# Mutational scan inferred binding energetics and structure in intrinsically disordered protein CcdA

***Soumyanetra Chandra[1], Kavyashree Manjunath[2], Aparna Asok[1], Raghavan Varadarajan[1*]***

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

***[2] Institute for Stem Cell Science and Regenerative Medicine, Bangalore 560065, India***

***[*]Author for correspondence***

***Raghavan Varadarajan, Molecular Biophysics Unit, Indian Institute of Science, Bangalore-560012, India, Email : varadar@iisc.ac.in, Phone: +91-80-22932612, Fax: +91-80-23600535***

# Abstract

Unlike globular proteins, mutational effects on the function of Intrinsically Disordered Proteins (IDPs) are not well-studied. Deep Mutational Scanning of a yeast surface displayed mutant library yields insights into sequence-function relationships in the CcdA IDP. The approach enables facile prediction of interface residues and local structural signatures of the bound conformation. In contrast to previous titration-based approaches which use a number of ligand concentrations, we show that use of a single rationally chosen ligand concentration can provide quantitative estimates of relative binding constants for large numbers of protein variants. This is because the extended interface of IDP ensures that energetic effects of point mutations are spread over a much smaller range than for globular proteins. Our data also provides insights into the much-debated role of helicity and disorder in partner binding of IDPs. Based on this exhaustive mutational sensitivity dataset, a model was developed to predict mutational effects on binding affinity of IDPs that form alpha-helical structures upon binding.



**Keywords**: protein-protein interaction, saturation mutagenesis, single-site mutations, affinity prediction, secondary structure prediction, α-helix

# Introduction

Deciphering sequence function relationships in proteins has been one of the key agendas in biological research in the last few decades. Successful prediction of mutational effects on protein function has important implications for studying genetically driven human diseases as well as in protein engineering and design. The advent of Next Generation sequencing (NGS) and high throughput screening techniques, allowing simultaneous probing of large numbers of protein variants, has greatly facilitated investigations of genotype-phenotype relationships in proteins. Popularly dubbed as Deep Mutational Scanning (DMS) (Fowler and Fields, 2014), this high throughput approach commonly employs screening platforms such as cell based *in vivo* phenotypic assays (Adkar et al., 2012; Hietpas et al., 2011; Jiang et al., 2013; Melamed et al., 2013; Roscoe et al., 2013) or display platforms like yeast surface display (Boder and Wittrup, 1997; Heyne et al., 2020), phage display (Ernst et al., 2010; Ravn et al., 2013), mammalian surface display(Forsyth et al., 2013) and ribosome display (Larman et al., 2012). DMS methods simultaneously evaluate functional features of large libraries, by exploiting variant enrichment following selection, or distribution of the variants amongst different bins based on a functional attribute. Several groups have successfully studied mutational effects on protein-protein interactions (PPIs)(DeBartolo et al., 2012; Ernst et al., 2010; Fowler et al., 2010; Fujino et al., 2012; McLaughlin et al., 2012; Starita et al., 2013; Whitehead et al., 2012) using DMS and attempted epitope mapping(Van Blarcom et al., 2015; Datta et al., 2020; Doolan and Colby, 2015; Kowalsky et al., 2015; Mata-Fink et al., 2013; Najar et al., 2017) as well as quantifying interactions for large numbers of variants(Adams et al., 2016; Jenson et al., 2018).

There has been a plethora of studies employing DMS in globular and structured proteins. In contrast, how amino acid sequence specifies the disorder and yet codes for a biologically relevant function in Intrinsically Disordered Proteins (IDPs) still remains a puzzle. A large part of the genome encodes proteins with functionally important, intrinsically disordered regions (IDRs) (Uversky and Dunker, 2010). IDP(R)s lack well-folded structure in their free, native form. Mutations in the IDRs have been implicated in several human diseases including cancer and neurodegenerative ailments (Uversky, 2014; Uversky et al., 2008). The IDRs are often responsible for binding target proteins with high specificity, and are known to become structured upon binding in several cases (Dyson and Wright, 2005; Mittag et al., 2010).A large number of IDRs fold into helices upon binding, as evident from available structures of IDP-target complexes.

To probe residue specific contributions towards function of an intrinsically disordered protein we studied CcdA, a 72 residue protein antitoxin coded by the *ccdA* component of *ccdAB*, an *E. coli* F-plasmid borne type II toxin-antitoxin system (Miki et al., 1984). A cognate toxin protein, CcdB is expressed from the same operonic system. CcdB exerts its toxicity and mediates cell death by binding to bacterial DNA Gyrase as well as its corresponding DNA adduct (Bernard and Couturier, 1992; Critchlow et al., 1997; Dao-Thi et al., 2005), in the absence of the CcdA antitoxin. The antitoxin CcdA, when present, prevents cell death by binding tightly to CcdB and preventing its binding to Gyrase. CcdA can also rejuvenate Gyrase by extracting CcdB from both the binary CcdB-Gyrase and ternary CcdB-Gyrase-DNA complexes(Aghera et al., 2020; De Jonge et al., 2009; Maki et al., 1996). CcdA has a structured, DNA-binding, N-terminal domain (residues1-39) and an intrinsically disordered C-terminal domain (residues 40-72). The CcdA N-terminal domain is responsible for dimerization (Madl et al., 2006; Van Melderen et al., 1996)

and binds to the operator/promoter region of the *ccdAB* operon, playing an important role in autoregulation of *ccdAB* transcription (Afif et al., 2001; Dao-Thi et al., 2002; Vandervelde et al., 2017). The CcdA C-terminal domain, responsible for binding the cognate toxin CcdB, is intrinsically disordered and lacks structure in its free form(Bernard and Couturier, 1991; Dao-Thi et al., 2002; De Jonge et al., 2009; Madl et al., 2006). The available crystal structure of the CcdA C-terminus bound to CcdB dimer reveals that the CcdA C-terminal domain folds into a bent alpha helical region (residues 39-63) and an extended arm(residues 64-72), upon binding to CcdB (De Jonge et al., 2009). The availability of extensive structural information for CcdA in its bound form makes it an ideal model candidate for performing saturation mutagenesis, with the aim of understanding functional consequences of sequence manipulation in an IDP.

We present here a DMS study of CcdA, which explores mutational effects in an IDP on its interaction with a protein partner, using a facile binding energetics quantification methodology, BAMseq (Binding Affinity Measurement using deep sequencing). The BAMseq methodology integrates site-directed saturation mutagenesis and deep sequencing with Fluorescence Activated Cell Sorting (FACS) coupled to Yeast Surface Display (YSD). Unlike most other affinity measuring DMS methods, BAMseq uses the ratio of fluorescence signal of ligand binding to that of the surface expression of query protein, as a measure of ligand binding affinity. BAMseq allows parallel high-throughput estimation of binding affinity for diverse and numerous variants of a protein at a single ligand concentration, in contrast to titration based approaches which examine binding at multiple ligand concentrations (Adams et al., 2016). We successfully scored ~1290 single mutants of CcdA based on their relative binding affinity for CcdB. These scores helped elucidate residue specific contributions to CcdA function and identified the CcdA residues crucial for CcdB binding. We also deduced the apparent dissociation constant, $K_d$ and

change in Gibbs Free Energy of binding, $\Delta G°_{bind}$ from the DMS scores for all single-site CcdA mutants in the library. We successfully predicted local secondary structural signatures for the intrinsically disordered CcdA C-terminal domain in its bound form, solely from analysis of the mutational effects across the protein length. This suggests that DMS data can be used to improve or validate structural predictions for IDP bound conformations. This is important, because while there have been revolutionary advancements in deep learning based structural predictions of proteins (Baek et al., 2021; Jumper et al., 2021), IDP structural predictions by these methods have been deemed unreliable (Chakravarty and Porter, 2022; Ruff and Pappu, 2021; Strodel, 2021). Based on the mutational tolerance scores for CcdA single variants, we also developed an interaction weighted mutational penalty-based model, IPApred (Intrinsically disordered Protein Affinity prediction), to predict mutational effects on binding affinity for IDPs involved in PPIs. IPApred showed encouraging results when applied to prediction of mutational effects on partner binding in a number of IDPs.

# Results

### CcdA SSM Library Probed by FACS coupled to YSD

A single-site saturation mutagenesis (SSM) library of all possible single mutants of CcdA, where all 72 positions were individually mutated to all other amino acids, was constructed and cloned in the yeast surface display vector, pETcon (Addgene plasmid # 41522). The CcdA molecules were displayed on the yeast surface as a genetic fusion to Aga-2p, a yeast surface protein (Swers et al., 2004). The surface expression level of CcdA variants was monitored by the epitope tag c-myc fused to its C-terminus. CcdB binding was probed using biotinylated CcdB by FACS (Fig 1).We observed that single mutations in CcdA protein generally fail to completely abolish CcdB binding (Fig 2A-B). The available complex structure of the CcdA C-terminal domain bound to

CcdB (PDB ID:3G7Z) (De Jonge et al., 2009) (Fig 2C-D)shows an extensive interface surface area of ~2850Å$^2$ between the CcdA C-terminus and the CcdB dimer, explaining the nominal effects of CcdA single mutants on binding energetics.

**Scoring Mutants for their Binding Affinity and Surface Expression using FACS coupled to deep sequencing**

One way to quantify affinity is to monitor the diagonal axis of the Binding versus Expression FACS plot, sorting the library into parallel diagonal gates, as used in SORTCERY(Reich et al., 2015, 2016). Affinity quantification from the diagonal axis is both experimentally and mathematically intricate and is unsuitable for libraries where mutational effects are quite moderate, as seen in the case of CcdA (Fig 2A). Previously, binding affinity for a variant library was also quantified by sorting based on the binding fluorescence, at multiple target concentrations in a titration like strategy(Adams et al., 2016). The requirement of sorting into multiple gates at multiple target concentrations makes this method both labour intensive and expensive. Instead we used a simpler approach, measuring the cell count as a function of Binding:Expression ratio (Alexa 633:Alexa 488 fluorescence) at a single ligand concentration, where the x-axis (Binding:Expression ratio) can be taken as an approximate measure of the CcdB binding affinity of displayed CcdA molecules (Fig 2B). Sorting the population into vertical bins across this x-axis separates the library into compartments with varying degrees of binding affinity. After introduction of bin specific barcodes that provide bin identification, the *ccdA* amplicon DNA from all the sorted populations was pooled and subjected to deep sequencing (MacrogenNovaSeq6000). The raw read distribution across bins in the deep sequencing data was used to reconstruct the frequency distribution of each mutant across the vertical bins. We verified

for a small number of mutants, that the reconstituted frequency distribution obtained from deep sequencing data analysis resembles the individual FACS profiles of the respective single mutants (Supplementary Fig S1A-B).

We recently used a similar approach to estimate mean fluorescence intensity, MFI of CcdB mutants using YSD and showed that this simplified approach yields MFI values similar to more complicated Maximum Entropy based approaches (Ahmed et al., 2022).The MFI value of the frequency distribution across the Binding: Expression ratio axis is assigned as $MFI^{ratio}$ for ~1290 single mutants of CcdA (94 % of the total library diversity) (see Methods and Fig 1). The deep sequencing and $MFI^{ratio}$ assignment were carried out in two biological replicates, which were highly correlated (Supplementary FigS1C). The $MFI^{ratio}$ values were normalized with respect to the WT $MFI^{ratio}$ and averaged across the replicates to obtain the deep sequencing derived Mutational Tolerance Scores ($MTS_{seq}$) for CcdA mutants. The $MTS_{seq}$ score provides a measure of the CcdB relative binding affinity.

All mutations bearing synonymous codons with respect to the WT sequence were neutral in the YSD binding assay. We also found that mutants coded by degenerate codons had similar inferred relative binding affinity, indicating that the mutational effects on binding observed in surface displayed CcdA are solely due to the primary amino acid sequence and not the nucleic acid sequences. We also characterized surface expression of the CcdA library by sorting into bins, based on the fluorescence associated with the C-terminal c-myc tag (Alexa 488 fluorescence) (Supplementary Fig S1D). Expression scores obtained from deep sequencing for the CcdA mutants showed absence of significant mutational effects on surface expression (Supplementary Fig S2), unlike effects seen in case of a notable fraction of yeast surface displayed globular protein variants(Ahmed et al., 2021; Park et al., 2006). This suggests that sequence variation has

a minimal effect on IDP proteolysis, which plays an essential role in TA system activation under stress conditions. Unlike globular proteins, where mutations affecting protein structure and stability may lead to increased aggregation and/or proteolysis and therefore lower expression levels on the yeast surface, expression profiles for intrinsically disordered proteins appear to be minimally affected by mutations. Surprisingly, certain mutants at C-terminal domain positions (especially residues 55-59) show an elevated expression level on the yeast surface (Supplementary Fig S2), indicating that the WT sequence of this C-terminal domain stretch might be responsible for promoting disorder but is otherwise destabilizing, and is rescued by mutations.

**Experimental validation of BAMseq methodology**

We experimentally determined the dissociation constant, $K_d$for twenty individual single mutants and WT CcdA (Supplementary Figs S3A and S4) by CcdB titration of surface displayed CcdA molecules, using FACS. Surface displayed WT CcdA exhibited a $K_d$ value of 0.25$\pm$0.05 nM for CcdB binding (Supplementary Fig S3A and Supplementary Table S1), in our YSD system. It has been previously reported that the CcdB dimer has two non-identical sites that bind two molecules of CcdA with distinct binding affinities, one binding the full length C-terminal domain of CcdA with a $K_d$ of 20 pM, while another binds only residues 40-61 of another CcdA C terminal domain with a $K_d$ of 13 μM (De Jonge et al., 2009).However, recent Surface Plasmon Resonance (SPR) and fluorescence kinetic studies of CcdA-$CcdB_2$ interactions showed a $K_d$ value of ~0.3 nM for the high affinity site (Aghera et al., 2020) consistent with the present value. Since full length CcdA protein is intrinsically disordered, it is unstable and aggregation prone upon purification (Dao-Thi et al., 2002), which significantly complicates investigation of its interaction with partners using classical techniques like ITC (Isothermal Titration Calorimetry) or SPR. Instead,

the YSD based titration experiment used here serves as a facile approach to experimentally quantify CcdB binding affinity for a finite number of individual CcdA mutants. Several studies have previously established that binding constant values derived from YSD based FACS titration agree well with those obtained from classical methods like ITC, SPR and BLI as well as other techniques such as equilibrium competition titrations, fluorescence quenching titrations and stopped flow kinetics consistently(Bacon et al., 2020; Lipovsek et al., 2007; Rathore et al., 2018; Uchański et al., 2019).

