. Red and blue stars on top show binding hot-spots and cold spots, respectively. Similar maps were obtained for all four PPIs.

We thus had in our hands four enrichment values from four affinity gates for each of the 228 BPTI mutants (**Figure 2**). Closer examination of the data showed that enrichment values in HI and LO affinity gates exhibited pseudo-symmetry, with highly enriched mutations in the HI gate being highly depleted in the LO gate and vice versa. The enrichment value maps could be used to define binding hot-spots for the BPTI/BT interactions (such as position 15, 16 indicated as red starts on top of **Figure 2**) and more tolerant to mutations positions (such as 11 and 34 indicated as blue stars on **Figure 2**). However, these maps were not sufficient in determining exact $\Delta\Delta G_{bind}$ values for each of the mutation. In fact, we noticed that some mutations that showed enrichment value of ~1 in the HI gate, that should correspond to the WT-like affinity, were determined to be destabilizing when measured with purified proteins (for example, G12A with experimental $\Delta\Delta G_{bind}$ of +4.35 kcal/mol[28-31]). This over-prediction of neutral and affinity-enhancing mutations

by our NGS results was due to the fact that in the YSD selection experiment we used much higher concentration of BT compared to the K_d of BPTI/BT interaction, shifting the equilibrium towards protein binding even for those BPTI mutants that possessed weaker affinities compared to the WT protein. To overcome this problem, we introduced the second innovation, described below.

Normalizing NGS data to get quantitative $\Delta\Delta G_{bind}$ measurements   Our   second innovation was to obtain a normalization formula for converting the enrichment data from four affinity gates into one $\Delta\Delta G_{bind}$ value. For this purpose, we collected all available $\Delta\Delta G_{bind}$ experimental data for binding of BPTI single mutants to BT, comprising 29 data points[28-31]. Plotting the experimental $\Delta\Delta G_{bind}$ vs. enrichment values for each of the four affinity gates showed that $\Delta\Delta G_{bind}$ was linearly dependent on the natural log of enrichment values in HI and LO gates (R-value of 0.87 for each of the gates, **Figure S3**). The NGS values from HI and LO gates were denoted further as functions X1 and X4, respectively. $\Delta\Delta G_{bind}$ showed a more complicated two-valued function behavior for WT and SL gates. This was expected since for these gates, the highest enrichment values were observed in the narrow range of $\Delta\Delta G_{bind}$ values but decreased for mutations that showed both higher and lower affinities compared to that narrow range of values. To eliminate this complicated multi-variable behavior and at the same time to utilize the additional information from WT and SL gates, we constructed two additional functions that multiplied enrichment values from HI and SL gates (HI x SL, denoted further as X2) and enrichment values from WT and LO gates (WT x LO, denoted further as X3). These two functions, X2 and X3, were linearly dependent on $\Delta\Delta G_{bind}$. We next used a linear regression to produce the best possible fit of experimental $\Delta\Delta G_{bind}$ values using the linear combination of four functions (X1, X2, X3, X4) arriving to the normalization formula that converts NGS enrichment values into $\Delta\Delta G_{bind}$ for the BPTI/BT complex in this particular experiment:

$$\Delta\Delta G_{bind} = -0.164\ X1 + 0.358\ X2 + 0.72\ X3 + 1.96\ X4 + 5.38 \tag{1}$$

Using 29 data points and leave-one-out cross-validation approach, where each data point was predicted without the enrichment information for that particular data point, we were able to predict experimental $\Delta\Delta G_{bind}$ values with very high accuracy over the range of more than 12 kcal/mol (**Figure 3**; R = 0.90, σ =1.5 kcal/mol). We used the obtained normalization formula to

8

convert the enrichment values to $\Delta\Delta G_{bind}$ values for all the single BPTI mutants in the library. Since the enrichment values for these mutants are within the range explored in our small experimental data set, the calculated $\Delta\Delta G_{bind}$ values are expected to have the same accuracy as reported above.
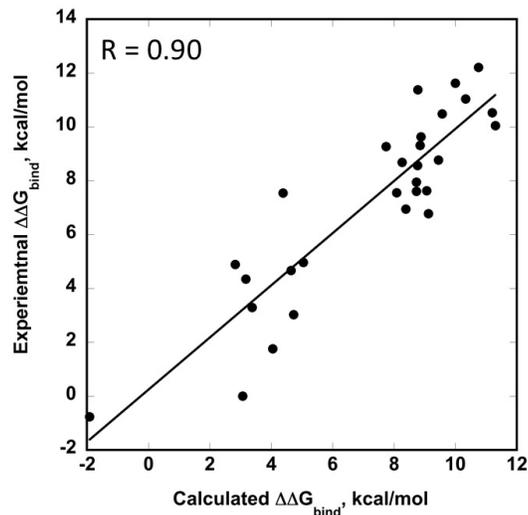
Figure 3: Correlation between the calculated $\Delta\Delta G_{bind}$ using the enrichment values from the NGS results and the $\Delta\Delta G_{bind}$ measurements for purified protein variants of BPTI interacting with BT. Leave-one-out cross validation was used to produce this graph.

