Stabilizing proteins through saturation suppressor mutagenesis.

Shahbaz Ahmed[1], Kavyashree Manjunath[2], Raghavan Varadarajan[1,*]

[1]Molecular Biophysics Unit, Indian Institute of Science, Bangalore-560012, India

[2]The Institute of Stem Cell Science and Regenerative Medicine, Bangalore 560065, India

*Author for correspondence

Telephone: +91-80-2293-. 2612

Fax: +91-80-23600535

Email: varadar@iisc.ac.in

Running Title: Screening for stability

Number of manuscript pages: 36

Number of supplementary material pages: 7

Number of figures: 12

Abstract

While there have been recent, transformative advances in the area of protein structure prediction, prediction of point mutations that improve protein stability remains challenging. It is possible to construct and screen large mutant libraries for improved activity or ligand binding, however reliable screens for mutants that improve protein stability do not exist, especially for proteins that are well folded and relatively stable. We demonstrate that incorporation of a single, specific destabilizing, (parent inactive) mutation into each member of a single-site saturation mutagenesis library followed by screening for suppressors, allows for robust and accurate identification of stabilizing mutations. When coupled to FACS sorting of a yeast surface display library of the bacterial toxin CcdB, followed by deep sequencing of sorted populations, multiple stabilizing mutations could be identified after a single round of sorting. Multiple libraries with different parent inactive mutations could be pooled and simultaneously screened to further enhance the accuracy of identification of stabilizing mutations. Individual stabilizing mutations could be combined to result in a multimutant with increase in thermal melting temperature of about 20 °C and enhanced tolerance to high temperature exposure. The method employs small library sizes and can be readily extended to other display and screening formats to rapidly isolate stabilized protein mutants.

Keywords: Yeast surface display, Stabilizing mutation, Hyper stable mutants, Second-site suppressor mutagenesis

Introduction

Directed evolution has drastically reduced the time required to engineer desired function into proteins (Hermes *et al*, 1990; Lehtovaara *et al*, 1988; Tao & Cornish, 2002; Turner, 2009). Enzymes and other proteins with altered function or binding specificity have been evolved using yeast surface display (YSD), phage display (Shusta *et al*, 2008; Wu *et al*, 2016) or other *in vivo* functional screens (Waldo *et al*, 1999; Wigley *et al*, 2001; Maxwell *et al*, 1999). Phage display utilizes its surface proteins pIII and pVIII, which are fused to the protein of interest (Mimmi *et al*, 2019). Phage display can be used to generate libraries of very high diversity (Schwimmer *et al*, 2013) which can be screened for binding to a target ligand. Agglutinin based Aga2p is a widely used system to display proteins on the yeast cell surface (Shusta *et al*, 2008). Aga2p is a small protein, covalently linked via disulphide linkages to the yeast cell surface protein Aga1(Boder & Wittrup, 1997). Different populations in a yeast library can be enriched using FACS. Relative to the phage display library, sizes are lower (~10^{11} vs 10^{7}). However, with YSD, eukaryotic post translation modifications are possible. While screening for mutants with improved binding or enzymatic activity is straight forward, it is non-trivial to screen for mutants with improved stability. Some prior studies have suggested that stabilized mutants are expressed at higher levels than wild type (WT), however, several other studies have not observed this (Shusta *et al*, 1999; Park *et al*, 2006; Piatesi *et al*, 2006). In several cases it was observed that for stable proteins, mutants with improved stability are expressed at a similar level to WT (Esteban & Zhao, 2004; Hagihara & Kim, 2002; Park *et al*, 2006; Piatesi *et al*, 2006) and hence expression alone cannot be used to discriminate mutants with higher stability from mutants with a slightly destabilized phenotype. We recently showed that the amount of active protein on the yeast cell surface (detected by the amount of bound ligand) correlates better with *in vitro* thermal stability or *in vivo* solubility than the amount of total protein on the yeast cell surface, for destabilized mutants (Ahmed *et al*, 2021). However, mutants above a certain stability threshold show similar expression and ligand binding to wild type. An alternative approach to isolate stabilizing mutations is to introduce a destabilizing mutation hereafter

referred to as parent inactivation mutation (PIM) and then create mutant libraries in this background to screen for suppressors (Araya *et al*, 2012; Blacklow *et al*, 1991; Brachmann *et al*, 1998; Foit *et al*, 2009; Sahoo *et al*, 2015). This methodology has required multiple rounds of enrichment to isolate stable mutants. The reversion of PIM to wild type or non-destabilizing mutants during library generation or enrichment enriches for mutants lacking the PIM instead of desired PIM-suppressor pairs. Additionally, in multi-round format, this methodology potentially allows isolation of only a few stabilizing mutations, and does not distinguish between allele specific and global suppressors. It is also unclear whether it will always be possible to isolate suppressors for every PIM. In the present study, we have modified this approach by introducing PIMs in the background of a single-site, saturation mutagenesis (SSM) library of bacterial toxin CcdB, sorted different populations and reconstructed the MFI of each mutant as described (Ahmed *et al*, 2021). We use both the binding MFI and expression MFI as criteria to differentiate between stabilized mutants, WT-like and destabilized mutants. Using this current methodology, we could readily identify putative stabilized mutants with a minimal number of false positives.

Results

Second site Saturation Suppressor Mutagenesis (SSSM) methodology.

In a previous study, when an SSM library of the bacterial toxin CcdB was used to isolate stabilized mutants based on their binding to GyA14 or binding/expression ratio on the yeast cell surface, a large number of false positives were obtained (Ahmed *et al*, 2021). We found that if the stability of displayed protein crosses a threshold, it is difficult to identify stabilized mutants solely based on the extent of ligand binding or expression as mutants show a similar level of binding and expression on the yeast cell surface irrespective of the stability. To minimize such threshold effects, we used single-site, saturation suppressor mutagenesis (SSSM) methodology (Figure 1), as described previously (Sahoo *et al*, 2015). Using this methodology, two different types of suppressors can be identified. Proximal

suppressors reverse the destabilizing effect of the PIM by locally compensating packing defects caused by the PIM. Proximal suppressors are allele specific and do not show stabilizing effects as a single mutant in the absence of the PIM (Nikolova, 2000; Sahoo *et al*, 2015). Distal suppressors are located far from the PIM and often act as global suppressors, reversing the effects of multiple individual PIMs. A distal suppressor also typically stabilizes the protein relative to WT, in the absence of the PIM.

Selection of PIM and sorting of Second Site Saturation Mutagenesis (SSSM) library of CcdB

CcdB is a bacterial toxin that causes bacterial cell death by binding and poisoning DNA gyrase (Bernard & Couturier, 1992). When expressed on the surface of yeast, properly folded CcdB can be detected by binding to FLAG tagged GyrA14 followed by incubations of cells with anti FLAG primary and fluorescently labeled secondary antibody (Sahoo *et al*, 2015). We selected four PIMs based on their *in vitro* thermal stability and *in vivo* activity in *E.coli* (Adkar *et al*, 2012; Sahoo *et al*, 2015). PIM V18D is completely inactive and highly aggregation prone due to a charged mutation in the core of the protein. V18G and V20G are partially inactive *in vitro* due to the formation of a cavity in the core of the protein, with V20G showing higher activity *in vivo* compared to V18G (Adkar *et al*, 2012). L36A is the most active amongst the four PIMs, the mutant protein is partially aggregated (Sahoo *et al*, 2015). SSSM libraries were constructed by introducing each PIM individually in the background of the single-site saturation mutagenesis library (Sahoo *et al*, 2015). These SSSM libraries showed variable expression and binding depending on the PIM present in the single mutant library (Supplementary Figure S1). Different populations of SSSM libraries were sorted based on the expression and binding histograms of the libraries (Supplementary Figure S2). Populations were subjected to deep sequencing and the mean fluorescence intensities of binding (MFI_{seq} (bind)) and expression (MFI_{seq} (expr)) were estimated as described (Ahmed *et al*, 2021). Values of MFI_{seq} (bind) and MFI_{seq} (expr) were used to predict stabilized mutants. Putative stabilized mutants were identified as those for which either binding or expression is enhanced by at least twenty five percent compared to the PIM alone. A subset of the predictions were validated by expressing and purifying the individual

point mutation in the WT background and measuring the T_m using thermal shift assay (TSA). In case of the V18D library, we observed that most mutations have low values of normalized MFI_{seq} (bind) as well as MFI_{seq} (expr). This might be because highly destabilizing PIM cannot be rescued by a single suppressor mutation. We observed a low sensitivity and specificity but reasonable accuracy of prediction for stabilized mutants when predictions were made based on either MFI_{seq} (bind) or MFI_{seq} (expr) (Supplementary Figure S3, Table 1) for this library. While the overall correlation between MFI_{seq} values of expression or binding with stability of the mutants was poor (Figure 2A, 2E), nevertheless the best binders showed significant stabilization. Since the results were not as promising as the other libraries, the V18D library was not included in subsequent analyses. A similar analysis was performed for the other three libraries and we found a high sensitivity of prediction for stabilized mutants for these libraries when MFI_{seq} (bind) was used as the criterion (Table 1). SSSM libraries with V20G or V18G as PIM displayed the highest sensitivity of prediction of stabilizing mutations (Figure 3A, Supplementary Figure S4A). We found some false positives for the L36A SSSM library (Supplementary Figure S5A), which had similar expression and binding compared to the single-site saturation mutant library of WT CcdB (Supplementary Figure S1). When stabilized mutant predictions were made based on the MFI_{seq} (expr), we observed lower sensitivity compared to the predictions based on MFI_{seq} (bind) (Figure 3B, Supplementary Figure S4B, S5B, Table 1) for all libraries. We also found a good correlation between the MFI_{seq} (bind) and the stability of stabilized and marginally destabilized ($0>\Delta T_m>-5$) mutants (Figure 2).

We hypothesized that the second mutation which alleviates the destabilizing effect of a PIM may act as a global suppressor and can therefore alleviate the destabilizing effects of other PIMs. To confirm this, common suppressors were shortlisted which were present in different PIM libraries (Figure 4). True global suppressors were identified as those which alleviated the destabilizing effect of at least two PIMs. Using this criterion to identify stabilized mutants we found a sensitivity of 1 for the prediction. We also found that at several positions, there were multiple mutants with stabilizing

phenotypes, which indicates that the WT is not the most preferred in terms of stability, this information can also be used as a criterion to find the stabilizing mutation. Interestingly none of the suppressors were found at the residues directly involved in GyrA binding. In any case, we have recently shown that such active-site residues can be identified based on the pattern of MFI_{seq} (bind) and MFI_{seq} (expr) in single-site saturation mutagenesis libraries and removed from the set of putative global suppressors (Ahmed *et al*, 2021).