The change in Gibbs free energy of CcdB binding ($\Delta G°_{bind}$) was calculated for each of the twenty CcdA mutants using the fitted $K_d$ values (Supplementary Table S1). These experimentally determined $\Delta G°_{bind}$ values correlate well with the DMS derived $MTS_{seq}$ scores (Pearson correlation, r= -0.84) (Supplementary Fig S3B) confirming the accuracy and precision of mutational scores derived from our BAMseq methodology. The experimental dissociation constant ($K_d$) values show a non-linear relationship to the $MTS_{seq}$ scores and fit better to an inverse first order equation (Supplementary Fig S3C), consistent with the ratio of binding to expression on yeast cell surface being inversely proportional to the $K_d$ values. Theoretically, binding: expression fluorescence signal is found to be proportional to $[C] / (K_d + [C] + [A])$(see Supplementary Methods), where $K_d$ = dissociation constant, [C] = CcdB (ligand) concentration used in the FACS experiment and [A]= concentration of free CcdA. Simulation of theoretical curves helps to determine the ideal ligand concentrations that can be used in BAMseq experiments for estimating binding affinities accurately (Supplementary Fig S5).

**Determination of apparent dissociation constant ($K_d$) for all single-site mutants of CcdA**

We further developed the BAMseq method to allow determination of binding constants from the DMS derived $MTS_{seq}$ scores, since the $MTS_{seq}$score only provides a relative measure of the CcdB binding affinities of CcdA mutants. The highly correlated $MTS_{seq}$ scores and experimental $\Delta G°_{bind}$ values of twenty experimentally validated mutants provide the necessary standard curve(Supplementary Fig S3B), facilitating extrapolation of apparent $\Delta G°_{bind}$ and apparent $K_d$ for mutants from the $MTS_{seq}$ scores. A similar method of extrapolation of binding energetics from DMS has been described previously(Jenson et al., 2018). Using this method, we could assign apparent $\Delta G°_{bind}$ and $K_d$ values for all ~1290 CcdA single mutants that lack prior experimental binding energetic information. The apparent $\Delta G°_{bind}{}^{seq}$ predicted by BAMseq show a normal distribution with a shoulder at the lower CcdB binding affinity range, indicating a distinct population of binding defective CcdA mutants (Supplementary Fig S3D).

**Mutational sensitivity landscape for CcdA confirms the redundancy of the N-terminal domain in CcdB binding activity**

In our CcdA DMS study, the N-terminal region (1-39) shows high tolerance to mutations, with $MTS_{seq}$ values close to that of WT CcdA, except for a few mutants that display modest impairment in CcdB binding (Fig 3A). Minor binding defects upon mutation occur primarily at buried positions (Fig 3A), when mapped onto the available DNA bound, CcdA dimer structure (Madl et al., 2006). This may result from disruption of the hydrophobic core upon mutations in the structured N-terminal domain, leading to the possible aggregation of the displayed full length CcdA protein on the yeast surface, consequently decreasing CcdB binding. It is noteworthy that these mutations show no distinguishable effects on surface expression (Supplementary Fig S2) indicating an absence of enhanced proteolysis or degradation of the variant protein on the yeast surface. Our DMS results clearly indicate that mutations in the N-terminal residues fail to impair

the CcdB binding activity and therefore can be inferred to have minimal contributions to the binding energetics or binding interface, consistent with the NMR and proteolysis studies of CcdB bound CcdA molecules, carried out previously (Madl et al., 2006). We also see no effects on CcdB binding of the displayed CcdA molecule upon mutations at positions involved in dimerization of the N-terminal domain (Madl et al., 2006; Van Melderen et al., 1996) (Fig 3A), confirming that dimerization of the N-terminus of CcdA is non-essential for CcdB binding.

The detrimental mutational effects on binding activity, as revealed by our DMS study of the CcdA, are restricted largely to the C-terminal intrinsically disordered domain (residues 40-72) and generally to the residues inferred to be CcdB interacting, from the available CcdA-CcdB complex structure(De Jonge et al., 2009) (Fig 3B). CcdA typically exhibits a high mutational tolerance throughout its length, barring a few positions, which can therefore be deduced as functionally important, CcdB binding residues.

**Prediction of binding interface from mutational landscape of an IDP**

To establish DMS as a viable method for prediction of interface residues in IDPs, we studied the relationship between overall mutational tolerance of CcdA C-terminal residues, with several structural parameters such as the change in Accessible Surface Area upon binding (ΔASA), residue depth (both commonly used to quantify residue specific contribution to interaction) or the number of bonds and interactions derived from the available complex structure of CcdA-CcdB (Supplementary Table S2). ΔASA is calculated by subtracting the accessible surface area (Lee and Richards, 1971) of CcdA in complex with CcdB from the accessible surface area of free structured CcdA. The Residue Depth is a measure of the depth of a residue from the exposed surface (Chakravarty and Varadarajan, 1999; Tan et al., 2011), with a higher depth in the bound

form signifying interaction with partner. The mutational tolerance averaged across all substitutions ($MTS_{seq}^{avg}$) of each residue position correlates fairly well with the respective ΔASA, number of interactions and residue depth (Fig 4A-C).

Functionally important residues in proteins are generally found to be conserved across species in order to maintain optimal activity. Surprisingly, the $MTS_{seq}^{avg}$ of CcdA C-terminal residues show only a modest correlation (r = -0.45) with the percent conservation of the *E. coli* WT residues in CcdA molecules across prokaryotes (Fig 4D, Supplementary Table S2). At some positions, mutations that are poorly tolerated in *E. coli* CcdA are found in other species. This might be due to co-evolution, with cognate CcdB harboring compensating mutations. In case of *E.coli* CcdB protein, CcdA interacting residues are also found to be poorly conserved across homologs.

Statistical k-means clustering of average mutational effects in CcdA into two classes, yields class-1 (similar to WT) and class-2 (lower binding affinity) having mean $MTS_{seq}$ values of 1.02 ± 0.152 and 0.49 ± 0.165 respectively. We categorize mutants having $MTS_{seq}$ lower than 0.9 (≈ $mean^{class2} + 2.5\ \sigma^{class2}$) as CcdB binding defective. A high cutoff value for this classification was chosen deliberately to allow identification of mutants with moderate binding defects, as mutational effects are observed to be typically mild in case of the CcdA library. CcdA residues can also be classified as CcdB interacting and non-interacting based on the ΔASA values obtained from the CcdA-CcdB structure (De Jonge et al., 2009). When compared to these structure based classes, the contributing and non-contributing classes of residues in CcdA identified solely based on a $MTS_{seq}^{avg}$ cutoff of 0.9 yields reasonable sensitivity (89%), specificity (80%) and accuracy (76%) calculated as described previously(Bhasin and Varadarajan, 2021; Tripathi et al., 2016) (Fig 4E-F).

While all the structure inferred non-interacting positions with the exception of Gly 63 (discussed later) were mutation tolerant, only 75% of the structure inferred CcdB interacting positions showed a significant decrease in CcdB binding activity upon mutation (Supplementary Fig S6).Despite the high residue burial upon binding and presence of contacts observed in the complex structure, these residues in CcdA exhibit high tolerance to mutations and thus minimal contribution to CcdB binding. CcdA mutational tolerance is non-uniform along its length. With the exception of W44, the 40-49 residue and the 60-64 residue stretches, as well as the C-terminal 71D and 72W residues in the CcdA C-terminal domain show high tolerance to mutations, irrespective of individual residue specific interactions with CcdB identified from the crystal structure (Supplementary Fig S6). This indicates that the residue contribution to CcdB binding in CcdA is higher at the interior of the interface and particularly at regions containing adjacently placed functionally important sites. Our data therefore suggests that in case of extensive interfaces as seen in CcdA, prediction of mutational effects solely from residue specific interactions derived from the crystal structure is likely to be inaccurate.

**Dissecting the plausible role of intrinsic disorder and helicity in CcdB binding of CcdA**

The CcdA C-terminal domain is intrinsically disordered and folds into a helical structure upon CcdB binding. We found that mutational tolerance in CcdA correlates negatively with predicted disorder (predicted using IUPred algorithm (Mészáros et al., 2018)) especially for mutations at CcdB interacting sites of CcdA (Supplementary Figure S7A-B). Intrinsic disorder might be essential for CcdA to maintain a disordered state, required for fast and efficient antitoxin degradation necessary in TA modules, and also to allow binding to multiple partners as observed in case of many IDPs. However, the present data indicates that increased disorder leads to impairment in binding affinity towards the partner as well as reduced activity. Similar findings

have also been reported in previous DMS studies on IDPs, namely α–synuclein and a transcription factor, Gcn4 (Newberry et al., 2020; Staller et al., 2018). Since many IDPs form helical structure upon partner binding, we studied the relationship between helix propensity and the observed effects of mutations. We observed that amino acid residues that promote helix formation are enriched in the known α-helical stretches in both the N- and C- terminal domains of CcdA (Supplementary Fig S7C). However, we found no apparent relationship between mutational effects on CcdB binding activity in CcdA variants and the relative helix propensity of the mutant amino acid residue with respect to that of the WT residue (Supplementary Fig S7D). If helix formation is a prerequisite for partner binding in IDPs, mutations that decrease helical propensity in these regions can be expected to hinder interactions with partner protein. However, in the case of CcdA the average mutational tolerance of each residue position in CcdA shows only a weak correlation with the helix propensity of the wildtype residues (Supplementary Fig S7E), indicating that partner binding and activity of CcdA do not depend strongly on inherent helix forming capability of the disordered regions.

**Interface Identification by Cysteine and Charged Scanning Methods**

A popular method of identifying residue contributions to partner binding in protein interfaces, is Alanine Scanning, where each residue is mutated to alanine to investigate the effect of mutations on protein-protein interactions (PPIs)(Cunningham and Wells, 1989). However Ala scanning performs poorly in many cases(Gray et al., 2017).Ser, Asp, Cys, Asn, His Scanning methodologies have been advocated as alternatives in identification of active site residues in globular proteins or in mapping epitopes (Das et al., 2020; Gray et al., 2017; Kanaya et al., 1990; Pál et al., 2005). We correlated the $MTS_{seq}$ patterns for each of the 20 substituted amino acids with the average residue $MTS_{seq}$ in CcdA. We found charged and cysteine mutants best

recapitulated the overall mutational tolerance of the different positions in the CcdA C-terminal domain (Supplementary Fig 8A). Further, the single mutant $MTS_{seq}$ scores for Cys and charged mutants correlate well with the residue depth of CcdA residues in the complex (Supplementary Fig 8B). While the charged substitutions possibly alter the primary interactions at functionally important residues, Cysteine, an aliphatic residue is not expected to alter partner binding. The disruptive effect of Cysteine substitutions predominantly at the interacting positions, might be due to its potential to form intermolecular disulfide bonds, in the oxidizing extracellular environment, masking the interacting face of the extended helical CcdB binding C-terminal domain of CcdA. In contrast, the Ala mutants produce relatively mild effects on binding affinity and fail to distinguish interacting from non-interacting residues in CcdA (Supplementary Fig S8).

**Secondary structural footprint deduced from mutational tolerance pattern**

The mutational sensitivity pattern of the CcdA C-terminus derived from BAMseq data shows an apparent oscillating pattern which is clearer for substitutions to charged residues, $MTS_{seq}$ (charged) (Fig 5A-B). It has been previously shown that mutational effects of charged residues on aggregation propensity of the membrane binding amphipathic helix α-synuclein resemble a waveform(Newberry et al., 2020). Though the intrinsically disordered CcdA C-terminal domain is not predicted to be amphipathic based on the hydrophobic moment calculated as described previously (Eisenberg, 1984) or according to estimations by the AMPHIPASEEK web server (Combet et al., 2000), we observe that the effects of charged substitutions on the binding activity in the CcdB interacting and non-interacting face of the CcdA C-terminal helix are clearly distinguishable. We therefore attempted to determine the periodicity in the mutational pattern and thus infer local structural features attained by the intrinsically disordered CcdA upon

binding. To correct for the unequal contribution to binding across the length of the CcdA protein, we subtracted the mutational tolerance averaged for charged substitutions, $MTS_{seq}$ (charged) at each position i by the same quantity averaged over a five residue sliding window centered around position i. We tested three, five, seven, and nine residue window averages and chose to use a five-residue window average, since it allows for clear detection of possible phase changes in the waveform. Identification of such discontinuity is essential in correctly predicting structures of IDPs in their partner bound form, where in many cases, the IDPs wrap around the globular partner protein upon binding, and there are drastic angular changes in the protein backbone.

When fitted to a single sinusoidal wave equation, the corrected $MTS_{seq}$ (charged) shows a poor fit. Upon identification of a possible phase change in the oscillating pattern at residue 63, we fit the corrected charged mutational tolerance values of the two residue segments to separate sinusoidal wave equations. The first 23 residue stretch of the C-terminal domain (residues 40-62) showed a reasonably good fit (R= 0.8) with a periodicity of 3.53$\pm$ 0.04, consistent with an α-helix containing 3.6 amino acids per turn (Fig 5C). However, the later 10 residue stretch (residues 63-72) fit better to a different waveform with periodicity of 4.1$\pm$ 0.2, with a good fit (R = 0.79), and a phase difference with the first waveform (Fig 5C). This indicates that the first 23 amino acids (region 40-62) form an α-helix while the next stretch does not, which is consistent with the available structure of CcdB bound CcdA C-terminal domain. Residues 40-62 form a bent helix, while residues 63-72 remain mostly unstructured with a central helical turn at residues 66-69 in the CcdB bound, crystal structure (De Jonge et al., 2009). The troughs in the waveform consist of residue positions 41, 42, 44, 45, 48, 51, 52, 55, 58, 59, 64, 65, 66 and 69 in CcdA, all of which (with the exception of 42) are known to interact with CcdB based on the available CcdA-CcdB crystal structure (De Jonge et al., 2009). This data can be used to further

validate or as an alternative approach to examining average mutational sensitivity as described in Figure 4. Only one face of the helix formed upon binding interacts with partner in IDP, resulting in a clear periodicity in the mutational effects. This periodicity when captured in terms of mutational scores allows facile identification of secondary structure and residues that constitute the binding interface. This demonstrates the potential of DMS for elucidation of local structural features of intrinsically disordered proteins and extended interfaces. The amplitude of the mutational tolerance waveform varies across the protein length, owing to differential contributions of individual residues to binding. The $MTS_{seq}$ scores for individual charged substitutions also show a similar oscillating pattern to the averaged values, (Supplementary Fig S9) and can be used to predict periodicity and structure.