Figure 4 shows the $\Delta\Delta G_{bind}$ values for all single BPTI mutants at 12 binding interface positions interacting with BT, producing a quantitative binding landscape for this PPI. As can be seen, the majority of the mutations at all positions in this PPI are highly destabilizing, producing destabilization as high as almost 12 kcal/mol. The most non-tolerant to substitution positions are 15 and 16 that lie in the core of the binding interface (**Figure 1B**). However, at position 15, one mutation, K15R, was determined to substantially stabilize the complex, in agreement with experimental results on purified proteins[30]. **Figure 4** also shows that the same type of mutations (e. g. hydrophobic or polar) frequently produce similar changes in $\Delta\Delta G_{bind}$ for the same position. We thus established that the BPTI/BT complex with the K_d of 10^{-14} M is highly optimized by nature, with most single mutations in BPTI leading to high destabilization of the PPI, and very few neutral and affinity-enhancing mutations.
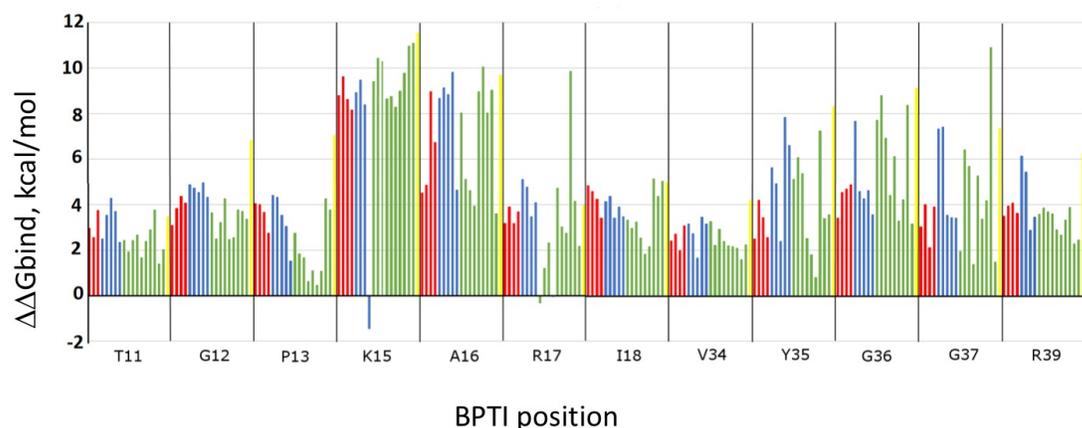
Figure 4: Changes in $\Delta\Delta G_{bind}$ for all single mutants of BPTI interacting with BT. Each bar represents a mutation to one amino acid including hydrophobic amino acids (green), polar amino acids (red), charged amino acids (blue) and Cys (yellow). X-axis shows WT residue followed by position.

In summary, we report a novel approach that allows us for the first time to produce quantitative binding landscapes based on binding selections into several affinity gates, NGS of the selected mutants, and normalization procedure using a small data set of experimentally determined $\Delta\Delta G_{bind}$ values. Very recently, a similar direction has been taken by Keating and colleagues to design mutations and improve affinity in peptide/protein interactions[32]. Unlike our study, the authors used IC50 value of yeast titrations to normalize NGS data and to predict $\Delta\Delta G_{bind}$ for various peptide mutants with K_ds in the medium affinity range. A similar thinking by two independent studies attest to attractiveness of our methodology for binding landscape mapping. Yet our work presents a simpler normalization procedure based on actual *in vitro* affinity measurements and allows to explore PPIs at the limits of the Kd spectra. We demonstrate superior correlations between $\Delta\Delta G_{bind}$ predictions and actual *in vitro* measurements over a much larger range compared to all previously reported approaches.

To achieve good prediction accuracy, the experimental data points used for normalization should show large spread in $\Delta\Delta G_{bind}$ values, including both affinity-enhancing and affinity-reducing mutations. More affinity gates could be used in future experiments, although in this case normalization would require a larger set of experimental data; at least five data points per each parameter should be used in the normalization function to avoid overfitting. Our methodology could be applied to study the evolutionary paths of any PPI regardless of its K_d value and to compare binding landscapes of various PPIs. The approach could be easily extended to

studies of double and higher-order mutational steps, providing more comprehensive information on PPI evolution and facilitating new modeling and protein engineering studies.

References

1   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* **128**, 3-13, (2017).

2   Gonzalez, M. W. & Kann, M. G. Chapter 4: Protein interactions and disease. *PLoS Comput Biol* **8**, e1002819, (2012).

3   Benedix, A., Becker, C. M., de Groot, B. L., Caflisch, A. & Bockmann, R. A. Predicting free energy changes using structural ensembles. *Nat Meth* **6**, 3-4, (2009).

4   Sharabi, O., Shirian, J. & Shifman, J. M. Predicting affinity- and specificity-enhancing mutations at protein-protein interfaces. *Biochem Soc Trans* **41**, 1166-1169, (2013).

5   Vangone, A. & Bonvin, A. M. Contacts-based prediction of binding affinity in protein-protein complexes. *Elife* **4**, e07454, (2015).