Additive effect of stabilizing mutations

While single mutations usually do not enhance the stability of proteins to a great extent, individual stabilizing mutations can be combined resulting in multi-mutants with enhanced stability. We therefore constructed multiple double and triple mutants and measured their thermal stabilities using a thermal shift assay. All the double, triple and multi-mutants showed a higher T_m than the WT (Figure 5A). We also observed a good correlation (r =0.98) between ΔT_m of double/triple mutants with the sum of ΔT_m of individual mutants (Figure 5B). When more than 7 mutants were combined such additive effects were not observed. While combining 7 or more stabilizing mutations, we did not consider any centroid-centroid distance cut-off, which indicates possible epistatic interactions between residues in close proximity.

Thermal aggregation analysis of stabilized mutants

To ascertain the ability of mutations to prevent loss of function from transient high temperature exposure, CcdB mutants were incubated at elevated temperatures for 1 hour, this can result in unfolding of CcdB and subsequent irreversible aggregation (Bajaj *et al*, 2004). The fraction of active protein remaining after incubation was assessed by its ability to bind GyrA at room temperature using SPR. In the case of WT, we did not observe a large decrease in the fraction of active protein except when incubated at temperatures at and above 60 ^{o}C (Figure 6A). After incubation at 80 ^{o}C, the protein was completely denatured, and did not show any binding with Gyrase. In contrast, one stabilized single

mutant R10G and three multi-site mutants retained significant activity after incubation at 80 ^{o}C for one hour (Figure 6B). Six mutants showed a higher fraction of active protein than WT after heating at 60 ^{o}C (Figure 6C). Three single mutants and two double mutants showed no reduction in active fraction of protein. Surprisingly, the triple mutant, (R10G/A37V/R86H) which showed very high thermal stability, aggregated completely at 80 ^{o}C showing that thermal stability and thermal tolerance need not always be correlated.

Sorting of multiple libraries simultaneously

When the suppressor mutations were identified using both enhanced binding to ligand relative to the corresponding PIM, as well as alleviating the destabilizing effect of at least two PIMs as the criteria, predictions were highly specific. Hence, to find a larger number of stabilizing mutants, it is desirable to screen multiple PIM libraries. However, screening multiple individual libraries is laborious. In most FACS based library screens, individual libraries are sorted for multiple rounds to enrich for the best binders. If this approach is applied to pooled libraries with multiple PIMs, the most stable PIM will dominate, resulting in the enrichment of mutants only from this library. To confirm this, we therefore sorted the pooled library as explained (Supplementary Figure S6A) and we found that mutants with the highest binding were largely from the L36A library as this PIM has the weakest effect on binding (Supplementary Figure S6B). To overcome this problem, we instead sorted multiple populations based on binding of the pooled library (Supplementary Figure S6C) and compared the relative binding of putative (PIM, suppressor) pairs with those of individual PIMs. We found a good correlation between the MFI_{seq} (bind) of mutants from the pooled library, with those from individually analysed libraries (Supplementary Figure S6D), with an increase in the correlation as the read cut-off increased. This suggests that single round sorting of YSD pooled SSSM libraries can rapidly identify stabilized mutations.

Discussion

Directed evolution has drastically reduced the time required to design proteins with new activities. Directed evolution is performed most often in conjunction with display techniques like phage and yeast surface display, which involve selection of better binders to a target of interest from large libraries. While this approach readily selects for high affinity binders, selecting for stable proteins is more difficult. Phage display libraries have very high diversity (Chao *et al*, 2004). However, in phage display experiments, there is less control over the selection of populations during enrichment and effects of post-translational modification including glycosylation cannot be studied. Yeast display enables the selection of populations during enrichment using FACS (Shusta *et al*, 2008). Previously, to find stabilized variants of proteins with high intrinsic stability, YSD libraries were subjected to thermal stress, to enrich for more stable variants followed by sorting to identify variants which retained binding to a conformation specific ligand (Pavoor *et al*, 2012). While this is potentially useful, yeast cells cannot replicate after high temperature exposure, and hence the method requires repeated rounds of plasmid isolation, PCR amplification and retransformation in yeast cells after each round of enrichment. Also, if a protein exhibits reversible thermal unfolding, enrichment of stabilized mutants in such cases will be difficult. Nevertheless, it would be interesting to couple this methodology to the single round sorting and analysis described in the present study. In a recent report, expression level was used to identify a few stabilized mutants of SARS CoV-2 RBD (Starr *et al*, 2020). In the present study, we did not see a good correlation between expression level of individual mutants and thermal stability of stabilized mutants. Instead, we observed that MFI_{seq} (bind) is a better predictor of protein stability than MFI_{seq} (expr). Previously, we found that for stable proteins it is difficult to isolate stabilizing mutants using yeast surface display, as the surface expression and binding of all the mutants above a threshold stability are often similar (Ahmed *et al*, 2021). To overcome this problem, we introduced a PIM in the SSM library of CcdB which reduced binding of all the mutants present in the library and then selected for suppressors which showed improved binding compared to the PIM. We

hypothesized that if a stabilizing mutation alleviates the destabilizing effect of at least two PIMs, it is likely to be a true global suppressor. We therefore introduced four different PIMs with varied stability in the SSM library.

For libraries with the V18G, V20G and L36A PIMs used in the study, a high specificity for the prediction was observed. These PIM libraries also showed a good correlation between protein stability and MFIseq (bind) (Figure 2). The identified mutants also showed additive stabilization when double or triple mutants were combined.

We also found that if a greater number of PIM libraries are screened, it enhances the prediction of stabilized mutants. To decrease the time and effort involved in screening multiple libraries, we pooled multiple libraries. The pooled library showed similar reconstructed MFIs of mutants to those obtained from individually analysed libraries, validating the approach used. Importantly, the size of libraries employed here is relatively small, identical to single-site saturation mutagenesis libraries, with a total size of 32N, where N is the length of the protein sequence. It can thus easily be extended to other library technologies including lentiviral and transposon libraries, phage and mammalian display.

The methodology has the following limitations. The methodology requires a conformation specific ligand to differentiate between the destabilized PIM and stabilized PIM-suppressor pair. For some proteins hyper glycosylation on the yeast cell surface may interfere with the binding between yeast surface displayed protein and its ligand. As an alternative to ligand binding, protease resistance can also be potentially used to select for stabilized mutants (Rocklin *et al*, 2017; Golinski *et al*, 2021). However, the aggregation and unfolding on the yeast cell surface may limit the cleavage of such proteins. Use of a protease assay to screen for stabilized mutants, assumes that protease cleavage occurs primarily in the unfolded state (Park & Marqusee, 2005). While this is true for small well folded proteins, for larger proteins, initial sites of cleavage may occur at surface loops, complicating interpretation of protease screening results. Overall the present methodology offers a robust pathway

to identify stabilizing mutations in any protein of interest for which a surface display based binding screen is available.

Materials and methods

Bacterial strains, yeast strains and plasmids

E.coli strain Top10Gyrase has a mutation in the gyrA gene which prevent CcdB toxicity. The EBY100 strain of *Saccharomyces cerevisiae* has TRP1 mutation which make it auxotrophic for tryptophan and transformants can be selected on minimal media. WT and mutant *ccdB* genes were coned in pBAD24 plasmid for controllable expression in *E.coli*. pPNLS shuttle vector was used to display ccdB mutants for yeast cell surface expression.

Purification of WT and mutant CcdB proteins

CcdB WT and mutant proteins were purified as described (Chattopadhyay & Varadarajan, 2019), briefly, overnight grown culture was diluted 100 folds in 300 ml LB media containing ampicillin (100µg/ml). The cells were grown and induce at an OD_{600} ~0.5 for 3 hours at 37 ^{o}C. The cells were harvested after induction and lysed using sonication. The soluble fraction was separated using centrifugation and incubated with Affigel-15 coupled to CcdA peptide (residue 45-72) for 2 hours at 4 ^{o}C. The unbound fraction was removed, and beads were washed with bicarbonate buffer (50 mM NaHCO3 and 500 mM NaCl). The proteins were eluted with glycine (200 mM, pH 2.5) and collected in tubes containing an equal volume of HEPES buffer (400 mM, pH 8)

Second site Saturation Suppressor Mutagenesis (SSSM) library generation

A ccdB site saturation mutagenesis (SSM) library was generated through an inverse PCR based approach (Jain & Varadarajan, 2014). The second site suppressor mutant library for CcdB was generated by introducing PIMs in the SSM library as described (Sahoo *et al*, 2015).

Yeast surface expression and sorting of SSSM library

Yeast surface display and flowcytometric analysis were performed as explained earlier (Ahmed *et al*, 2021). Briefly, *Saccharomyces cerevisiae* EBY100 cells containing WT ccdB or mutant plasmids were grown in SDCAA media for 16 hours and induced in SGCAA media for an additional 16 hours at 30 ^{o}C. Ten million cells were taken for FACS sample preparation. The Second site Suppressor mutant library of CcdB was sorted based on 1D sorting of surface expression and binding to1000 nM GyrA14. The estimated K_d for WT CcdB to GyrA14 is 4 nM. A higher CcdB concentration was employed to ensure that even destabilized mutants where only a small fraction was properly folded, showed detectable binding to CcdB. Surface expression was quantitated by fluorescent antibody binding to HA tagged CcdB. Binding to GyrA14 was detected by anti FLAG antibody as described previously (Sahoo *et al*, 2015). The sorting of CcdB libraries was performed using a BD Aria III cell sorter. In the case of simultaneous sorting of multiple SSSM libraries, each library samples were prepared separately as explained above and pooled before sorting.

Sample preparation for deep sequencing

Deep sequencing samples were prepared as explained earlier (Ahmed *et al*, 2021). Briefly, sorted cells were grown on SDCAA agar plates, colonies were scraped, and pooled plasmids were extracted. The ccdB gene was PCR amplified using the primers having multiplex identifier (MID) sequence at the 5'end, that bind upstream and downstream of the ccdB gene. The DNA was PCR amplified for 15 cycles, equal amounts of DNA from each sorted population were pooled, gel extracted, and the library was generated using TruSeq™ DNA PCR-Free kit from Illumina. The sequencing was done on an Illumina HiSeq 2500 250PE platform at Macrogen, South Korea.