**Optimal CcdB binding requires a bent alpha-helix**

While all other non-interacting residues are tolerant to mutations, the non-interacting residue Glycine 63 shows a binding defect upon most substitutions (Fig 3B). This is because Glycine 63 has a non-canonical backbone conformation in the bound state ($\Phi,\Psi$ = 143, 144) (Fig 5D), which is uncharacteristic of an α-helix, and unachievable by all other amino acid residues. G63 is responsible for breaking the helix (40-62) and allowing the subsequent unstructured tail (64-72) to bend and bind CcdB optimally (Fig 5E). The CcdB dimer is known to have two asymmetrical binding sites for CcdA, one where only the CcdA C-terminal stretch 40-63 binds with low affinity ($K_d$ = 13 μM) and the second, where the whole C-terminal 40-72 stretch binds with a high affinity ($K_d$ = 20 pM) (De Jonge et al., 2009). The 40-63 stretch as well as the 62-72 stretch have both been shown individually to bind the CcdB dimer with low affinity using ITC experiments(De Jonge et al., 2009). The Glycine residue is thus evidently responsible for providing a flexible hinge between the alpha helical region (39-63) and the extended arm (64-

72), a structural feature which appears to be crucial for high affinity CcdB binding. Upon analysis, we further identified similar glycine residues with positive Φ values of torsion angle in antitoxins of many type II TA systems (Table 1), that are positioned specifically to create a drastic angular change in the protein backbone molding the geometry of the typically extended interfaces involved in IDP antitoxins interacting with their cognate toxin partners.

**Performance of available computational predictors of binding affinity changes in mutants**

Several commonly used predictors of mutational effect on binding affinity, including BindProfX, PRODIGY, SAAMBE-3D etc fail to make predictions in cases like CcdA (interacting with a dimeric CcdB), where proteins with more than one protomer are involved in the interaction. The predictors mCSM-PPI (Rodrigues et al., 2019), BeAtMuSiC (Dehouck et al., 2013) and MutaBind (Li et al., 2016) do not have this limitation and showed modest correlations with Pearson correlation co-efficient (r) values of 0.49, 0.58 and 0.65 respectively with the experimental $\Delta\Delta G°_{bind}$ (= $\Delta G°_{bind}^{mut}$-$\Delta\Delta G°_{bind}^{WT}$) values for twenty single CcdA mutants (Supplementary Fig S10A-C). However, these predicted mutational effects are significantly larger than observed experimentally, for CcdA. The BAMseq derived apparent $\Delta G°_{bind}^{seq}$ deduced for all single mutants by BAMseq was compared to predictions by the available computational tool BeAtMuSiC, that predicts binding affinity changes upon all possible single mutations (Dehouck et al., 2013) and a modest correlation with r = 0.45 was observed (Supplementary Fig S10D). Comparison of apparent $\Delta\Delta G°_{bind}^{seq}$ deduced for all single CcdA mutants by BAMseq to the predicted $\Delta\Delta G°_{bind}$ values using BeAtMuSiC indicates that many positions with high mutational tolerance are erroneously predicted to have severely depleted binding affinity by BeAtMuSiC, and also the program fails to successfully predict effects in the most mutation sensitive positions (residues 64-67) of CcdA (Supplementary Fig S11). These

observations highlight the drawbacks of using the currently available predictors of mutational effects on energetics of PPIs in case of IDPs.

**Prediction of Mutant activity in IDPs based on DMS data**

In the present study, we have endeavored to delineate the contribution of every amino residue in the CcdA molecule to its CcdB binding activity. Based on the observed dependence ofmutational tolerance in CcdA on residue burial and physico-chemistry of substitutions, we have designed a predictive model, IPApred (Intrinsically disordered Protein Affinity prediction) (Fig 6). This method allows prediction of the functional effects of single mutations on the binding affinity, in terms of an arbitrary Mutational Tolerance Score ($MTS_{pred}$) for IDPs forming extended interfaces with partners. For each mutant, the $MTS_{pred}$ score is calculated by accounting for the type of substitution, and the residue burial upon binding of the WT residue. The amino acids can be classified into six categories namely charged, polar, aromatic, aliphatic, Proline and Glycine. We have found distinct tolerance patterns for substitutions across these categories in CcdA and calculated category penalties from the CcdA tolerance landscape to account for the effects due to changed chemistry upon mutation. Owing to the fair correlation between positional mutational tolerance and the residue depth upon binding, we have also introduced residue depth to weigh mutational effects at each position. This requires the availability of a complex structure of the interacting proteins. Owing to lack of prolines in the CcdA WT sequence, our model currently does not predict effects of substituting prolines. The IPApred was developed using a training set comprising 60% of the CcdA DMS dataset and was evaluated on a test set of the remaining 40% of the data. The predicted score, $MTS_{pred}$ correlated well (r= 0.65) with the experimental $MTS_{seq}$ for the CcdA mutants test set (Fig 7A). Based on the good correlation between the $\Delta MTS_{pred}$ ($MTS_{pred}^{mut} – MTS_{seq}^{WT}$) values from IPApredand the experimentally determined $\Delta\Delta G°_{bind}^{mut-WT}$

in the set of validated twenty CcdA mutants(Fig 7B), we further developed our approach to predict $\Delta\Delta G°_{bind}^{pred}$ both for CcdA and other systems. Our predictions for CcdA also correlated moderately with predictions by BeAtMuSiC server (Fig7C).

Validaion of our affinity prediction model was difficult owing to the lack of adequate and comprehensive literature on single site mutational effects on binding affinity in intrinsically disordered proteins. We therefore proceeded to test the performance of IPApred on the limited datasets of single site mutational effects of small numbers of alanine/glycine/valine substituted variants in disordered protein domains in humans, namely c-Myb(Giri et al., 2013), ACTR (transcriptional co-activator for thyroid hormone and retinoid receptors)(Dogan et al., 2013) , NCBD domain of CBP (CREB binding protein)(Dogan et al., 2013) and Hif 1α (hypoxia inducible factor 1α)(Lindström et al., 2018). Our predictions showed better correlation to the literature derived experimental data, relative to predictions of BeAtMuSiC (Fig 7E-L). In case of Hif 1α protein, the disordered C-terminal activation domain forms helical stretches interrupted by large unstructured regions upon binding to its partner, the Taz1 domain of CBP protein (Fig 7D). We noticed that while IPApred accurately predicted mutational effects in the helix forming regions of Hif 1α, our model performed poorly in case of substitutions at residues that constitute the unstructured regions (Fig 7G-H). While BeAtMuSiC excels at predicting binding affinity changes in structured globular proteins, this does not fare as well for PPIs in case of disordered proteins/regions. In case of ACTR, and NCBD experimental values correlated poorly with the predictions made by our method, IPApred (Supplementary Fig S12), as well as the web server, BeAtMuSiC. This is because of a single outlier in each case, namely the ACTR L32A mutation and the NCBD L77A mutations (Supplementary Fig S12). Upon removal of these two mutants, we find that the correlations improve significantly and that the IPApred methodology

consistently fares better than the BeAtMuSiC in these small datasets (Fig7G-H). However, CcdA DMS data driven IPApred isunable to capture stabilizing mutational effects on binding energetics, since CcdA single mutants rarely show better binding to partner relative to WT. We also observed that the quantitative effects of mutations on $\Delta\Delta G°_{bind}$ were heavily dependent on the length of IDP stretch and interface size of the PPI. Therefore, though MTSpred provides an accurate relative binding affinity estimate (relative to WT), the predicted $\Delta\Delta G°_{bind}$values may besmaller or larger than those observed experimentally for IDPs with larger or smaller interfaces respectively. More systematic studies of binding affinity in large variant libraries of various IDP systems are therefore necessary to understand effects of mutations and interface size on IDP binding energetics and to build more accurate prediction tools.

We also used IPApred to test mutational predictions on PPIs for all single-site mutants in three other intrinsically disordered proteins, human oncoprotein p53 and *Mtb* antitoxins, VapB5 and MazEmt9. These IPApred predicted scores were compared to the respective binding affinity predictions by BeAtMuSiC, in the light of absence of exhaustive experimental binding data for all single mutants of IDPs in the literature. Overall, IPApred prediction shows a fair correlation with the BeAtMuSiC predictions for these large number of variants (Supplementary Fig S13), indicating applicability of IPApred for prediction of binding energetic effects for all possible single-site mutants of IDPs.

## Discussion

Deep mutational scanning of a yeast surface displayed library of single-site CcdA variants reveal that mutations in itsN-terminal DNA-binding domain do not affect the CcdB binding activity, consistent with available structural and functional information. The C-terminal intrinsically

disordered domain is sensitive to mutations, especially at sites known to be involved in CcdB binding. The mutational effects at the CcdB interacting sites of CcdA are smaller than those commonly observed in globular proteins, likely due to the highly extended interface for the CcdA-CcdB complex that involves about ~32% of all CcdA residues contacting the CcdB dimer. In contrast, active sites of globular proteins typically encompass smaller interface areas and a significantly lower fraction of interacting residues, for example only ~8% of residues in the globular protein CcdB interact with its cellular target bacterial DNA-Gyrase. In globular proteins, interface (active-site) residues show high sensitivity to aliphatic mutants (Tripathi et al., 2016), which is not seen in case of CcdB-interacting positions in the IDP, CcdA C-terminal domain. CcdB-interacting residues of CcdA behave similarly to buried residues in globular proteins, tolerating aliphatic substitutions but not polar or charged mutations (Tripathi et al., 2016). This possibly originates from the largely hydrophobic nature of CcdA-CcdB interactions, which account for nearly 50% of the ΔASA (De Jonge et al., 2009). However, unlike buried sites in globular proteins which are all typically buried deep in the hydrophobic core, interacting residues in IDP interfaces undergo transient burial upon interaction, residing at low residue depth from the complex surface. This expectedly produces a milder energetic effect of single substitutions in disordered CcdA including at the interacting positions. At positions harboring wildtype, aliphatic or aromatic residues in CcdA, charged substitutions are poorly tolerated. At wildtype charged and polar residues, the tolerance pattern generally lacks any apparent amino acid preference. We also observed no prominent effect of the size of substitutions on the CcdB-binding affinity in CcdA.

Despite the high tolerance to mutations in CcdA, our methodology BAMseq is able to detect small changes in binding affinity in a high-throughput manner by integrating YSD coupled

FACS methodology to deep sequencing. Previous YSD FACS based method to estimate binding affinity such as Tite-seq, involves experiments at multiple (~11) ligand concentrations(Adams et al., 2016). In contrast, BAMseq involves sorting the library at a single ligand concentration into bins across a binding : expression axis. Carrying out DMS based binding analysis at multiple concentrations may facilitate accurate measurement of binding energetics in globular proteins where single-site mutations can have large effects on binding affinity. In contrast, in cases like CcdA where $K_d$ values for single-site mutants vary over a relatively smaller range of values (0.1-7 nM for CcdA mutants) experiments at a single ligand concentration suffice for accurate measurement of binding affinity to target protein. Based on analysis of theoretical curves of binding : expression fluorescence intensity as a function of Gibbs free energy of binding (and dissociation constant) (Supplementary Fig S5), we found that CcdB (ligand) concentrations in the range of 2-10 nM are suitable for the single ligand concentration approach used in $BAM_{seq}$ to parallelly estimate binding affinities of CcdA mutants. A similar approach can be applied to select a single or minimal number of ligand concentrations for $K_d$ estimation in other systems..

A mutational tolerance score ($MTS_{seq}$) calculated for each single mutant describes the mutational effects on IDP-target interactions accurately and elucidates sequence–function relationships in IDPs. Further, this scores also allows parallel and high throughput quantification of binding energetics of all protein variants using an internal standard. While wild type CcdA shows a $\Delta G°_{bind}$ value of -13.08 $\pm$ 0.05 kcal/mole, $\Delta G°_{bind}$ for the CcdA single mutants vary over a range of -11 to -14 kcal/mole.In case of mutations with very low surface expression, the Binding:Expression fluorescence parameter fails to represent the actual affinity to target. While some mutations in globular proteins are known to affect surface expression in YSD studies, we observed no such effects on expression in case of IDP CcdA mutants. Unlike the well-studied

mutational effects on protein folding and stability in globular proteins, we find no significant mutational effects on stability, proteolysis and/or levels of protein in the IDP, CcdA. This is evident from our exhaustive screening of the expression levels of CcdA mutants relative to WT, on the yeast surface.

The averaged $MTS_{seq}$ score (across all substitutions) for each residue position provides a good estimate of the residue's contribution to CcdB binding energetics. We developed a method for identification of functionally important residues in IDPs using the deep mutational scan information, which shows high accuracy, sensitivity and specificity. Surprisingly, evolutionary conservation shows poor correlation with CcdA mutational effects, indicating the potential disadvantages in predicting functionally important residues from evolutionary information in case of disordered antitoxins and other similar protein segments. While there is an overall good agreement between experimental mutational tolerance and the residue's interactions in the complex, there are a few residues that show extensive residue burial in the complex but are surprisingly tolerant to mutations. These are found either close to the N or C-terminus, or at stretches which have a small number of contiguous interacting positions across the primary sequence. Our findings highlight the caveats associated with assuming all interactions identified from the crystal structure are functionally important, especially in case of high affinity interactions involving large interfaces.

Generally, proline residues are found to be abundant in intrinsically disordered proteins (Mateos et al., 2020). The CcdA C-terminus however lacks any wild type proline residue. CcdB binding was observed to decrease upon introduction of proline at most C-terminal domain positions (excepting the 42, 50 and C terminus 70-72 residue positions), including the CcdB non-interacting sites (Fig 3). Proline is known to disrupt helices except when located at the first three

N-terminal positions(Bajaj et al., 2007; Kim and Kang, 1999). This is owing to its lack of the amide hydrogen that forms an essential intra-molecular H bond in helices with the i-4 C=O group. The decreased activity of proline mutants indicates that CcdB binding of CcdA, though mediated by an intrinsically disordered domain, may require structuring into α-helical structures upon binding. A similar disruption in IDP function due to proline substitutions throughout the protein length, has also been reported in an amphiphilic helix forming IDP, α-synuclein (Newberry et al., 2020). However, we find poor correlation between helix propensity and observed mutational tolerance in CcdA. This is consistent with previous findings refuting the pre-requisite of an intrinsic helical tendency for optimal partner binding in IDPs(Rogers et al., 2014). Our results suggest that though the intrinsically disordered CcdA C-terminal domain may not require a transiently populated helical structure in the free form to bind CcdB, permanent disruption in the helical structure formed upon binding by introduction of backbone-modulating proline residues may impede CcdB interaction. We also suggest that though the disorder encoded into the primary sequence of an IDP might be essential for functions like prompt bio-removal, promiscuous binding etc., its contribution to partner binding is minimal.