6   Sharabi, O., Dekel, A. & Shifman, J. M. Triathlon for energy functions: who is the winner for design of protein-protein interactions? *Proteins* **79**, 1487-1498, (2011).

7   Raucci, R., Laine, E. & Carbone, A. Local Interaction Signal Analysis Predicts Protein-Protein Binding Affinity. *Structure* **26**, 905-915 e904, (2018).

8   Moal, I. H., Moretti, R., Baker, D. & Fernandez-Recio, J. Scoring functions for protein-protein interactions. *Curr Opin Struct Biol* **23**, 862-867, (2013).

9   Moal, I. H. & Fernandez-Recio, J. Intermolecular Contact Potentials for Protein-Protein Interactions Extracted from Binding Free Energy Changes upon Mutation. *J Chem Theory Comput* **9**, 3715-3727, (2013).

10 Fleishman, S. J., Whitehead, T. A., Strauch, E. M., Corn, J. E., Qin, S., Zhou, H. X., Mitchell, J. C., Demerdash, O. N., Takeda-Shitaka, M., Terashi, G., Moal, I. H., Li, X., Bates, P. A., Zacharias, M., Park, H., Ko, J. S., Lee, H., Seok, C., Bourquard, T., Bernauer, J., Poupon, A., Aze, J., Soner, S., Ovali, S. K., Ozbek, P., Tal, N. B., Haliloglu, T., Hwang, H., Vreven, T., Pierce, B. G., Weng, Z., Perez-Cano, L., Pons, C., Fernandez-Recio, J., Jiang, F., Yang, F., Gong, X., Cao, L., Xu, X., Liu, B., Wang, P., Li, C., Wang, C., Robert, C. H., Guharoy, M., Liu, S., Huang, Y., Li, L., Guo, D., Chen, Y., Xiao, Y., London, N., Itzhaki, Z., Schueler-Furman, O., Inbar, Y., Potapov, V., Cohen, M., Schreiber, G., Tsuchiya, Y., Kanamori, E., Standley, D. M., Nakamura, H., Kinoshita, K., Driggers, C. M., Hall, R. G., Morgan, J. L., Hsu, V. L., Zhan, J., Yang, Y., Zhou, Y., Kastritis, P. L., Bonvin, A. M., Zhang, W., Camacho, C. J., Kilambi, K. P., Sircar, A., Gray, J. J., Ohue, M., Uchikoga, N., Matsuzaki, Y., Ishida, T., Akiyama, Y., Khashan, R., Bush, S., Fouches, D., Tropsha, A., Esquivel-Rodriguez, J., Kihara, D., Stranges, P. B., Jacak, R., Kuhlman, B., Huang, S. Y., Zou, X., Wodak, S. J., Janin, J. & Baker, D. Community-wide assessment of protein-interface modeling suggests improvements to design methodology. *J Mol Biol* **414**, 289-302, (2011).

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

12 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* **87**, 7150-7154, (1990).

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

14   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* **97**, 8950-8954, (2000).

15   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* **22**, 1-10, (2014).

16   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* **281**, 22378-22385, (2006).

17   Sharabi, O., Erijman, A. & Shifman, J. M. Computational methods for controlling binding specificity. *Methods Enzymol* **523**, 41-59, (2013).

18   Fack, F., Hugle-Dorr, B., Song, D., Queitsch, I., Petersen, G. & Bautz, E. K. Epitope mapping by phage display: random versus gene-fragment libraries. *J Immunol Methods* **206**, 43-52, (1997).

19   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* **113**, 8705-8710, (2016).

20   Clackson, T. & Wells, J. A. A hot spot of binding energy in a hormone-receptor interface. *Science* **267**, 383-386, (1995).

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

22   Boder, E. T. & Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat Biotechnol* **15**, 553-557, (1997).

23   Wrenbeck, E. E., Faber, M. S. & Whitehead, T. A. Deep sequencing methods for protein engineering and design. *Curr Opin Struct Biol* **45**, 36-44, (2017).

24   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* **30**, 543-548, (2012).

25   Forsyth, C. M., Juan, V., Akamatsu, Y., DuBridge, R. B., Doan, M., Ivanov, A. V., Ma, Z., Polakoff, D., Razo, J., Wilson, K. & Powers, D. B. Deep mutational scanning of an antibody against epidermal growth factor receptor using mammalian cell display and massively parallel pyrosequencing. *MAbs* **5**, 523-532, (2013).

26   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* **84**, 1914-1928, (2016).

27   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* **175**, 171-177, (2011).

28   Beeser, S. A., Goldenberg, D. P. & Oas, T. G. Enhanced protein flexibility caused by a destabilizing amino acid replacement in BPTI. *J Mol Biol* **269**, 154-164, (1997).

29 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* **289**, 175-186, (1999).

30 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* **48**, 419-428, (2001).

31 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* **35**, 11435-11446, (1996).

32 Jenson, J. M., Xue, V., Stretz, L., Mandal, T., Reich, L. L. & Keating, A. E. Peptide design by optimization on a data-parameterized protein interaction landscape. *Proc Natl Acad Sci U S A* **115**, E10342-E10351, (2018).