Analysis of deep sequencing data

Deep sequencing data for the ccdB mutants were processed as described (Ahmed *et al*, 2021). Briefly, the paired end reads were assembled using the PEAR v0.9.6 (Paired-End Read Merger) tool (Zhang *et*

al, 2014). Following assembly, reads were filtered to eliminate those that do not contain the relevant MID and/or primers along with the reads having mismatched MID's. Only those reads which have bases with Phred score ≥ 20 are retained. Reads which pass the assembling and filtering step were binned according to the respective MIDs. Binned reads were aligned with the wild type ccdB sequence using the Water v6.4.0.0 program (Smith & Waterman, 1981) and reformatted. Finally the reads were classified based on insertions, deletions and substitutions (single, double, etc mutants).

MFI reconstruction from deep sequencing data

The MFI of each mutant was reconstructed as described (Ahmed *et al*, 2021). Briefly, reads of each mutant were normalized across different bins (Equation 1). The fraction of each mutant (X*i*) distributed across each bin was calculated (Equation 2). The reconstructed MFI (MFI_{seq}) of individual mutant was calculated by the summation of the product, obtained upon multiplying the fraction (X*i*) of that mutant in bin (*i*) with the MFI of the corresponding bin obtained from the FACS experiment (F*i*), across the various bins populated by that mutant (Equation 3). The normalized MFI of each mutant was calculated from the reconstructed MFI of each mutant (Equation 4).

Normalized read of mutant in bin *i* (N*i*) = $\frac{No.of\ mutant\ i\ in\ bin\ i}{\sum Reads\ in\ bin\ i}$ … Equation 1

Fraction of mutants in each gate (X*i*) = $\frac{Ni}{\sum_1^n Ni}$ … Equation 2

Reconstructed MFI = $\sum_1^n Fi * Xi$ … Equation 3

Normalized MFI = $\frac{Reconstructed\ MFI\ of\ mutant\ i}{Reconstructed\ MFI\ of\ mutant\ WT}$ … Equation 4

Stabilized mutants were classified as those which showed at least 25% enhanced binding or expression when present as PIM-suppressor pair compared to the PIM alone.

Protein thermal stability measurement.

This was carried out as described (Chattopadhyay & Varadarajan, 2019). Briefly, a solution of total volume 20 µL containing 10 µM of the purified CcdB protein and 2.5x Sypro orange dye in buffer (200 mM HEPES, 100 mM glycine), pH 7.5 was heated from 15 ^{o}C to 90 ^{o}C with 0.5 ^{o}C increment every 30 seconds on an iCycle iQ5 Real Time Detection System (Bio-Rad, Hercules, CA). The normalized fluorescence data was plotted against temperature (Niesen *et al*, 2007).

Thermal aggregation studies of CcdB mutants.

CcdB mutants and WT proteins (500 nM, 200 µL for each of the proteins) were subjected to incubation at four different temperatures, 20, 40, 60 and 80 ^{o}C on an iCycle iQ5 Real Time Detection System (Bio-Rad, Hercules, CA). The temperature was gradually increased to the desired temperature at a rate of 3 ^{o}C/min and samples were kept at the desired temperature for 1 hour. The heated protein was then cooled down to 4 ^{o}C at the rate of 3 ^{o}C/min. The aggregated protein was removed using centrifugation at 18000g. The fraction of active protein remaining was measured by binding to GyrA14 on a Biacore 2000 SPR platform. The percentage of active protein at different temperatures was calculated using the following equation.

$$\%Active = \frac{Binding\ of\ CcdB\ after\ incubation\ at\ temperature\ T\ (RU)}{Binding\ of\ CcdB\ after\ incubation\ at\ 20\ ^{O}\mathrm{C}\ \ (RU)} * 100$$

Statistical Analysis

All the data were plotted using the GraphPad Prism software 9.0.0. The correlation coefficients between deep sequencing replicates were estimated using the GraphPad Prism software 9.0.

Data Availability

All the data are in the manuscript.

Competing Interest Statement: The authors claim no conflict of interest.

Author contribution

R.V. and S.A. designed the experiments. S.A. performed all the experiments. R.V. and S.A. analyzed all the data. K.M. wrote the software and carried out the processing of the deep sequencing data. S.A. analyzed the deep sequencing data of CcdB. R.V. and S.A. wrote most of the manuscript.

Acknowledgements

This work was funded in part by a grant to RV from the Department of Science and Technology, grant number-EMR/2017/004054, DT.15/12/2018), Government of India and Department of Biotechnology, grant no. BT/COE/34/SP15219/2015 DT. 20/11/2015, Ministry of Science and Technology, Government of India. We also acknowledge funding for infrastructural support from the following programs of the Government of India: DST FIST, UGC Centre for Advanced study, Ministry of Human Resource Development (MHRD), and the DBT IISc Partnership Program. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication. S.A. acknowledges Council of Scientific & Industrial Research for his fellowship (SPM-07/079(0218)/2015-EMR-I). K.M. is thankful to Department of Science and Technology (DST) Science and Engineering Research Board for financial support, sanction order no: PDF/2017/002641. Aparna Ashok is duly acknowledged for FACS. Nonavinakere Seetharam Srilatha is duly acknowledged for the SPR experiments.

References

Adkar B V., Tripathi A, Sahoo A, Bajaj K, Goswami D, Chakrabarti P, Swarnkar MK, Gokhale RS & Varadarajan R (2012) Protein model discrimination using mutational sensitivity derived from deep sequencing. *Structure* **20:** 371–381 Available at: https://www.sciencedirect.com/science/article/pii/S0969212612000068 [Accessed June 6, 2019]

Ahmed S, Manjunath K & Varadarajan R (2021) Prediction of residue-specific contributions to binding and thermal stability using yeast surface display. *bioRxiv***:** 2021.05.31.446445 Available at: http://biorxiv.org/content/early/2021/06/01/2021.05.31.446445.abstract

Araya CL, Fowler DM, Chen W, Muniez I, Kelly JW & Fields S (2012) A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. *Proc. Natl. Acad. Sci.* **109:** 16858–16863 Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.1209751109 [Accessed July 11, 2019]

Bajaj K, Chakshusmathi G, Bachhawat-Sikder K, Surolia A & Varadarajan R (2004) Thermodynamic characterization of monomeric and dimeric forms of CcdB (controller of cell division or death B protein). *Biochem. J.* **380:** 409–17 Available at: http://www.ncbi.nlm.nih.gov/pubmed/14763902 [Accessed July 27, 2019]

Bernard P & Couturier M (1992) Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *J. Mol. Biol.* **226:** 735–745 Available at: https://www.sciencedirect.com/science/article/pii/002228369290629X [Accessed June 8, 2019]

Blacklow SC, Liu KD & Knowles JR (1991) Stepwise improvements in catalytic effectiveness: independence and interdependence in combinations of point mutations of a sluggish triosephosphate isomerase. *Biochemistry* **30:** 8470–6 Available at: http://pubs.acs.org/doi/abs/10.1021/bi00098a026 [Accessed July 26, 2019]

Boder ET & Wittrup KD (1997) Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **15:** 553–7 Available at: http://www.ncbi.nlm.nih.gov/pubmed/9181578 [Accessed July 11, 2019]

Brachmann RK, Yu K, Eby Y, Pavletich NP & Boeke JD (1998) Genetic selection of intragenic suppressor mutations that reverse the effect of common p53 cancer mutations. *EMBO J.* **17:** 1847–59 Available at: http://emboj.embopress.org/cgi/doi/10.1093/emboj/17.7.1847 [Accessed July 27, 2019]

Chao G, Cochran JR & Wittrup KD (2004) Fine epitope mapping of anti-epidermal growth factor receptor antibodies through random mutagenesis and yeast surface display. *J. Mol. Biol.* **342:** 539–550

Chattopadhyay G & Varadarajan R (2019) Facile measurement of protein stability and folding kinetics using a nano differential scanning fluorimeter. *Protein Sci.* **28:** 1127–1134 Available at: https://onlinelibrary.wiley.com/doi/abs/10.1002/pro.3622 [Accessed July 26, 2019]

Esteban O & Zhao H (2004) Directed evolution of soluble single-chain human class II MHC molecules. *J. Mol. Biol.* **340:** 81–95 Available at: https://www.sciencedirect.com/science/article/pii/S0022283604005017?via%3Dihub#BIB25 [Accessed June 6, 2019]

Foit L, Morgan GJ, Kern MJ, Steimer LR, von Hacht AA, Titchmarsh J, Warriner SL, Radford SE & Bardwell JCA (2009) Optimizing Protein Stability In Vivo. *Mol. Cell* **36:** 861–871 Available at: http://www.ncbi.nlm.nih.gov/pubmed/20005848 [Accessed July 27, 2019]

Golinski AW, Mischler KM, Laxminarayan S, Neurock NL, Fossing M, Pichman H, Martiniani S & Hackel BJ (2021) High-throughput developability assays enable library-scale identification of producible protein scaffold variants. *Proc. Natl. Acad. Sci.* **118:** Available at: https://www.pnas.org/content/118/23/e2026658118 [Accessed July 15, 2021]

Hagihara Y & Kim PS (2002) Toward development of a screen to identify randomly encoded, foldable

sequences. *Proc. Natl. Acad. Sci. U. S. A.* **99:** 6619–6624 Available at: http://www.ncbi.nlm.nih.gov/pubmed/11997470 [Accessed June 6, 2019]

Hermes JD, Blacklow SC & Knowles JR (1990) Searching sequence space by definably random mutagenesis: improving the catalytic potency of an enzyme. *Proc. Natl. Acad. Sci. U. S. A.* **87:** 696–700 Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.87.2.696 [Accessed July 22, 2019]

Jain PC & Varadarajan R (2014) A rapid, efficient, and economical inverse polymerase chain reaction-based method for generating a site saturation mutant library. *Anal. Biochem.* **449:** 90–8 Available at: http://www.ncbi.nlm.nih.gov/pubmed/24333246

Lehtovaara PM, Koivula AK, Bamford J & Knowles JK (1988) A new method for random mutagenesis of complete genes: enzymatic generation of mutant libraries in vitro. *Protein Eng.* **2:** 63–8 Available at: http://www.ncbi.nlm.nih.gov/pubmed/3151018 [Accessed July 22, 2019]

Maxwell KL, Mittermaier AK, Forman-Kay JD & Davidson AR (1999) A simple in vivo assay for increased protein solubility. *Protein Sci.* **8:** 1908–1911 Available at: http://www.ncbi.nlm.nih.gov/pubmed/10493593 [Accessed July 27, 2019]