Low mutational tolerance of the CcdB non-interacting Glycine 63 residue of CcdA to all substitutions helps to identify it as an interface shaping residues. This can be rationalized by the positive Φ value of backbone conformation at this position, which allows the helix (40-62) in WT CcdA to break and the 64-72 extended arm to wrap around and bind the CcdB dimer. While BAMseq successfully identifies the previously unknown, functionally important Glycine 63 residue, it fails to recognize it as a non-interacting residue, lacking physical contact with CcdB. We have located several glycine residues with similar positive Φ torsional angle and putative backbone bending roles, widespread in several type II antitoxins (Table 1).

Despite immense advances in protein structural modeling for globular proteins, prediction of partner bound structure and function of IDPs remains a major challenge (Baek et al., 2021; Chandra et al., 2021; Jumper et al., 2021). In the current study we inferred that the CcdA C-terminal region 40-62 forms a helical structure, based on the ~3.6 residue periodicity observed in the tolerance pattern of charged substitutions. While the mutational pattern of the 63-72 residue stretch also shows periodicity, the periodicity of ~4.1 however suggests a predominantly non-helical stretch. Since the mutations in the N-terminal domain of CcdA show no effect on CcdB binding affinity, BAMseq could shed no light on its local structural features (Supplementary Fig S9A). Based on the mutational data, we suggest that both Cysteine scanning and charged residue scanning mutagenesis might be useful in identifying functionally important positions at extended interfaces, and predicting structure attained upon binding in IDPs (Supplementary Figs S8 and S9). Such Single Residue Scanning methods are comparatively cheaper and simpler than full blown saturation mutagenesis studies. Cys mutants produce relatively smaller effects than charged substitutions making the latter more suitable for rapid identification of interface or secondary structure in IDPs. We have recently carried out charged scanning mutagenesis of another antitoxin, MazE6 from *Mycobacterium tuberculosis* to predict functional residues and structural features(Chandra et al., 2021), though in the absence of an experimental structure, the predictions remain to be validated.

Based on the available exhaustive mutational landscape of CcdB binding to the CcdA protein, we developed a prediction method, IPApred to predict mutational effects on partner binding for single mutations in IDPs. IPApred requires as input an IDP-target complex structure to calculate the residue depth of WT residues. We applied IPApred to a number of IDP domains that fold to predominantly α-helical structure upon partner binding and have extended interfaces. We

validated our predictions using experimental binding affinities and compared performance of IPApred to that BeAtMuSiC server for a handful of IDPs where structures are known and experimental binding energetics are available for sets of single-mutant variants. We found IPApred to consistently perform better than BeAtMuSiC in case of these IDRs that attain helical structure upon target binding. Since mutational effects on IDP-target binding affinity appear to be dependent on the size of the binding interface, a method like IPApred that provides arbitrary predictive scores (like $MTS_{pred}$) is a useful approach. The MTSpred scores can be converted to $\Delta\Delta G°$ values by using a small experimental dataset of single mutants of the IDP system involved as an internal standard to obtain a mutational prediction on binding energetics for any single-site mutant. Using CcdA experimental affinity values for the same purpose may produce inaccurate quantitative estimates of $\Delta\Delta G°$ values in cases of some IDPs.

IPApred thus provides arbitrary scores depicting relative binding affinity and apparent $\Delta\Delta G°_{bind}$ upon mutation, calculated using a small set of experimental binding affinity measurements for the system of interest. This helps in accounting for the differences in the extent of mutational effects on binding affinity in IDPs with interfaces of various dimensions. Currently IPApred does not correct for the non-uniform mutational tolerance pattern observed in case of CcdA. We believe predictions can be significantly improved by incorporating high throughput mutational data on binding effects in other IDPs, but such data are currently unavailable.

The CcdA mutational study along with investigations of mutational effects in other disordered antitoxin systems (Chandra et al., 2021) indicate that the severity of mutational effects on partner binding is highly dependent of the size of the interaction interface and strength of the binding affinity in the WT IDP. Studies of more IDP-partner interactions will help develop more accurate models to predict the severity of mutational effects on binding affinity. The current results also

highlight how mutational effects on IDP function differ both qualitatively and quantitatively from observed and well documented sequence-function relationships in well-folded, globular proteins. Further DMS studies to elucidate sequence-activity relationships in other IDPs will help to decipher the role of protein disorder in PPIs and to develop improved methodology for prediction of mutational effects on binding and function in IDPs. Structure prediction in proteins with few homologs, which is common in IDPs, remains challenging since even recently developed, highly successful deep learning based methods (Baek et al., 2021; Jumper et al., 2021) work best for multiple sequence alignments, containing large numbers of sequences. Our IDP affinity prediction method in combination with deep learning based predictions of protein-protein complex structure (Jumper et al., 2021; Mirdita et al., 2022) can be a powerful tool to decipher the functional effects of mutations in intrinsically disordered functional domains in proteins, where structural and/or functional information is often largely unavailable.

## Materials and Methods

Brief descriptions are provided below, more details are available in the Supplemental Methods Section

**Cloning and CcdA library preparation**

Wildtype and single mutants of CcdA were cloned into the YSD plasmid vector, pETcon (Addgene plasmid # 41522), using two or three fragment homologous recombination(Swers et al., 2004) in yeast *S. cerevisiae* EBY100 strain employing a high efficiency LiAc/ssDNA/PEG chemical transformation method (Gietz and Schiestl, 2007). For preparation of the single-site

saturation mutagenesis library of CcdA in pETcon vector, we used primers with NNK codons, and the cloning was carried out using homologous recombination in yeast.

**CcdB Purification and Biotinylation**

CcdB was purified using an immobilized CcdA peptide affinity matrix(Tripathi et al., 2016). The purified CcdB protein was then biotinylated using EZ-Link Sulfo-NHS-SS Biotin (Thermo-Fisher). 100 μL of 170 μMCcdB was incubated at 4°C for 2 hours in 1X PBS reaction buffer with 2 mM biotin. The non-reacted biotin was then removed by passing the biotinylated sample through a PD SpinTrap G-25 column (Sigma-Aldrich) with a molecular weight cutoff of 5000 Da. ESI-MS of the biotinylated CcdB sample revealed that the majority of CcdB molecules were biotinylated at 1-3 sites. Biotinylated CcdB was quantified by absorbance measurements and the concentration was verified by Tricine-SDS PAGE using known concentrations of lysozyme protein as a standard.

**Yeast Surface Expression and FACS sorting of CcdA Library**

CcdA was expressed and displayed on the yeast surface(Chao et al., 2006). Upon induction, the CcdA displayed on yeast cells was labeled with anti-myc primary antibody and biotinylated CcdB (2nM for sorting and varying concentrations between 0.01pM and 200nM for titration experiments) followed by suitable fluorophore tagged secondary antibodies to probe surface expression and CcdB binding of surface displayed CcdA molecules. To estimate binding affinity, the binding:expression signal was used for sorting the library into vertical gates (bins). To additionally quantify the expression levels, the library was also sorted based on expression signals.

Yeast plasmid was purified from each gated population. The CcdA gene was amplified with primers containing gate specific barcodes before pooling and subjecting to deep sequencing at Illumina NovaSeq6000 150 PE Platform (Macrogen, South Korea)

**Determination of Mutational Tolerance Score ($MTS_{seq}$) for each mutant**

NGS data for the CcdA mutant library was processed using an in-house developed pipeline (see Supplementary Methods Section). The number of reads in each gate was obtained for each single mutant after data analysis which was then further normalized before the real abundance of mutants across gate is represented. The read numbers were thus converted to frequency distributions of mutants across gates. (see Supplementary Methods for details on normalization).

The $MFI^{ratio}$or the mean binding: expression ratio fluorescence intensity for a mutant was then calculated as follows:

$$MFI^{ratio}_{x} = \Sigma_i[\left(D_i^x\right) * F_i] \qquad \text{Equation 1}$$

where, x and i represents mutant and gate identity respectively. $F_i$ is the mean fluorescence for gate i and $D_i^x$ is the normalized fraction of a mutant x in gate i, accurately representing the frequency distribution of mutants across gates (see Supplementary Methods).

The$MFI^{ratio}$ was calculated for mutants with >100 total reads and averaged across the replicates. The averaged $MFI^{ratio}$ normalized with respect to WT $MFI^{ratio}$ is then assigned as the $MTS_{seq}$(Mutational Tolerance Score derived from deep sequencing) to each mutant. Therefore,

$$MTS_{seq} = MFI^{ratio} \text{ of mutant} / MFI^{ratio} \text{ of WT} \qquad \text{Equation 2}$$

The $MTS_{seq}$ values depicted in the heatmaps are the $MTS_{seq}$ values averaged over replicates 1 and 2 to maximize the information obtainable from the dataset. The $MTS_{seq}$ values used in correlation studies are $MTS_{seq}$ values of Replicate 1, taken from an individual biological experiment and analysis to investigate the performance of single experiment measurements in predicting functional residues etc. The same correlations were also evaluated for $MTS_{seq}$ values of Replicate 2 and showed comparable Pearson correlation coefficients.

Using the same approach, we also calculated the Expression Scores ($ES_{seq}$) for all single mutants of CcdA from the raw mutant reads obtained in the gates P11-P15 (Supplementary Fig S1D).

**Determination of $K_d$ and $\Delta G°_{bind}$ for surface displayed WT CcdA and single mutants**

WT and single mutants of CcdA were cloned in the pETcon vector in *S. cerevisiae* EBY100 strain. The CcdA molecules were expressed, displayed and probed for CcdB binding with a range of concentrations (0.01pM - 200nM) of biotinylated CcdB and Streptavidin conjugated AlexaFluor-633 secondary antibody (1:2000). The mean Alexa-633 (binding) fluorescence was then fitted using SigmaPlot 14, to a one site ligand binding equation :

$y = x\, B_{max} / (x + K_d)$  Equation 3

where $B_{max}$ is the saturated Fluorescence Intensity and $K_d$ is the dissociation constant of binding (Rathore et al., 2018). The Gibbs free energy of binding ($\Delta G°_{bind}$) was calculated using the relationship $\Delta G°_{bind} = RT \ln K_d$, where R (ideal gas constant) = $1.9872 * 10^{-3}$kcal $K^{-1}$ $mole^{-1}$ and T (temperature) = 298.15 K. The YSD titration estimated $K_d$ and $\Delta G°_{bind}$ values of CcdB binding for WT and mutant CcdA molecules are denoted as experimentally derived binding energetic parameters ($K_d^{exp}$ and $\Delta G°_{bind}^{exp}$ ) in this work.

**Calculation of apparent $\Delta G°_{bind}$ using $MTS_{seq}$ values**

Based on the observed high correlation (r = -0.84) between the experimentally determined $\Delta G°_{bind}$ values of CcdA molecules (WT and single mutants) and the corresponding deep sequencing derived $MTS_{seq}$ values, we acquired a standard curve describing this relationship. The data points of $\Delta G°_{bind}^{exp}$versus $MTS_{seq}$ axis plot were fit to a straight line equation.This was used as an internal standard to calculate the apparent $\Delta G°_{bind}^{seq}$ from obtained from deep sequencing for mutants (with unknown experimental $K_d$ value) using the measured $MTS_{seq}$ values as follows.

$\Delta G°_{bind}^{seq} = - 2.04 * MTS_{seq} - 11.0035$ Equation 4

**IPApred method development for prediction of mutant binding affinity changes**

For the purpose of designing a predictive model, we used 60% of the CcdA mutational data (comprising of mutant data for the C-terminal domain, residues 40-72) as the training set, and the remaining 40% as the test set to validate the predictive models (Fig 6). The amino acids other than proline and glycine were classified into four categories namely charged, polar, aromatic and aliphatic. Besides these, proline and glycine were considered as individual classes. We calculated category penalties from the CcdA mutational tolerance landscape, to account for the effects due to changed chemistry upon mutation (see Supplementary Methods).We also tested and observed that incorporation of a residue specific penalty (expected to account for residue features like size of the functional side chains)into our predictive model did not improve the predictions. The residue specific penalty was therefore not used in the final model. The good correlation between structure derived residue depth and positional mutational tolerance was used to derive an internal

standard curve to estimate the overall sensitivity to mutations for a residue position (positional $MTS_{pred}$) from the residue depth parameter obtained from the protein-partner complex structure.

positional $MTS_{pred}$ = 1.315 – 0.0934 * residue depth Equation 5

To distinguish between partner interacting and non-interacting residues and to improve the predictor, mutational penalties were weighted by a scaling factor related to the residue depth parameter. After exploring various normalizations of the residue depth, we found the best results using the normalization described below.

The normalized residue depth of position j ,

$nRD^j = 2\ RD^j / RD^{max}$ Equation 6

Where $RD^j$ is residue depth of a particular residue j in the IDP, and $RD^{max}$ is the maximum residue depth observed in the same IDP segment. The normalization factor of 2 was found to perform best by parametric fitting of equation 6 below for the training set.A number of variations of the model were rigorously tested on the 40% CcdA DMS data (test set). The best performing mathematical formula for prediction of effect of an (a→b) category of mutation at position j, which was finally used in the current work is as follows:

Predicted score $^{j,(a\rightarrow b)}$= positional $MTS_{pred}^j$ + $nRD^j$* Category $Penalty^{(a\rightarrow b)}$ Equation 7

where Category $penalty^{(a\rightarrow b)}$ = category averaged $MTS_{seq}^{(a\rightarrow b)}$- WT $MTS_{seq}$

= category averaged $MTS_{seq}^{(a\rightarrow b)}$ - 1 Equation 8

category averaged $MTS_{seq}$ is the averaged $MTS_{seq}$ value for all substitutions of a→b category in the training set.