Mimmi S, Maisano D, Quinto I & Iaccino E (2019) Phage Display: An Overview in Context to Drug Discovery. *Trends Pharmacol. Sci.* **40:** 87–91 Available at: https://linkinghub.elsevier.com/retrieve/pii/S016561471830230X [Accessed July 11, 2019]

Niesen FH, Berglund H & Vedadi M (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2:** 2212–2221 Available at: http://www.ncbi.nlm.nih.gov/pubmed/17853878 [Accessed June 6, 2019]

Nikolova P V. (2000) Mechanism of rescue of common p53 cancer mutations by second-site suppressor mutations. *EMBO J.* **19:** 370–378

Park C & Marqusee S (2005) Pulse proteolysis: A simple method for quantitative determination of

protein stability and ligand binding. *Nat. Methods* **2:** 207–212 Available at: https://pubmed.ncbi.nlm.nih.gov/15782190/ [Accessed August 5, 2021]

Park S, Xu Y, Stowell XF, Gai F, Saven JG & Boder ET (2006) Limitations of yeast surface display in engineering proteins of high thermostability. *Protein Eng. Des. Sel.* **19:** 211–217 Available at: http://academic.oup.com/peds/article/19/5/211/1556411/Limitations-of-yeast-surface-display-in [Accessed June 6, 2019]

Pavoor T V., Wheasler JA, Kamat V & Shusta E V. (2012) An enhanced approach for engineering thermally stable proteins using yeast display. *Protein Eng. Des. Sel.* **25:** 625–630 Available at: /pmc/articles/PMC3449400/?report=abstract [Accessed November 12, 2020]

Piatesi A, Howland SW, Rakestraw JA, Renner C, Robson N, Cebon J, Maraskovsky E, Ritter G, Old L & Wittrup KD (2006) Directed evolution for improved secretion of cancer-testis antigen NY-ESO-1 from yeast. *Protein Expr. Purif.* **48:** 232–242 Available at: https://linkinghub.elsevier.com/retrieve/pii/S1046592806000337 [Accessed June 6, 2019]

Rocklin GJ, Chidyausiku TM, Goreshnik I, Ford A, Houliston S, Lemak A, Carter L, Ravichandran R, Mulligan VK, Chevalier A, Arrowsmith CH & Baker D (2017) Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science* **357:** 168–175 Available at: http://www.ncbi.nlm.nih.gov/pubmed/28706065

Sahoo A, Khare S, Devanarayanan S, Jain PC & Varadarajan R (2015) Residue proximity information and protein model discrimination using saturation-suppressor mutagenesis. *Elife* **4:** e09532 Available at: https://elifesciences.org/articles/09532 [Accessed July 18, 2019]

Schwimmer LJ, Huang B, Giang H, Cotter RL, Chemla-Vogel DS, Dy F V., Tam EM, Zhang F, Toy P, Bohmann DJ, Watson SR, Beaber JW, Reddy N, Kuan HF, Bedinger DH & Rondon IJ (2013) Discovery of diverse and functional antibodies from large human repertoire antibody libraries. *J. Immunol. Methods* **391:** 60–71

Shusta E, Pepper L, Cho Y & Boder E (2008) A Decade of Yeast Surface Display Technology: Where Are We Now? *Comb. Chem. High Throughput Screen.* **11:** 127–134

Shusta E V., Kieke MC, Parke E, Kranz DM & Wittrup KD (1999) Yeast polypeptide fusion surface display levels predict thermal stability and soluble secretion efficiency. *J. Mol. Biol.* **292:** 949–956 Available at: http://www.ncbi.nlm.nih.gov/pubmed/10512694

Smith TF & Waterman MS (1981) Identification of common molecular subsequences. *J. Mol. Biol.* **147:** 195–197 Available at: http://www.ncbi.nlm.nih.gov/pubmed/7265238 [Accessed June 6, 2019]

Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, Navarro MJ, Bowen JE, Tortorici MA, Walls AC, King NP, Veesler D & Bloom JD (2020) Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* **182:** 1295-1310.e20

Tao H & Cornish VW (2002) Milestones in directed enzyme evolution. *Curr. Opin. Chem. Biol.* **6:** 858–64 Available at: http://www.ncbi.nlm.nih.gov/pubmed/12470742 [Accessed July 28, 2019]

Tripathi A, Gupta K, Khare S, Jain PC, Patel S, Kumar P, Pulianmackal AJ, Aghera N & Varadarajan R (2016) Molecular Determinants of Mutant Phenotypes, Inferred from Saturation Mutagenesis Data. *Mol. Biol. Evol.* **33:** 2960–2975

Turner NJ (2009) Directed evolution drives the next generation of biocatalysts. *Nat. Chem. Biol.* **5:** 567–573 Available at: http://www.ncbi.nlm.nih.gov/pubmed/19620998 [Accessed July 28, 2019]

Waldo GS, Standish BM, Berendzen J & Terwilliger TC (1999) Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.* **17:** 691–695 Available at: http://www.nature.com/articles/nbt0799_691 [Accessed July 25, 2019]

Wigley WC, Stidham RD, Smith NM, Hunt JF & Thomas PJ (2001) Protein solubility and folding monitored in vivo by structural complementation of a genetic marker protein. *Nat. Biotechnol.* **19:**

131–136 Available at: http://www.nature.com/articles/nbt0201_131 [Accessed July 25, 2019]

Wu C-H, Liu I-J, Lu R-M & Wu H-C (2016) Advancement and applications of peptide phage display technology in biomedical science. *J. Biomed. Sci.* **23:** 8 Available at: http://www.ncbi.nlm.nih.gov/pubmed/26786672 [Accessed July 11, 2019]

Zhang J, Kobert K, Flouri T & Stamatakis A (2014) PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30:** 614–620 Available at: http://www.ncbi.nlm.nih.gov/pubmed/24142950 [Accessed June 6, 2019]

Figures

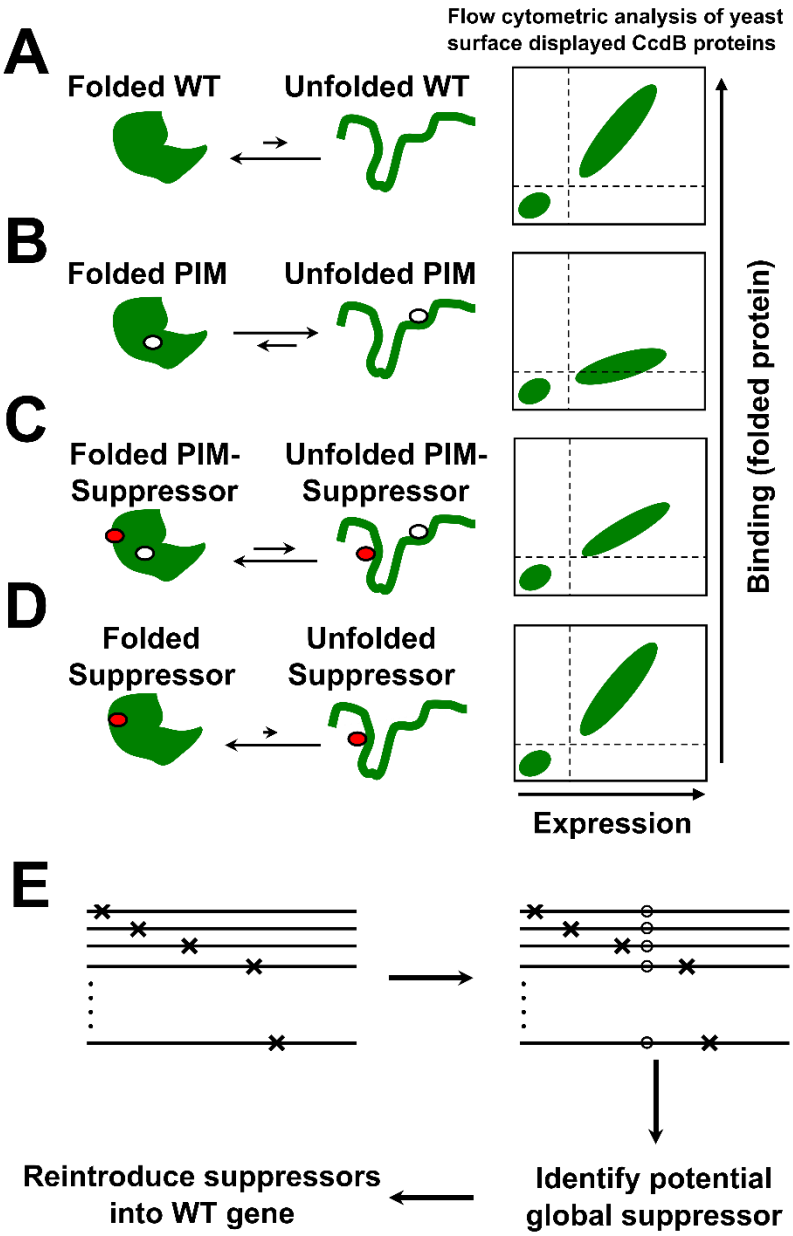


Figure 1: Schematic representation of Second Site Suppressor Mutagenesis (SSSM) methodology. Proteins exist in an equilibrium between folded and unfolded states. (A) WT proteins generally have the equilibrium shifted towards the folded state. Such proteins when expressed on the yeast cell surface show good expression and binding to their cognate ligand (B) Introduction of a Parent Inactive Mutation (PIM) shift the equilibrium towards the unfolded state, and decrease the amount of properly folded protein on the yeast cell surface, leading to decreased ligand binding. (C)

Second-site suppressor mutation, distal from and present in the background of the PIM will reduce the amount of unfolded protein present at equilibrium. Such double mutants have higher expression and binding on the yeast cell surface compared to the PIM alone. (D) Such global/distal suppressor mutations can stabilize multiple PIMs and also stabilize the WT protein, although the expression and binding of suppressor alone on the yeast cell surface is similar to the WT protein. (E) Saturation suppressor libraries are created by introducing a PIM (o) into the background of a single-site (x) saturation mutagenesis library, selecting for suppressors, and reintroducing identified suppressors in the background of the WT gene.