The predicted score is then corrected based on its observed correlation with the experimental DMS derived $MTS_{seq}$ values for the training set mutants to obtain the predicted mutational tolerance scores, MTSpred (Fig 6).

$$MTSpred^{j,(a\rightarrow b)} = 0.45 + \text{Predicted score}^{j,(a\rightarrow b)} * 0.62 \qquad \text{Equation 9}$$

MTSpred values are calculated and validated for single-site mutants in CcdA test set as well as in other IDP systems (Fig 6).

**<u>Prediction of $\Delta\Delta G°_{bind}$ values from predicted Mutational Tolerance Scores</u>**

Based on the good agreement between $\Delta MTS_{pred}$ ($MTS_{pred}^{mutant}$ - $MTS_{pred}^{WT}$) andthe experimentally determined $\Delta\Delta G°_{bind}$ values for single-site CcdA mutants, we used this relationship as a standard curve, to derive $\Delta\Delta G_{pred}$ values from the predicted $MTS_{pred}$ values as follows:

$$\Delta\Delta G_{pred}^{j,(a\rightarrow b)} = -1.31 * \Delta MTS_{pred}^{j,(a\rightarrow b)} - 0.01 \qquad \text{Equation 10}$$

**Author Contributions**: R.V and S.C. designed the experiments and wrote the manuscript. K. M. processed the deep sequencing data. A.A helped in acquiring FACS data. S.C. performed all experiments and analyzed the data.

**Acknowledgements:** SC acknowledges Ministry of Human Resource Development for her fellowships and thanks all the members of the RV lab for their valuable suggestions. We thank Munmun Bhasin for the useful discussions on protein modelling. 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.

**Data and Material availability:** The data relevant to the figures in the paper have been made available within the article and in the Supplementary Information section. All materials generated in this study are available from the lead contact (varadar@iisc.ac.in) without restriction.

**Conflict of interest statement:** The authors declare no conflict of interest.

**Funding:** This work was funded in part by a grant to RV from the Department of Biotechnology, grant number-BT/COE/34/SP15219/2015, DT.20/11/2015), Government of India, Science and Engineering Research Board (SERB), Government of India (grant number EMR/2017/0040S4 and SR/S2/JCB-10/2007 (JC Bose Fellowship)) and grants from DST and MHRD. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

# Figures

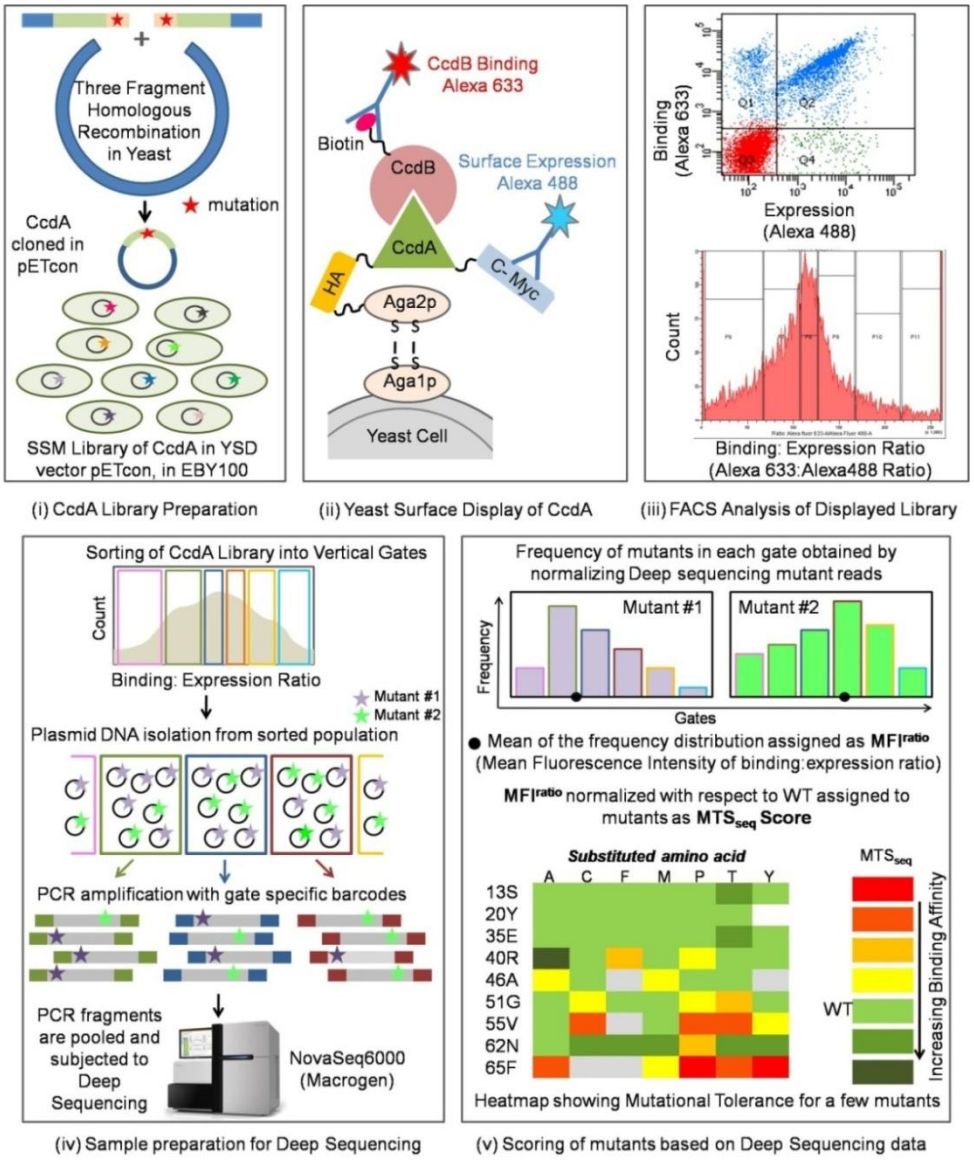


**Figure 1. Detailed workflow of deep mutational scanning employed in BAMseq.**A single-site saturation mutagenesis library of CcdA in yeast surface display vector pETcon was constructed using three fragment homologous recombination in yeast strain, *S. cerevisiae*. This library was then displayed on the yeast surface as a fusion to the Aga2p surface protein and probed by FACS for its surface expression as well as binding to its target partner, CcdB. The library of CcdA

variants is then sorted into vertical bins across an axis representing CcdB binding affinity. After introducing bin specific barcodes, the mutants from different bins are pooled and subjected to deep sequencing. After normalisation of the obtained raw data representing the mutant distribution across the affinity bins, a mutational tolerance score ($MTS_{seq}$) is assigned to each mutant (see Methods), which allows both ranking of mutants based on affinity as well as undestanding residue specific contributions to the binding enegetics.

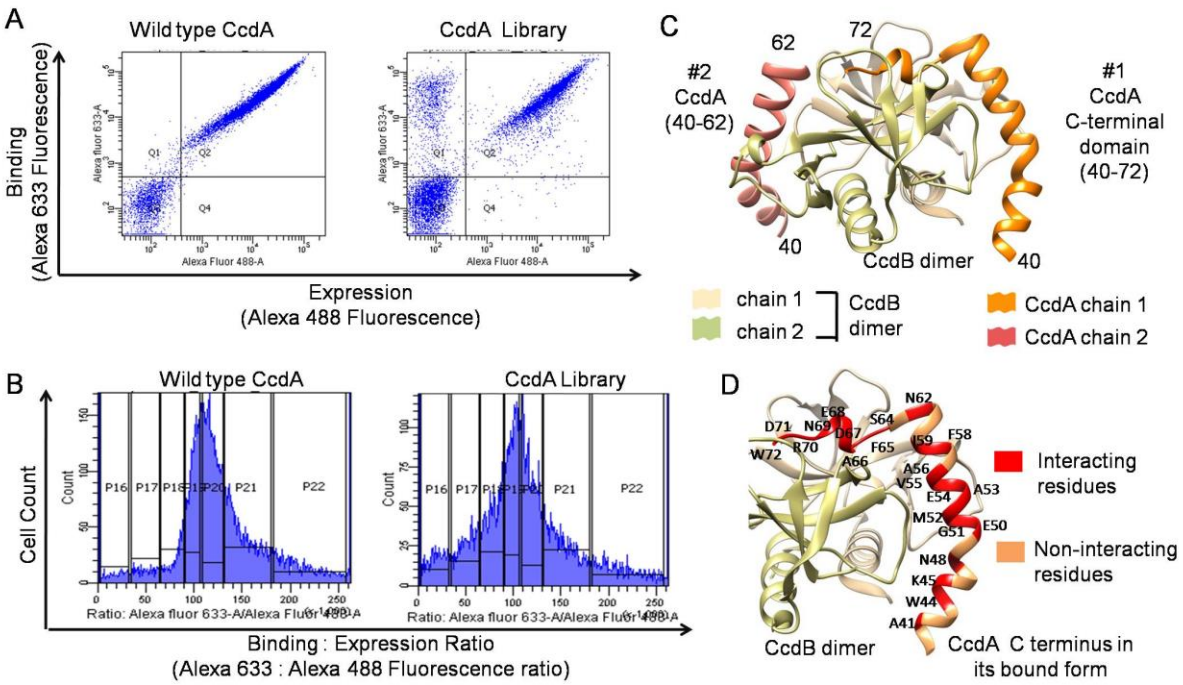


**Figure 2. CcdB binding activity of WT and mutant CcdA**. (A) FACS analysis of the surface displayed CcdA WT (left) and SSM Library (right) showing CcdB binding versus surface expression plots. The surface expression and CcdB binding of CcdA are monitored by Alexa-488 and Alexa-633 fluorescence intensity respectively. Most CcdA variants have similar binding affinity to WT as evident from the similarity of the two plots. (B) Cell count as a function of CcdB Binding : Expression Fluorescence Intensity ratio for CcdA Wild type and SSM library. This allows for sorting based on relative affinity towards CcdB. The CcdA library was sorted into seven vertical bins, P16-P22. (C) The crystal structure of the intrinsically disordered C-terminal domain of CcdA in complex with the CcdB dimer (PDB ID : 3G7Z). The CcdB dimer has two asymmetric binding sites for CcdA with differing affinity, one binding the full 40-72 residue C-terminal domain with high affinity (CcdA chain 1) and the other binding to only the 40-62 stretch of the C-terminal domain of a second CcdA molecule with lower affinity (CcdA chain 1). (D) A closer view of the high affinity binding site between the CcdA C-terminal

domain and the CcdB dimer, with CcdB interacting and non-interacting residues, as inferred from the crystal structure, highlighted in red and peach respectively. The interacting residues are labeled (WT residue identity and position).

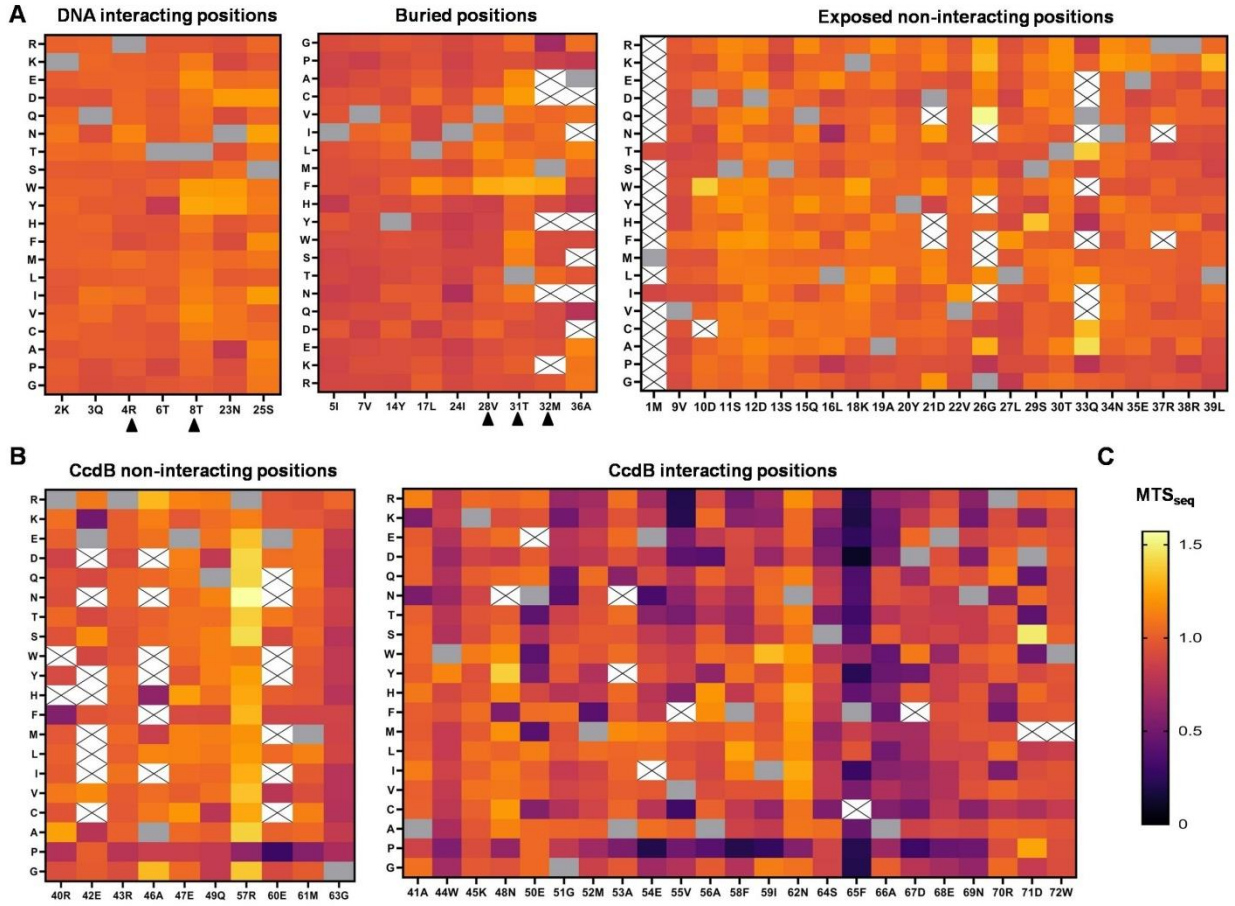


**Figure 3. Mutational tolerance of CcdA single mutants.** Mutational tolerance scores for (A) the structured N-terminal DNA binding domain of CcdA and (B) the intrinsically disordered, CcdB binding C-terminal domain of CcdA. The N-terminal domain residues are classified as DNA interacting, buried and non-interacting sites based on the DNA bound dimeric CcdA N-

terminal domain structure (PDB ID :2H3C)(Madl et al., 2006). The positions known to be involved in the N-terminal domain dimerization in the DNA bound CcdA dimer structure, are marked with black triangles. Residues in the C-terminal disordered domain of CcdA are classified as CcdB interacting and non-interacting based on the CcdAB complex structure (PDB ID :3G7Z). (C) The color key used depicts the ranges of $MTS_{seq}$ scores representing CcdB affinity. WT CcdA has $MTS_{seq}$ score of 1. The poor binders with lower tolerance scores are shown in dark colors, while lighter colors depict marginally better binders. Grey cells and blank cells (marked with X) indicate WT residue and mutants with no available data, respectively.

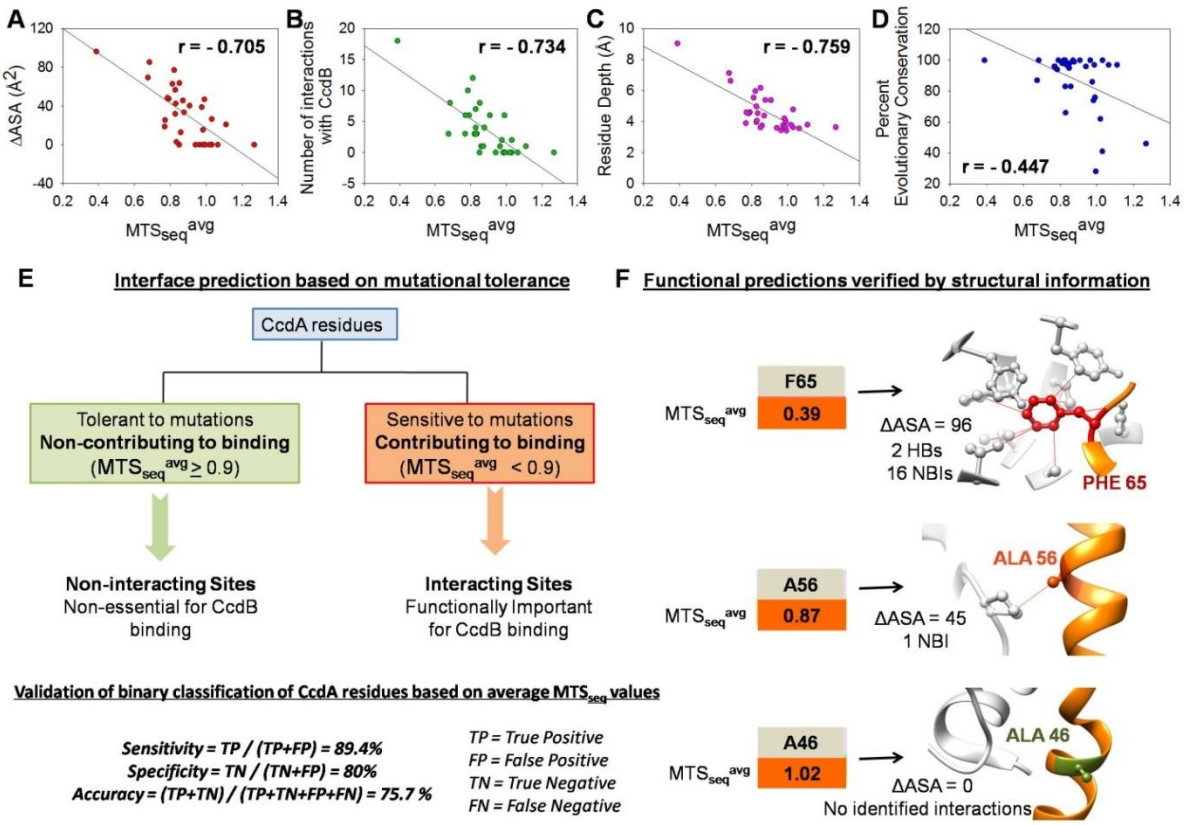


**Figure 4. Prediction of interface residues from Deep Mutational Scanning of CcdA.** Correlations of the average mutational tolerance scores ($MTS_{seq}^{avg}$) for residues in the CcdA C-terminal domain with (A) ΔASA (B) number of interactions each position is involved in with CcdB (C) Residue Depth respectively, calculated using the CcdAB complex structure (PDB ID: 3G7Z) and (D) the percent evolutionary conservation at each position across bacterial species. r is the Pearson Correlation Coefficient. The $MTS_{seq}^{avg}$ scores were calculated by averaging the $MTS_{seq}$ for all substitutions at each position and depict the average effect of mutations at each position. (E) Flowchart for prediction of functionally important residues in CcdA from the positional $MTS_{seq}^{avg}$ scores (F) Validation of interface predictions by structural analysis. HB and NBI refers to number of H-bonds and non-bonded interactions respectively that the WT residue forms with CcdB (grey) in the crystal structure of the complex.

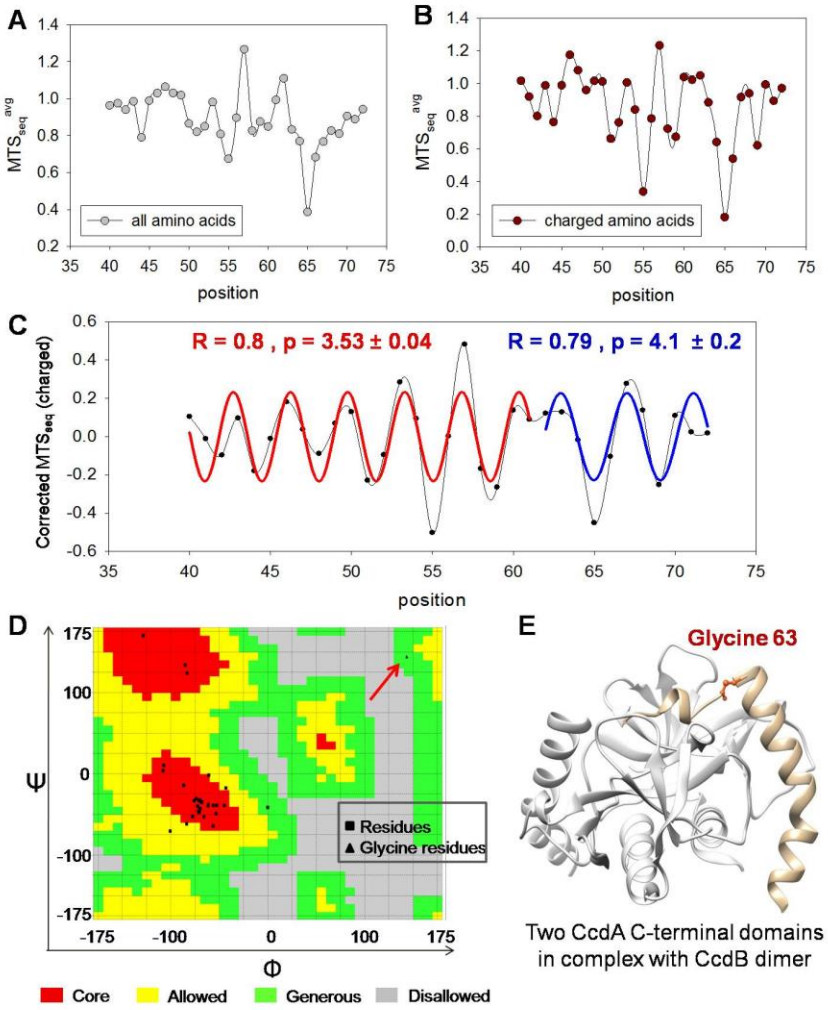


**Figure 5. Local structural insights from CcdA C-terminal mutational data.** (A) The mutational tolerance score averaged for all substitutions ($MTS_{seq}^{avg}$) as a function of residue position in the CcdA C-terminal domain (plotted as a spline curve) displays an oscillating pattern. (B) The mutational tolerance score averaged for charged substitutions, $MTS_{seq}$(charged) as a function of residue position. (C) Corrected mutational tolerance for charged substitutions plotted against position across the length of the CcdA C-terminal domain. The y axis is the difference between averaged mutational tolerance for charged substitutions for each position and the mutational tolerance for charged substitutions averaged over a five residue sliding window

[i.e. corrected $MTS_{seq}$(charged) = $MTS_{seq}$(charged) – average $MTS_{seq}$(charged) for a sliding window of five residues], to correct for the uneven distribution of mutational effects across the length of the C-terminal domain. The data shows periodicity and is fitted to a single sinusoidal wave form, $y=a \sin(2\pi x/b+c)$, where a = amplitude, b = periodicity and c = phase. A possible phase change is apparent at residue 63. The two C-terminal domain stretches separated by residue 63, when fitted to two separate wave equations (fitted curves shown in red and blue), show distinct periodicities of 3.54 and 4.1 in the 40-62 region and the 63-72 region respectively. (D) Ramachandran map for all C-terminal domain residues. The arrow shows the G63 torsional angle values highlighted with an arrow. (E) Back bone conformation of Glycine 63 mapped on the structure of the C-terminal domain of CcdA bound to CcdB dimer (PDB ID: 3G7Z).

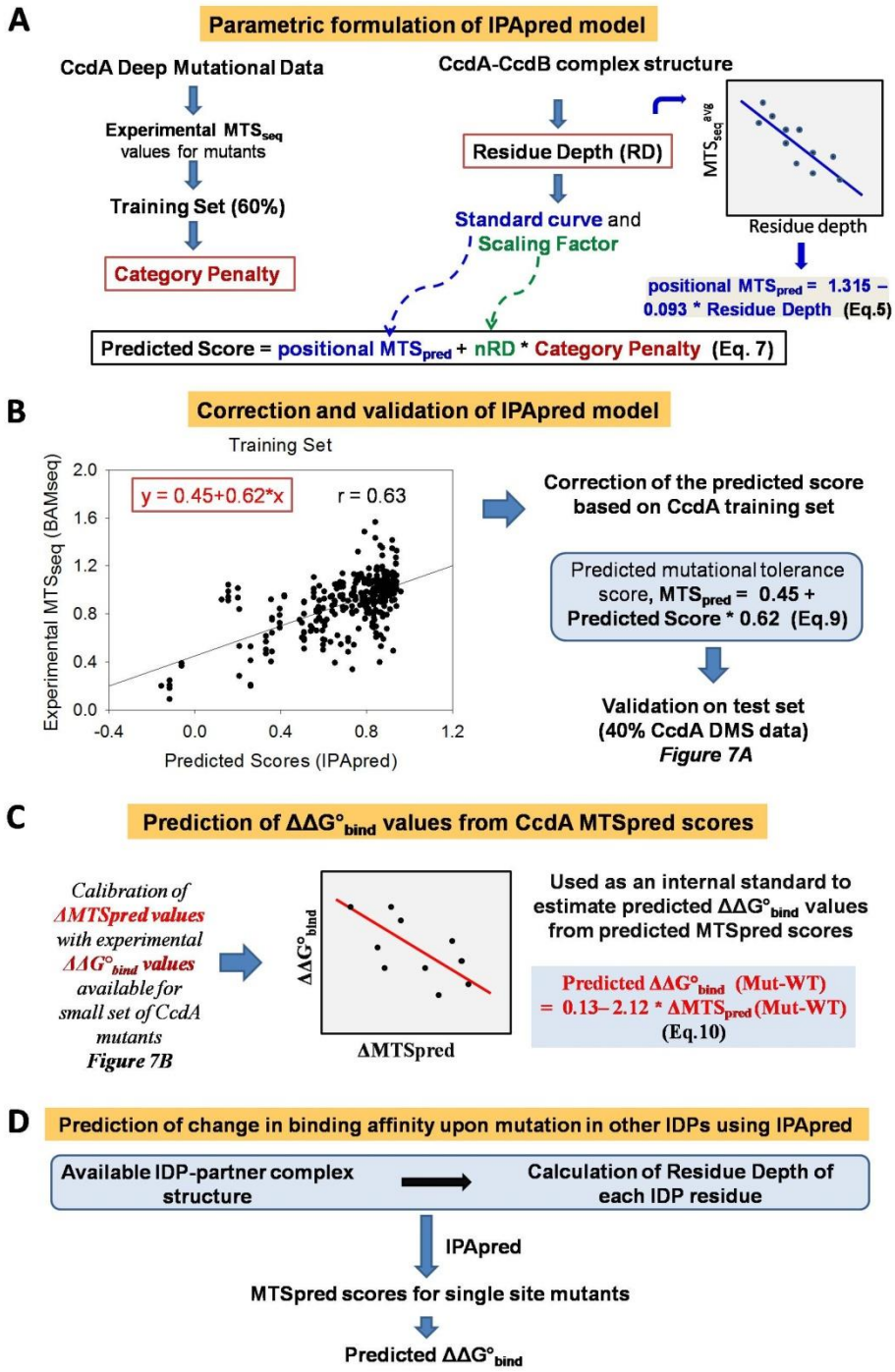


**Figure 6. IPApred method development for prediction of binding affinity changes in IDP.** (A) The category penalties for each of the 29 categories of substitution, and residue depth (Å) are derived from the CcdA DMS training set and available CcdAB complex structure (PDB id :

3G7Z) respectively. A mathematical formula (equation 7) for estimating relative binding affinities (predicted score) is optimized for best performance using the training set. This uses the positional MTSpred (obtained from dependence of experimental DMS scores on residue depth in CcdA) and category penalty weighted by normalized residue depth. (B) The predicted scores are corrected based on the observed correlation of predicted and available experimental scores in the training set to derive MTSpred which is the final predicted mutational score (equation 9). The method is then validated on the CcdA DMS test set shown in Figure 7A. (C) The linear fit to experimental $\Delta\Delta G°$ binding of twenty individual mutants as a function of $\Delta$MTSpred (MTSpred (mut) – MTSpred(WT)) is used to further estimate the predicted $\Delta\Delta G°_{bind}$ of mutants from MTSpred values (equation 10). (D) Finally the model is applied to other IDP systems. Residue depth is calculated from available IDP-partner structure and category penalties and corrections derived from CcdA are used to predict MTSpred values using the same equations (Eq. 9) as used for CcdA (A-C). The $\Delta\Delta G°_{bind}$ values are then predicted using these calculated MTSpred scores (Eq. 10).