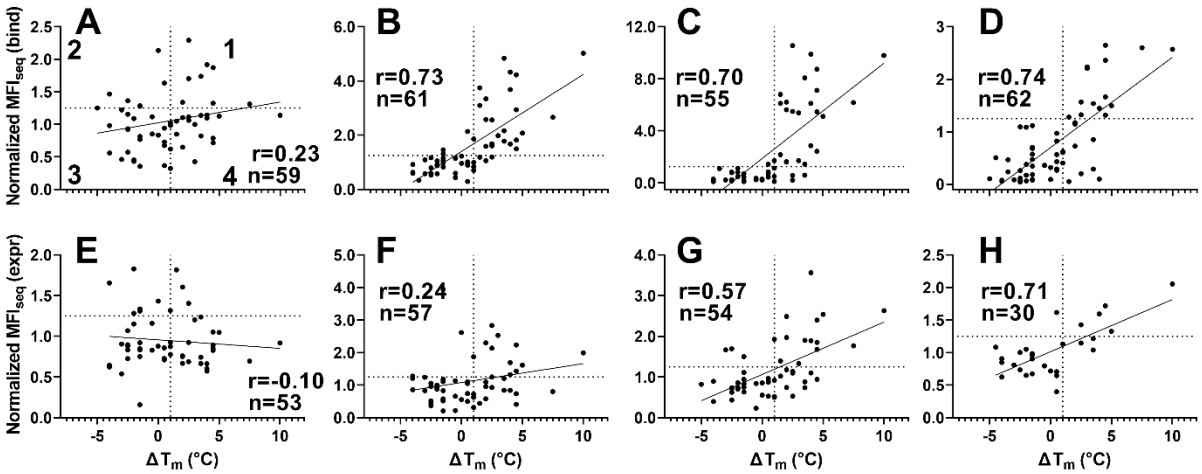


Figure 2: Correlation of CcdB mutant stability (ΔT_m of single mutant) with normalized MFI_{seq} (bind) or MFI_{seq} (expr) of (PIM, mutant) pair in PIM libraries. MFI_{seq} values of double mutants were normalized with the MFI_{seq} values of their respective PIM. First, second, third and fourth quadrants numbered in (A) represent true positive, false positive, true negative and false negative points respectively. Normalized MFI_{seq} (bind) correlation with thermal stability for V18D (A), V18G (B), V20G (C) and L36A (D) libraries. Normalized MFI_{seq} (expr) correlation with thermal stability for V18D (E), V18G (F), V20G (G) and L36A (H) libraries. Normalized MFI_{seq} (bind) correlates better than Normalized MFI_{seq} (expr) with thermal stability. For V18D while overall correlation is poor, those mutants with highest MFI_{seq} (bind) are stabilized.

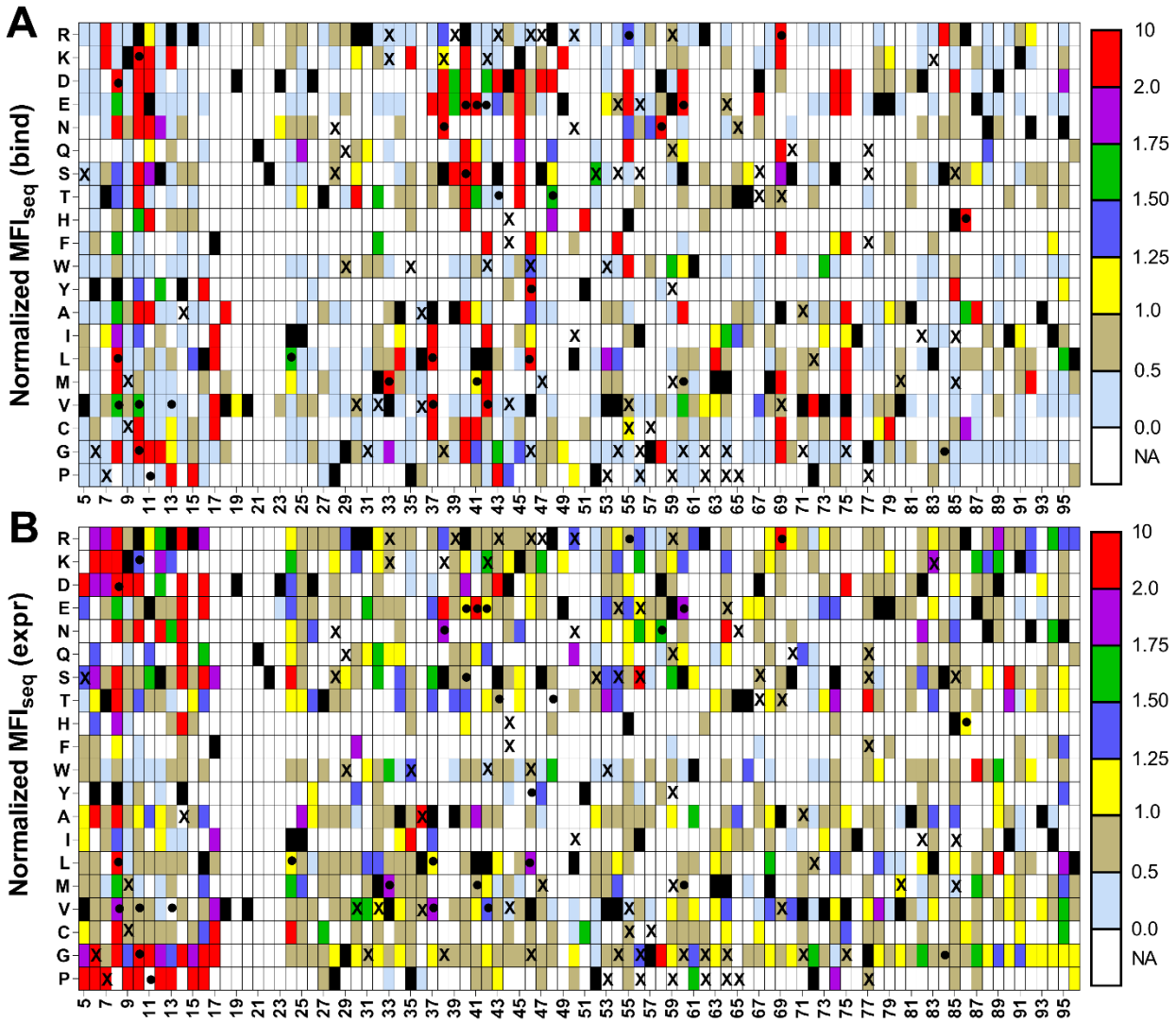


Figure 3: Heat map of normalized MFI_{seq} (bind) and normalized MFI_{seq} (expr) for V20G library. Normalized (with respect to V20G) values of MFI_{seq} (bind) (A) and MFI_{seq} (expr) (B) were categorized in different ranges. Mutants with normalized MFI_{seq} (bind) or MFI_{seq} (expr) greater than 1.25 were categorized as putative stabilized mutants. Black rectangle represents WT residues, mutants where no data are available are indicated with a white rectangle. A subset of mutants were purified and their thermal stability (T_m) *in vitro* was measured. Stable mutants (ΔT_m>1) are indicated with a "●" and destabilized mutants (ΔT_m<0) are indicated with an "X". It is clear that MFI_{seq} (bind) is superior to MFI_{seq} (expr) in identification of stabilized mutants.

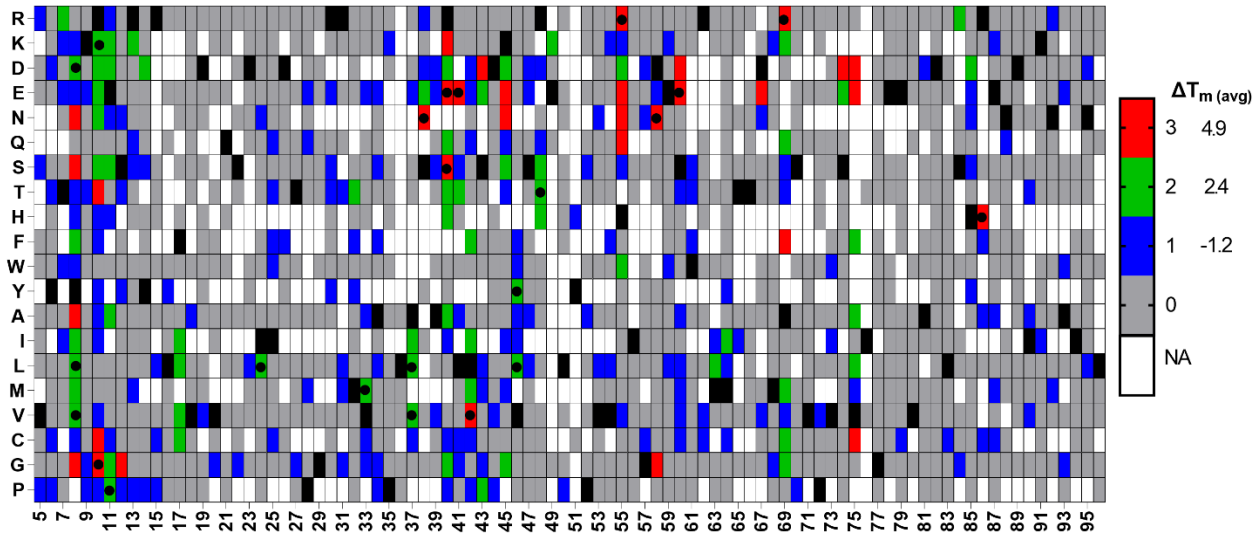


Figure 4: Heat map of putative stabilized mutants found in all the three V18G, V20G and L36A libraries based on MFI_{seq} (bind). Putative suppressor mutants in each library were given a score of one unit. Grey, blue red and green indicate that the mutant acted as a suppressor in zero, one, two and three different libraries. White indicate the mutants where no data are available.

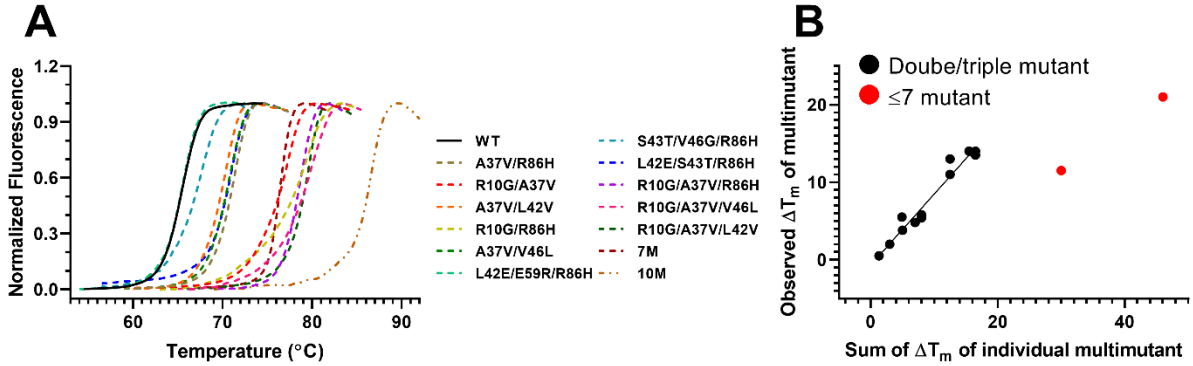


Figure 5: Thermal shift assay data for select CcdB double, triple and multi-site mutants. (A) Data for WT and mutants are shown in black and colour respectively. Multi-mutants showed higher thermal stability compared to individual mutants. Mutations present in 7M are Y8D/R10G/E11P/S12G/A37V/R40S/A69R and 10M are Y8D/R10G/E11P/S12G/A37V/R40S /L42V/V46L/A69R/R86H. (B) The multi-mutants showed an additive effect when two or three mutations were combined. Multi-mutant which contain seven or more mutations did not show completely additive stabilization.

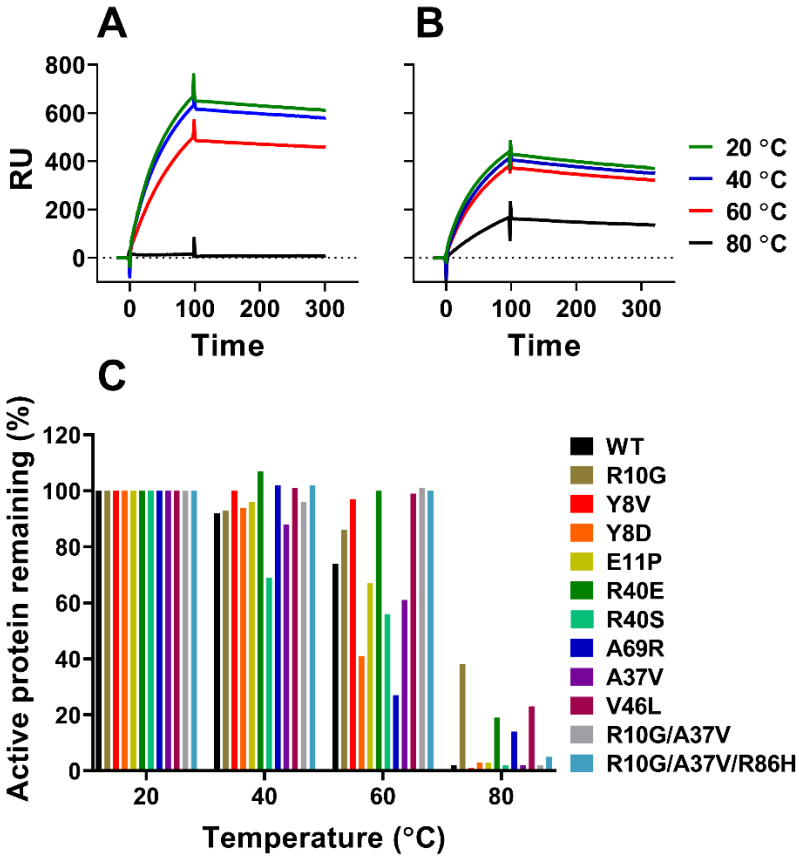


Figure 6: Estimation of the fraction of active protein after thermal stress. CcdB WT and mutants were incubated at 20, 40, 60 and 80 ^{o}C for one hour. The fraction of active protein was subsequently estimated by assaying binding to GyrA at 25 ^{o}C using SPR. A representative SPR sensorgram for (A) WT CcdB and (B) R10G CcdB showing the relative amount of active protein remaining in the samples after incubation at different temperatures for one hour. (C) Fraction (%) of active protein after incubation at indicated temperature for 1 hour.

Tables

Table 1:

Identification of stabilized mutant solely from normalized MFI_{seq} (bind) or MFI_{seq} (expr) data. Mutants were predicted to be stabilized if the corresponding MFI_{seq} value was greater than 25 percent. CcdB mutant thermal stability data was taken from previous studies (Tripathi *et al*, 2016; Ahmed *et al*, 2021) as well as additional mutant stability data measured in this study (Supplementary Table 1).

Parameter	**V18D library**		**V18G library**		**V20G library**		**L36A library**	
	Bind	Expr	Bind	Expr	Bind	Expr	Bind	Expr
Sensitivity[a]	0.27	0.15	0.80	0.44	0.79	0.54	0.63	0.56
Specificity[a]	0.73	0.88	0.89	0.79	0.98	0.78	1.00	0.91
Accuracy[a]	0.57	0.63	0.86	0.67	0.90	0.70	0.89	0.84

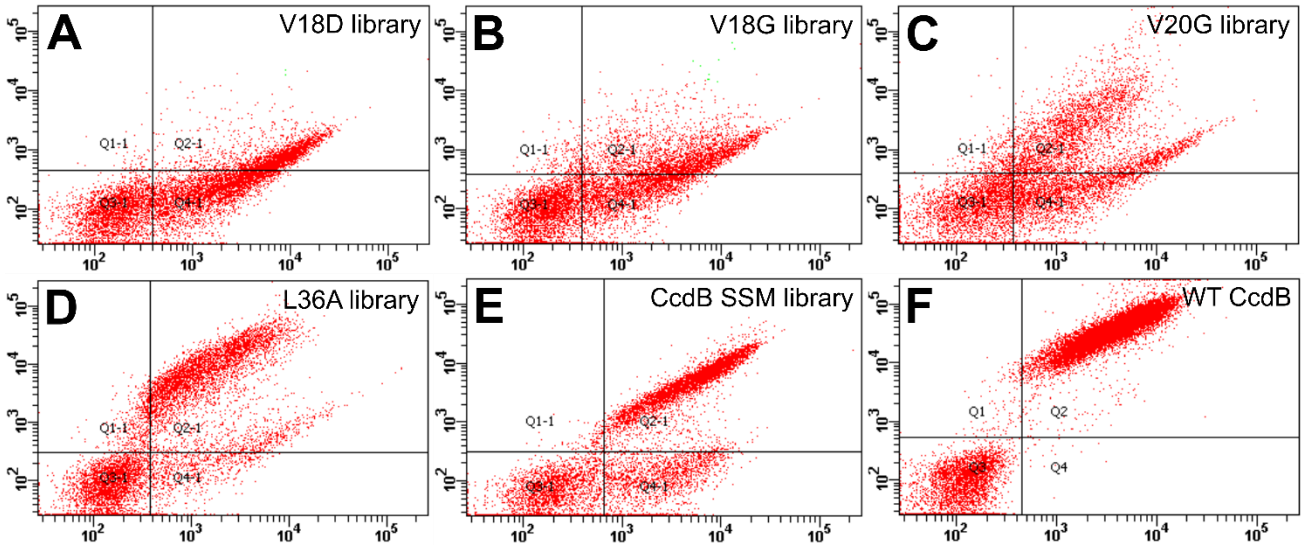
[a]The sensitivity, specificity and accuracy were calculated using the following formulae, where TP, TN, FP and FN correspond to true positive (ΔT_m (predicted) > 0, ΔT_m (observed) > 0) , true negative (ΔT_m (predicted) < 0, ΔT_m (observed)< 0), false positive (ΔT_m (predicted) > 0, ΔT_m (observed) < 0) and false negative (ΔT_m (predicted) < 0, ΔT_m (observed) > 0) respectively..