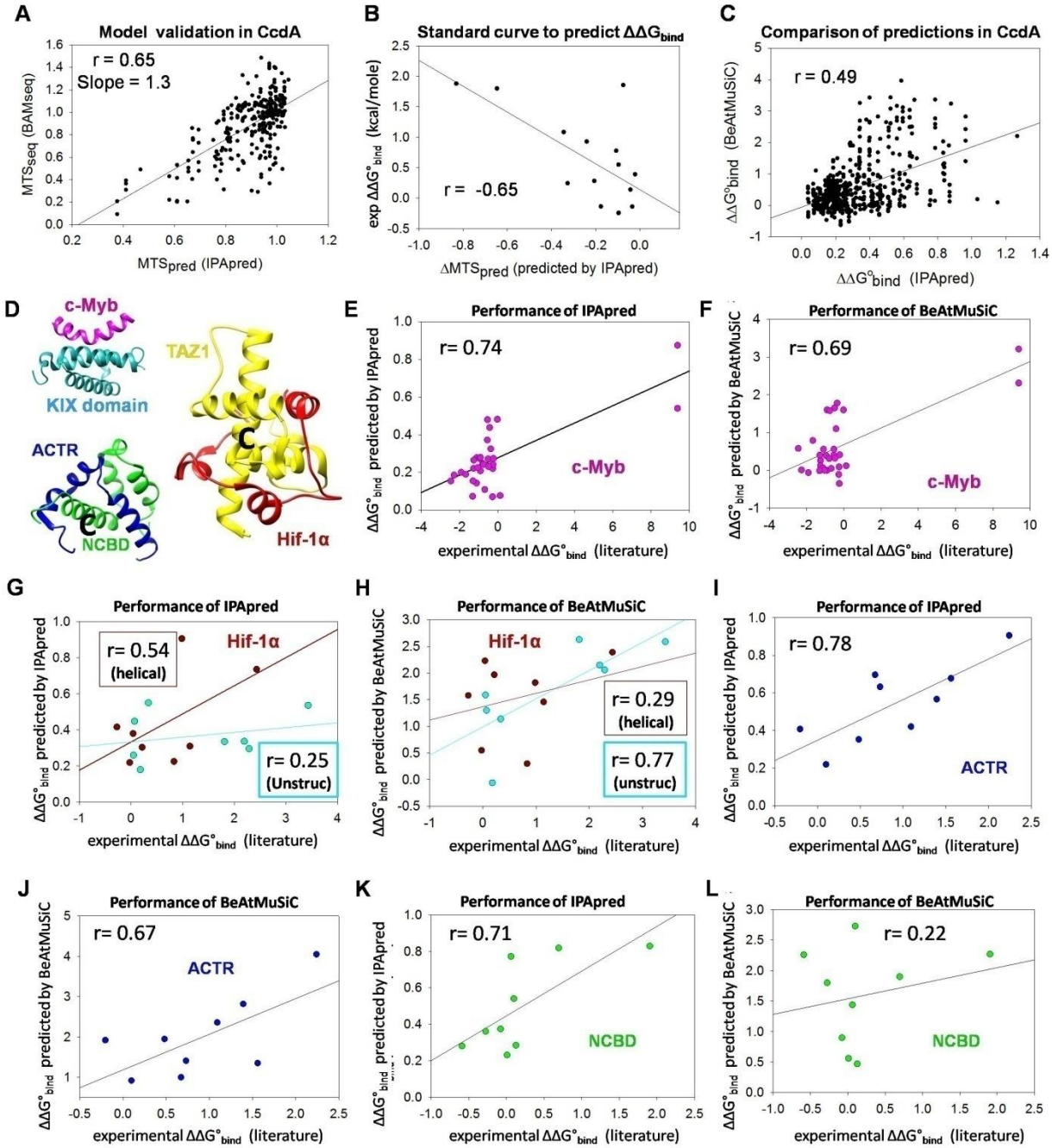


**Figure 7. Correlations of experimentally determined mutational effects on binding affinity with predictions by IPApred.** (A) Correlation between the mutational tolerance scores predicted by IPApred ($MTS_{pred}$) and the experimental mutational tolerance scores ($MTS_{seq}$) obtained from the DMS study for the test set of CcdA C-terminal domain mutants. (B)

Correlation of $\Delta MTS_{pred}$ ($MTS_{pred}^{mutant}$ - $MTS_{pred}^{WT}$) with $\Delta\Delta G°_{bind}$ values determined experimentally for 20 single-site CcdA mutants. The fitted straight line (where y is $\Delta\Delta G°_{bind}$ and x is $\Delta MTS_{pred}$) is further used as an internal standard to derive predicted $\Delta\Delta G°_{bind}$ values from $MTS_{pred}$ scores in other systems. (C) Correlation of the $\Delta\Delta G°_{bind}$ values predicted by IPApred for CcdA C-terminal domain mutants in the test set with the corresponding $\Delta\Delta G°_{bind}$ values predicted by BeAtMuSiC for the same mutants. (D) The available structures of intrinsically disordered transactivation domain of c-Myb protein interacting with the KIX domain of CBP protein (PDB id: 1SBO), intrinsically disordered ACTR domain of p160 protein bound to the molten globule like NCBD domain of CBP protein both of which fold into helices upon binding to each other (PDB id: 1KBH), and intrinsically disordered CAD domain of Hif-1α protein interacting with the TAZ domain of CBP protein (PDB id: 1L8C). (E,G,I and K) Correlation of the $\Delta\Delta G°_{bind}$ values predicted by IPApred for disordered protein domains, with available experimentally determined $\Delta\Delta G°_{bind}$ values for corresponding single mutants. (F,H,J and L) Correlation of the $\Delta\Delta G°_{bind}$ values for disordered protein domains predicted by BeAtMuSiC with available experimentally determined $\Delta\Delta G°_{bind}$ values for corresponding single mutants In case of Hif-1α, the mutant dataset was split based on their positions in the helical (brown) or unstructured (blue) stretches in the Hif-1α CAD domain (G-H). The IPApred method appears to perform better for helical stretches than unstructured regions, formed upon TAZ binding in Hif-1α. For ACTR and NCBD, the correlation of predicted and experimental $\Delta\Delta G°_{bind}$ values after removal of a single outlier in each case (indicated in Supplementary Fig S12) is shown (I-L).

## Tables

**Table 1. Predominance of glycine residues with positive torsion angles in bacterial type II antitoxins**

| | Residue/ Position[a] | Antitoxin | Complex PDB ID | Chain | TA module | Organism | Φ torsion angle (°) [b] |
|---|---|---|---|---|---|---|---|
| 1 | Gly 25 | Maz E | 6A6X | C | MazEF mt9 | *Mycobacterium tuberculosis* | +71 |
| 2 | **Gly 76** | Maz E | 7D2P | E | MazEF | *Deinococcus radiodurans* | +47 |
| 3 | **Gly 138** | Maz E | 6L2A | F | MazEF 4 | *Mycobacterium tuberculosis* | +133 |
| 4 | Gly 35 | Vap B | 6SD6 | A | VapBC | *Shigella sonnei* | +68 |
| 5 | **Gly 44** | Vap B | 6SD6 | A | VapBC | *Shigella sonnei* | +68 |
| 6 | **Gly 67** | Vap B | 3DBO | A | VapBC 5 | *Mycobacterium tuberculosis* | +88 |
| 7 | Gly 9 | Vap B | 6NKL | C | VapBC 5 | *Haemophilus influenzae* | +65 |
| 8 | Gly 35 | Vap B | 6NKL | C | VapBC 5 | *Haemophilus influenzae* | +91 |
| 9 | **Gly 55** | Vap B | 6NKL | C | VapBC 5 | *Haemophilus influenzae* | +117 |
| 10 | Gly 35 | Vap B | 3TND | B | VapBC | *Shigella flexneri* | +68 |
| 11 | **Gly 44** | Vap B | 3TND | B | VapBC | *Shigella flexneri* | +99 |

[a] Glycine residues in the C-terminal toxin binding domain of the antitoxins are highlighted in bold.

[b] Φ refers to the C-N-Cα-C backbone torsion angle for the Glycine residue calculated from complex structures using PDB Goodies (Hussain et al., 2002).

# References

Adams, R.M., Mora, T., Walczak, A.M., Kinney, J.B., and Supe, N. (2016). Measuring the sequence-affinity landscape of antibodies with massively parallel titration curves. 1–27. https://doi.org/10.7554/eLife.23156.

Adkar, B. V, Tripathi, A., Sahoo, A., Bajaj, K., Goswami, D., Chakrabarti, P., Swarnkar, M.K., Gokhale, R.S., and Varadarajan, R. (2012). Protein model discrimination using mutational sensitivity derived from deep sequencing. Structure *20*, 371–381. https://doi.org/10.1016/j.str.2011.11.021.

Afif, H., Allali, N., Couturier, M., and Van Melderen, L. (2001). The ratio between CcdA and CcdB modulates the transcriptional repression of the ccd poison-antidote system. Mol. Microbiol. *41*, 73–82. https://doi.org/10.1046/j.1365-2958.2001.02492.x.

Aghera, N.K., Prabha, J., Tandon, H., Chattopadhyay, G., Vishwanath, S., Srinivasan, N., and Varadarajan, R. (2020). Mechanism of CcdA-Mediated Rejuvenation of DNA Gyrase. Structure *28*, 562-572.e4. https://doi.org/10.1016/j.str.2020.03.006.

Ahmed, S., Manjunath, K., and Varadarajan, R. (2021). Prediction of residue-specific contributions to binding and thermal stability using yeast surface display. BioRxiv 2021.05.31.446445. https://doi.org/10.1101/2021.05.31.446445.

Ahmed, S., Bhasin, M., Manjunath, K., and Varadarajan, R. (2022). Prediction of Residue-specific Contributions to Binding and Thermal Stability Using Yeast Surface Display . Front. Mol. Biosci. *8*. .

Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. *38*, W529-33. https://doi.org/10.1093/nar/gkq399.

Bacon, K., Blain, A., Burroughs, M., McArthur, N., Rao, B.M., and Menegatti, S. (2020). Isolation of Chemically Cyclized Peptide Binders Using Yeast Surface Display. ACS Comb. Sci. *22*, 519–532. https://doi.org/10.1021/acscombsci.0c00076.

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., Wang, J., Cong, Q., Kinch, L.N., Schaeffer, R.D., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. Science *373*, 871–876. https://doi.org/10.1126/science.abj8754.

Bajaj, K., Madhusudhan, M.S., Adkar, B. V, Chakrabarti, P., Ramakrishnan, C., Sali, A., and Varadarajan, R. (2007). Stereochemical criteria for prediction of the effects of proline mutations on protein stability. PLoS Comput. Biol. *3*, e241. https://doi.org/10.1371/journal.pcbi.0030241.

Bernard, P., and Couturier, M. (1991). The 41 carboxy-terminal residues of the miniF plasmid CcdA protein are sufficient to antagonize the killer activity of the CcdB protein. Mol. Gen. Genet. MGG *226*, 297–304. https://doi.org/10.1007/BF00273616.

Bernard, P., and Couturier, M. (1992). Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. J. Mol. Biol. *226*, 735–745. https://doi.org/https://doi.org/10.1016/0022-2836(92)90629-X.

Bhasin, M., and Varadarajan, R. (2021). Prediction of Function Determining and Buried Residues Through Analysis of Saturation Mutagenesis Datasets . Front. Mol. Biosci. *8*, 58. .

Van Blarcom, T., Rossi, A., Foletti, D., Sundar, P., Pitts, S., Bee, C., Melton Witt, J., Melton, Z., Hasa-Moreno, A., Shaughnessy, L., et al. (2015). Precise and efficient antibody epitope determination through library design, yeast display and next-generation sequencing. J. Mol. Biol. *427*, 1513–1534. https://doi.org/10.1016/j.jmb.2014.09.020.

Boder, E.T., and Wittrup, K.D. (1997). Yeast surface display for screening combinatorial

polypeptide libraries. Nat. Biotechnol. *15*, 553–557. https://doi.org/10.1038/nbt0697-553.

Chakravarty, D., and Porter, L.L. (2022). AlphaFold2 fails to predict protein fold switching. Protein Sci. *31*, e4353. https://doi.org/10.1002/pro.4353.

Chakravarty, S., and Varadarajan, R. (1999). Residue depth: a novel parameter for the analysis of protein structure and stability. Structure *7*, 723–732. https://doi.org/10.1016/s0969-2126(99)80097-5.

Chandra, S., Chattopadhyay, G., and Varadarajan, R. (2021). Rapid Identification of Secondary Structure and Binding Site Residues in an Intrinsically Disordered Protein Segment. Front. Genet. *12*, 2173. https://doi.org/10.3389/fgene.2021.755292.

Chandra, S., Gupta, K., Khare, S., Kohli, P., Asok, A., Mohan, S.V., Gowda, H., and Varadarajan, R. (2022). Codon optimality is the primary contributor to the exceptional mutational sensitivity of CcdA antitoxin in its operonic context. BioRxiv https://doi.org/10.1101/2022.01.15.476443.

Chao, G., Lau, W.L., Hackel, B.J., Sazinsky, S.L., Lippow, S.M., and Wittrup, K.D. (2006). Isolating and engineering human antibodies using yeast surface display. Nat. Protoc. *1*, 755–768. https://doi.org/10.1038/nprot.2006.94.

Colby, D.W., Kellogg, B.A., Graff, C.P., Yeung, Y.A., Swers, J.S., and Wittrup, K.D. (2004). Engineering antibody affinity by yeast surface display. Methods Enzymol. *388*, 348–358. https://doi.org/10.1016/S0076-6879(04)88027-3.

Combet, C., Blanchet, C., Geourjon, C., and Deléage, G. (2000). NPS@: network protein sequence analysis. Trends Biochem. Sci. *25*, 147–150. https://doi.org/10.1016/s0968-0004(99)01540-6.

Critchlow, S.E., O'Dea, M.H., Howells, A.J., Couturier, M., Gellert, M., and Maxwell, A.

(1997). The interaction of the F plasmid killer protein, CcdB, with DNA gyrase: induction of DNA cleavage and blocking of transcription11J. Karn. J. Mol. Biol. *273*, 826–839. https://doi.org/https://doi.org/10.1006/jmbi.1997.1357.

Cunningham, B.C., and Wells, J.A. (1989). High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. Science *244*, 1081–1085. https://doi.org/10.1126/science.2471267.

Dao-Thi, M.-H., Charlier, D., Loris, R., Maes, D., Messens, J., Wyns, L., and Backmann, J. (2002). Intricate interactions within the ccd plasmid addiction system. J. Biol. Chem. *277*, 3733–3742. https://doi.org/10.1074/jbc.M105505200.