Sensitivity = $\frac{TP}{TP+FN}$

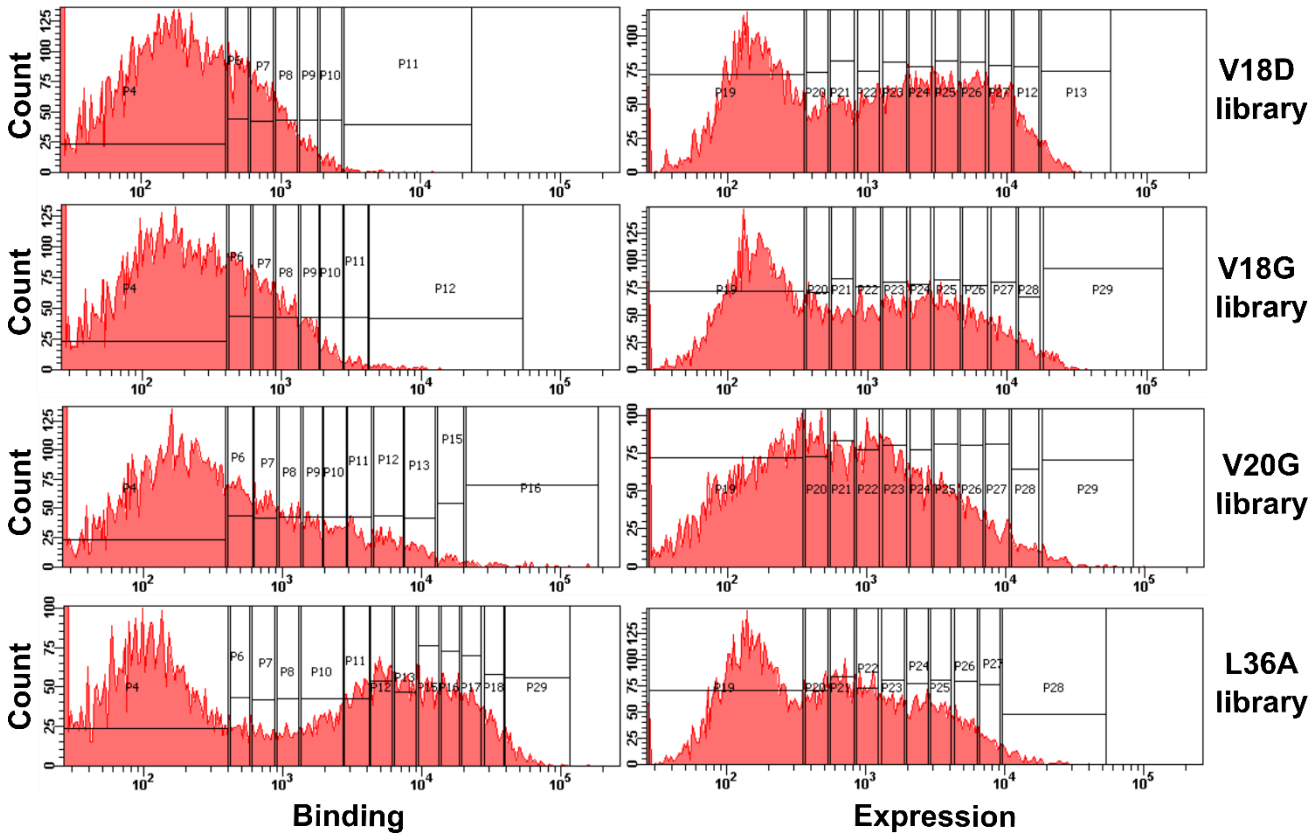
Specificity = $\frac{TN}{TN+FP}$

Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$

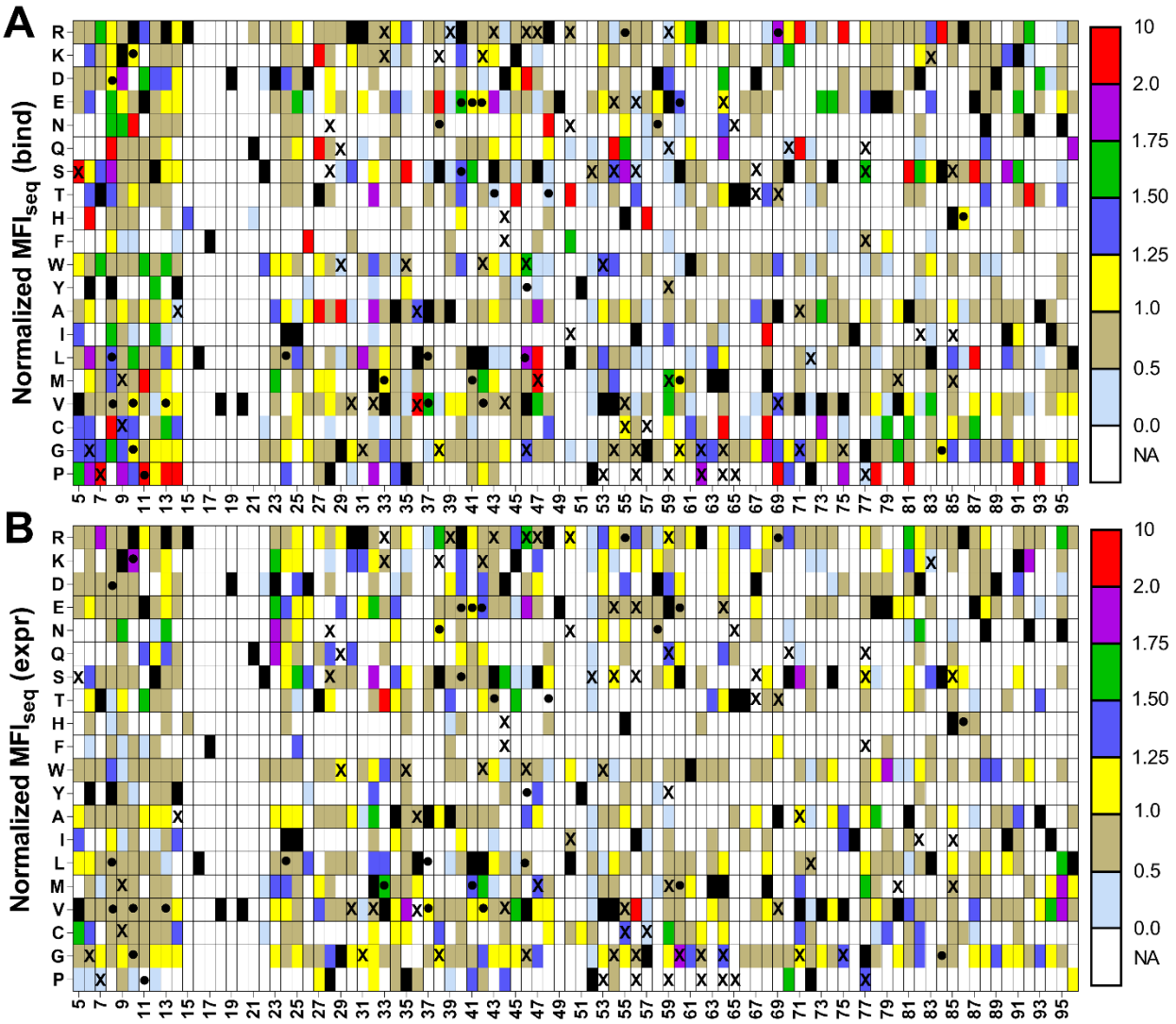
Supplementary Information



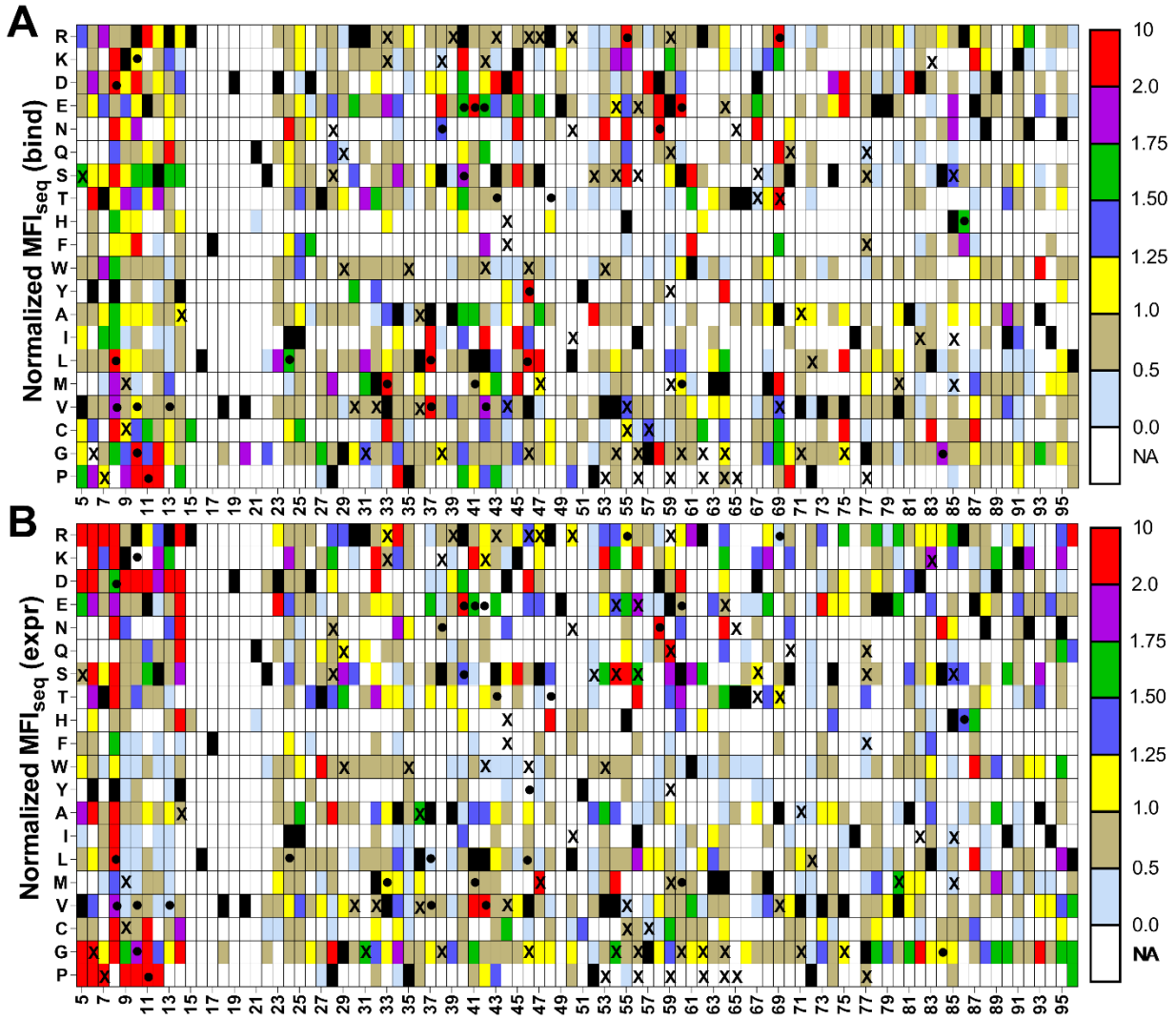
Supplementary Figure S1: FACS double plot showing the expression and binding to GyrA14 of CcdB mutant libraries in the background of different PIMs. FACS double plot of CcdB library containing PIMs (A) V18D, (B) V18G, (C) V20G and (D) L36A. Double plots of (E) CcdB single-site saturation mutagenesis library and (F) WT CcdB.



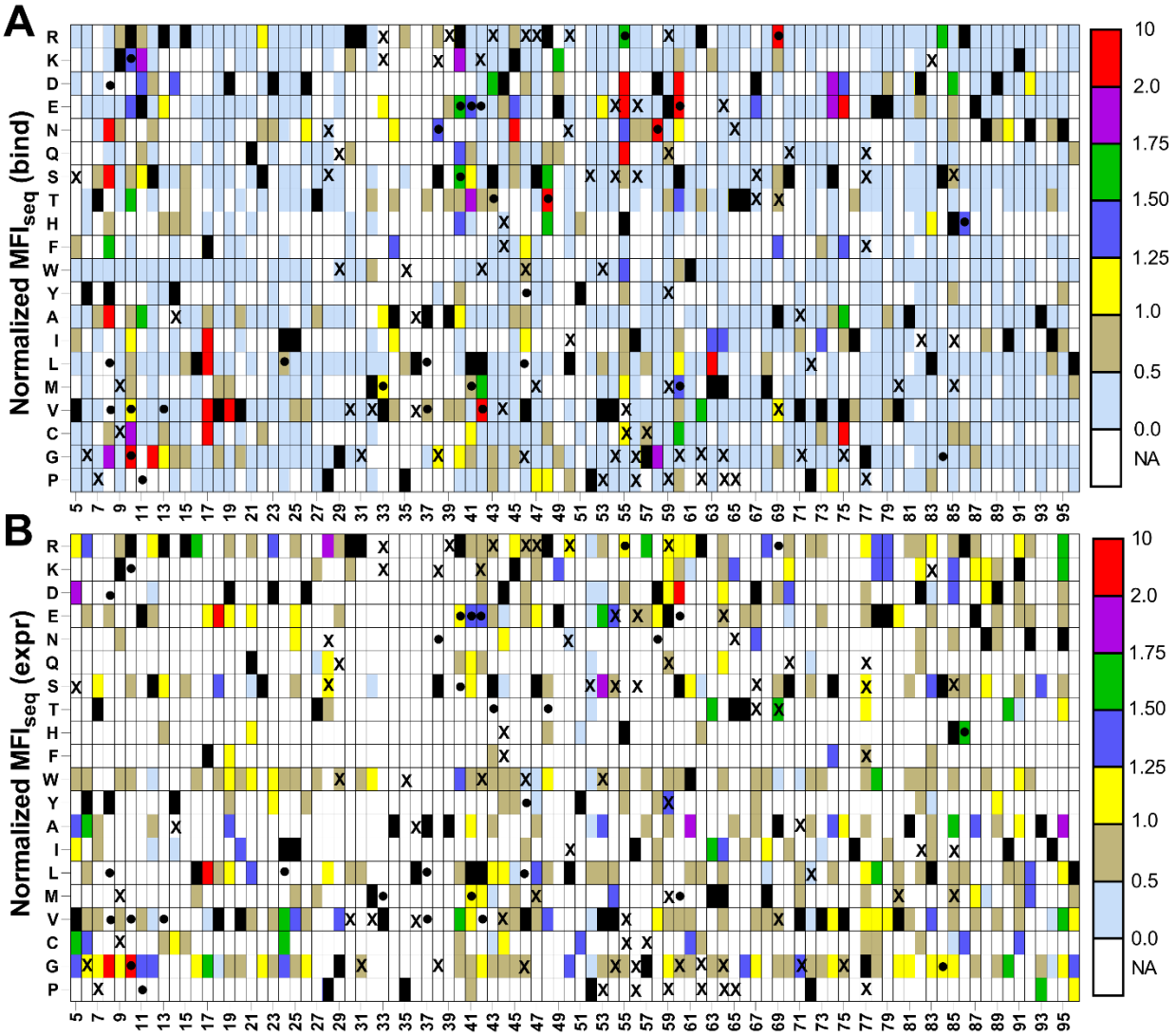
Supplementary Figure S2: Histograms of expression and binding to GyrA14 of PIM containing CcdB libraries. Left and right panels show binding and expression histograms. Different populations sorted into bins based on expression and binding are indicated by the vertical lines.



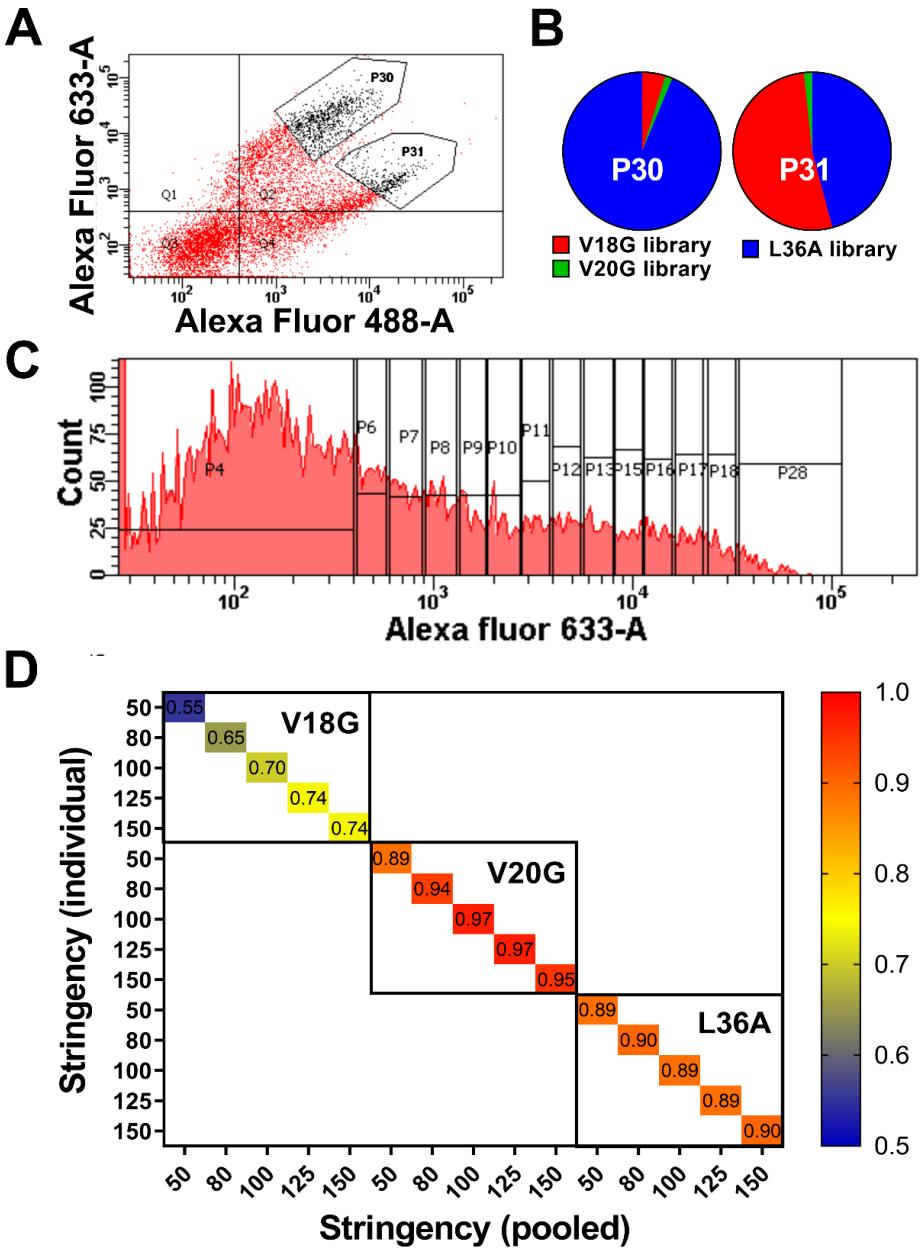
Supplementary Figure S3: Heat maps of normalized MFI_{seq} (bind) and MFI_{seq} (expr) for V18D library. Normalized (with respect to V18D) values of MFI_{seq} (bind) (A) and MFI_{seq} (expr) (B) were categorized in different ranges. Mutants with normalized MFI_{seq} (bind) or MFI_{seq} (expr) >1.25 were categorized as putative stabilized mutants. Black rectangle represents WT residue, mutants with no data available are indicated with a white rectangle. Several single mutants were characterized *in vitro* to estimate their stability, stabilized mutants having $\Delta T_m>1$ are indicated with a "●" and destabilized mutants $\Delta T_m<0$ are indicated with an "X".



Supplementary Figure S4: Heat maps of normalized MFI_{seq} (bind) and MFI_{seq} (expr). Normalized (with respect to V18G) values of MFI_{seq} (bind) (A) and MFI_{seq} (expr) (B) were categorized in different ranges. Mutants with normalized MFI_{seq} (bind) or MFI_{seq} (expr) >1.25 were categorized as putative stabilized mutants. Black rectangle represents WT residue, mutants with no data available are indicated with a white rectangle. Several single mutants were characterized *in vitro* to estimate their stability, stabilized mutants (ΔT_m>1) are indicated with a "●" and destabilized mutants with (ΔT_m<0) are indicated with an "X".



Supplementary Figure S5: Heat maps of normalized MFI_{seq} (bind) and MFI_{seq} (expr). Normalized (with respect to L36A) values of MFI_{seq} (bind) (A) and MFI_{seq} (expr) (B) were categorized in different ranges. Mutants with normalized MFI_{seq} (bind) or MFI_{seq} (expr) >1.25 were categorized as putative stabilized mutants. Black rectangle represents WT residue, mutants with no data available are indicated with a white rectangle. Several single mutants were characterized *in vitro* to estimate their stability, stabilized mutants (ΔT_m>1) are indicated with a "●" and destabilized mutants with (ΔT_m<0) are indicated with an "X".



Supplementary Figure S6: FACS of pooled libraries. (A) Dot plot showing the expression and binding of pooled library. Two different gates, P30 and P31 were used to sort the populations showing the highest expression and binding. (B) Pie chart of relative enrichment of mutants from each library after one round of sorting and deep sequencing in gates P30 and P31. (C) Sorting of pooled V18G, V20G and L36A library based on binding to GyrA14. (D) Heat map of correlation coefficient between the binding MFI of mutants calculated from an individual library and the pooled library at different stringencies, where stringency is the minimum number of reads per mutant.

Supplementary Table 1: Thermal stability of CcdB mutants estimated using TSA.

Mutant	ΔT_m	Mutant	ΔT_m	Mutant	ΔT_m	Mutant	ΔT_m
R62P	-24	D67S	-5	P28N	-1.5	Y8L	2
I56P	-17	I25A	-4.5	M32V	-1.5	R10V	2
P35W	-15.5	I25N	-4.5	V46G	-1.5	I24L	2
L50N	-15	S43Y	-4	G57C	-1.5	I24Y	2
V54S	-15	V46R	-4	E59Q	-1.5	V33M	2
P52S	-14.5	S47R	-4	A69V	-1.5	E11P	2.5
L50R	-14	V53W	-4	V75G	-1.5	R40E	2.5
T7P	-13	I25T	-3.5	S12A	-0.5	L42E	2.5
D82I	-13	S38K	-3.5	A37D	-0.5	S60M	2.5
D67T	-11.5	D44H	-3.5	A37Y	-0.5	L42V	3
L36V	-10	R86E	-3.5	S43A	-0.5	R48T	3
T65Q	-9.5	Y14A	-3	WT	0	101N	3
D44F	-7	L42K	-3	Y14W	0	A37V	3.5
M64P	-7	L50I	-3	S47M	0	H55R	3.5
G77P	-7	D67L	-3	K9R	0.5	S84G	3.5
D44V	-6.5	K9M	-2.5	K45Y	0.5	Y8D	4
P72L	-6.5	D26S	-2.5	V46W	0.5	I24K	4
R86D	-6.5	R30V	-2.5	R48G	0.5	V46L	4
A39R	-6	S38G	-2.5	A69T	0.5	R86H	4
E59P	-6	L42W	-2.5	R86K	0.5	S38N	4.5
V71A	-6	R62G	-2.5	Y8V	1	R40S	4.5
G77S	-6	H55C	-2	R13V	1	D58N	4.5
P28S	-5.5	E59R	-2	L41M	1	A69R	4.5
V53P	-5.5	S60G	-2	S43T	1	L41E	5
E59Y	-5.5	S70Q	-2	R10K	1.5	S60E	7.5
E79I	-5.5	K9C	-1.5	I24E	1.5	R10G	10
D26A	-5	R13F	-1.5	A37L	1.5		
P28Y	-5	D26V	-1.5	V46Y	1.5		