Dao-Thi, M.-H., Van Melderen, L., De Genst, E., Afif, H., Buts, L., Wyns, L., and Loris, R. (2005). Molecular basis of gyrase poisoning by the addiction toxin CcdB. J. Mol. Biol. *348*, 1091–1102. https://doi.org/10.1016/j.jmb.2005.03.049.

Das, R., Datta, R., and Varadarajan, R. (2020). Probing the structure of the HIV-1 envelope trimer using aspartate scanning mutagenesis. J. Virol. *94*, e01426-20. https://doi.org/10.1128/JVI.01426-20.

Datta, R., Roy Chowdhury, R., Manjunath, K., Hanna, L.E., and Varadarajan, R. (2020). A facile method of mapping HIV-1 neutralizing epitopes using chemically masked cysteines and deep sequencing. Proc. Natl. Acad. Sci. U. S. A. *117*, 29584–29594. https://doi.org/10.1073/pnas.2010256117.

DeBartolo, J., Dutta, S., Reich, L., and Keating, A.E. (2012). Predictive Bcl-2 family binding models rooted in experiment or structure. J. Mol. Biol. *422*, 124–144. https://doi.org/10.1016/j.jmb.2012.05.022.

Dehouck, Y., Kwasigroch, J.M., Rooman, M., and Gilis, D. (2013). BeAtMuSiC: Prediction of

changes in protein-protein binding affinity on mutations. Nucleic Acids Res. *41*, W333-9. https://doi.org/10.1093/nar/gkt450.

Dogan, J., Mu, X., Engström, Å., and Jemth, P. (2013). The transition state structure for coupled binding and folding of disordered protein domains. Sci. Rep. *3*, 2076. https://doi.org/10.1038/srep02076.

Doolan, K.M., and Colby, D.W. (2015). Conformation-Dependent Epitopes Recognized by Prion Protein Antibodies Probed Using Mutational Scanning and Deep Sequencing. J. Mol. Biol. *427*, 328–340. https://doi.org/https://doi.org/10.1016/j.jmb.2014.10.024.

Dyson, H.J., and Wright, P.E. (2005). Intrinsically unstructured proteins and their functions. Nat. Rev. Mol. Cell Biol. *6*, 197–208. https://doi.org/10.1038/nrm1589.

Eisenberg, D. (1984). Three-dimensional structure of membrane and surface proteins. Annu. Rev. Biochem. *53*, 595–623. https://doi.org/10.1146/annurev.bi.53.070184.003115.

Ernst, A., Gfeller, D., Kan, Z., Seshagiri, S., Kim, P.M., Bader, G.D., and Sidhu, S.S. (2010). Coevolution of PDZ domain-ligand interactions analyzed by high-throughput phage display and deep sequencing. Mol. Biosyst. *6*, 1782–1790. https://doi.org/10.1039/c0mb00061b.

Forsyth, C.M., Juan, V., Akamatsu, Y., DuBridge, R.B., Doan, M., Ivanov, A. V, Ma, Z., Polakoff, D., Razo, J., Wilson, K., et al. (2013). Deep mutational scanning of an antibody against epidermal growth factor receptor using mammalian cell display and massively parallel pyrosequencing. MAbs *5*, 523–532. https://doi.org/10.4161/mabs.24979.

Fowler, D.M., and Fields, S. (2014). Deep mutational scanning: a new style of protein science. Nat. Methods *11*, 801–807. https://doi.org/10.1038/nmeth.3027.

Fowler, D.M., Araya, C.L., Fleishman, S.J., Kellogg, E.H., Stephany, J.J., Baker, D., and Fields, S. (2010). High-resolution mapping of protein sequence-function relationships. Nat. Methods *7*,

741–746. https://doi.org/10.1038/nmeth.1492.

Fujino, Y., Fujita, R., Wada, K., Fujishige, K., Kanamori, T., Hunt, L., Shimizu, Y., and Ueda, T. (2012). Robust in vitro affinity maturation strategy based on interface-focused high-throughput mutational scanning. Biochem. Biophys. Res. Commun. *428*, 395–400. https://doi.org/10.1016/j.bbrc.2012.10.066.

Gietz, R.D., and Schiestl, R.H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat. Protoc. *2*, 31–34. https://doi.org/10.1038/nprot.2007.13.

Giri, R., Morrone, A., Toto, A., Brunori, M., and Gianni, S. (2013). Structure of the transition state for the binding of c-Myb and KIX highlights an unexpected order for a disordered system. Proc. Natl. Acad. Sci. *110*, 14942 LP – 14947. https://doi.org/10.1073/pnas.1307337110.

Gray, V.E., Hause, R.J., and Fowler, D.M. (2017). Using large-scale mutagenesis to guide single amino acid scanning experiments. BioRxiv 140707. https://doi.org/10.1101/140707.

Heyne, M., Papo, N., and Shifman, J.M. (2020). Generating quantitative binding landscapes through fractional binding selections combined with deep sequencing and data normalization. Nat. Commun. *11*, 297. https://doi.org/10.1038/s41467-019-13895-8.

Hietpas, R.T., Jensen, J.D., and Bolon, D.N.A. (2011). Experimental illumination of a fitness landscape. Proc. Natl. Acad. Sci. *108*, 7896 LP – 7901. https://doi.org/10.1073/pnas.1016024108.

Hussain, A.S.Z., Shanthi, V., Sheik, S.S., Jeyakanthan, J., Selvarani, P., and Sekar, K. (2002). PDB Goodies--a web-based GUI to manipulate the Protein Data Bank file. Acta Crystallogr. D. Biol. Crystallogr. *58*, 1385–1386. https://doi.org/10.1107/s090744490200985x.

Jenson, J.M., Xue, V., Stretz, L., Mandal, T., Reich, L.L., and Keating, A.E. (2018). Peptide design by optimization on a data-parameterized protein interaction landscape. Proc. Natl. Acad.

Sci. U. S. A. *115*, E10342–E10351. https://doi.org/10.1073/pnas.1812939115.

Jiang, L., Mishra, P., Hietpas, R.T., Zeldovich, K.B., and Bolon, D.N.A. (2013). Latent effects of Hsp90 mutants revealed at reduced expression levels. PLoS Genet. *9*, e1003600. https://doi.org/10.1371/journal.pgen.1003600.

De Jonge, N., Garcia-Pino, A., Buts, L., Haesaerts, S., Charlier, D., Zangger, K., Wyns, L., De Greve, H., and Loris, R. (2009). Rejuvenation of CcdB-Poisoned Gyrase by an Intrinsically Disordered Protein Domain. Mol. Cell *35*, 154–163. https://doi.org/10.1016/j.molcel.2009.05.025.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

Kanaya, E., Kanaya, S., and Kikuchi, M. (1990). Introduction of a non-native disulfide bridge to human lysozyme by cysteine scanning mutagenesis. Biochem. Biophys. Res. Commun. *173*, 1194–1199. https://doi.org/10.1016/s0006-291x(05)80912-x.

Kim, M.K., and Kang, Y.K. (1999). Positional preference of proline in alpha-helices. Protein Sci. *8*, 1492–1499. https://doi.org/10.1110/ps.8.7.1492.

Kowalsky, C.A., Faber, M.S., Nath, A., Dann, H.E., Kelly, V.W., Liu, L., Shanker, P., Wagner, E.K., Maynard, J.A., Chan, C., et al. (2015). Rapid fine conformational epitope mapping using comprehensive mutagenesis and deep sequencing. J. Biol. Chem. *290*, 26457–26470. https://doi.org/10.1074/jbc.M115.676635.

Larman, H.B., Xu, G.J., Pavlova, N.N., and Elledge, S.J. (2012). Construction of a rationally designed antibody platform for sequencing-assisted selection. Proc. Natl. Acad. Sci. U. S. A. *109*, 18523–18528. https://doi.org/10.1073/pnas.1215549109.

Laskowski, R.A., Jabłońska, J., Pravda, L., Vařeková, R.S., and Thornton, J.M. (2018). PDBsum: Structural summaries of PDB entries. Protein Sci. *27*, 129–134. https://doi.org/10.1002/pro.3289.

Lee, B., and Richards, F.M. (1971). The interpretation of protein structures: estimation of static accessibility. J. Mol. Biol. *55*, 379–400. https://doi.org/10.1016/0022-2836(71)90324-x.

Li, M., Simonetti, F.L., Goncearenco, A., and Panchenko, A.R. (2016). MutaBind estimates and interprets the effects of sequence variants on protein-protein interactions. Nucleic Acids Res. *44*, W494-501. https://doi.org/10.1093/nar/gkw374.

Lindström, I., Andersson, E., and Dogan, J. (2018). The transition state structure for binding between TAZ1 of CBP and the disordered Hif-1α CAD. Sci. Rep. *8*, 7872. https://doi.org/10.1038/s41598-018-26213-x.

Lipovsek, D., Lippow, S.M., Hackel, B.J., Gregson, M.W., Cheng, P., Kapila, A., and Wittrup, K.D. (2007). Evolution of an interloop disulfide bond in high-affinity antibody mimics based on fibronectin type III domain and selected by yeast surface display: molecular convergence with single-domain camelid and shark antibodies. J. Mol. Biol. *368*, 1024–1041. https://doi.org/10.1016/j.jmb.2007.02.029.

Madl, T., Melderen, L. Van, Mine, N., Respondek, M., Oberer, M., Keller, W., Khatai, L., and Zangger, K. (2006). Structural Basis for Nucleic Acid and Toxin Recognition of the Bacterial Antitoxin CcdA. 170–185. https://doi.org/10.1016/j.jmb.2006.08.082.

Maki, S., Takiguchi, S., Horiuchi, T., Sekimizu, K., and Miki, T. (1996). Partner switching mechanisms in inactivation and rejuvenation of Escherichia coli DNA gyrase by F plasmid proteins LetD (CcdB) and LetA (CcdA). J. Mol. Biol. *256*, 473–482. https://doi.org/10.1006/jmbi.1996.0102.

Mata-Fink, J., Kriegsman, B., Yu, H.X., Zhu, H., Hanson, M.C., Irvine, D.J., and Wittrup, K.D. (2013). Rapid conformational epitope mapping of anti-gp120 antibodies with a designed mutant panel displayed on yeast. J. Mol. Biol. *425*, 444–456. https://doi.org/10.1016/j.jmb.2012.11.010.

Mateos, B., Conrad-Billroth, C., Schiavina, M., Beier, A., Kontaxis, G., Konrat, R., Felli, I.C., and Pierattelli, R. (2020). The Ambivalent Role of Proline Residues in an Intrinsically Disordered Protein: From Disorder Promoters to Compaction Facilitators. J. Mol. Biol. *432*, 3093–3111. https://doi.org/https://doi.org/10.1016/j.jmb.2019.11.015.

McLaughlin, R.N.J., Poelwijk, F.J., Raman, A., Gosal, W.S., and Ranganathan, R. (2012). The spatial architecture of protein function and adaptation. Nature *491*, 138–142. https://doi.org/10.1038/nature11500.

Melamed, D., Young, D.L., Gamble, C.E., Miller, C.R., and Fields, S. (2013). Deep mutational scanning of an RRM domain of the Saccharomyces cerevisiae poly(A)-binding protein. RNA *19*, 1537–1551. https://doi.org/10.1261/rna.040709.113.

Van Melderen, L., Thi, M.H., Lecchi, P., Gottesman, S., Couturier, M., and Maurizi, M.R. (1996). ATP-dependent degradation of CcdA by Lon protease. Effects of secondary structure and heterologous subunit interactions. J. Biol. Chem. *271*, 27730–27738. https://doi.org/10.1074/jbc.271.44.27730.

Mészáros, B., Erdős, G., and Dosztányi, Z. (2018). IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res. *46*, W329–W337. https://doi.org/10.1093/nar/gky384.

Miki, T., Yoshioka, K., and Horiuchi, T. (1984). Control of cell division by sex factor F in Escherichia coli: I. The 42.84–43.6 F segment couples cell division of the host bacteria with replication of plasmid DNA. J. Mol. Biol. *174*, 605–625.

https://doi.org/https://doi.org/10.1016/0022-2836(84)90086-X.

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. Nat. Methods *19*, 679–682. https://doi.org/10.1038/s41592-022-01488-1.

Mittag, T., Kay, L.E., and Forman-Kay, J.D. (2010). Protein dynamics and conformational disorder in molecular recognition. J. Mol. Recognit. *23*, 105–116. https://doi.org/10.1002/jmr.961.

Najar, T.A., Khare, S., Pandey, R., Gupta, S.K., and Varadarajan, R. (2017). Mapping Protein Binding Sites and Conformational Epitopes Using Cysteine Labeling and Yeast Surface Display. Structure *25*, 395–406. https://doi.org/10.1016/j.str.2016.12.016.

Newberry, R.W., Leong, J.T., Chow, E.D., Kampmann, M., and DeGrado, W.F. (2020). Deep mutational scanning reveals the structural basis for α-synuclein activity. Nat. Chem. Biol. *16*, 653–659. https://doi.org/10.1038/s41589-020-0480-6.

Pál, G., Fong, S.-Y., Kossiakoff, A.A., and Sidhu, S.S. (2005). Alternative views of functional protein binding epitopes obtained by combinatorial shotgun scanning mutagenesis. Protein Sci. *14*, 2405–2413. https://doi.org/10.1110/ps.051519805.

Park, S., Xu, Y., Stowell, X.F., Gai, F., Saven, J.G., and Boder, E.T. (2006). Limitations of yeast surface display in engineering proteins of high thermostability. Protein Eng. Des. Sel. *19*, 211–217. https://doi.org/10.1093/protein/gzl003.

Rathore, U., Purwar, M., Vignesh, V.S., Das, R., Kumar, A.A., Bhattacharyya, S., Arendt, H., DeStefano, J., Wilson, A., Parks, C., et al. (2018). Bacterially expressed HIV-1 gp120 outer-domain fragment immunogens with improved stability and affinity for CD4-binding site neutralizing antibodies. J. Biol. Chem. *293*, 15002–15020.

https://doi.org/10.1074/jbc.RA118.005006.

Ravn, U., Didelot, G., Venet, S., Ng, K.-T., Gueneau, F., Rousseau, F., Calloud, S., Kosco-Vilbois, M., and Fischer, N. (2013). Deep sequencing of phage display libraries to support antibody discovery. Methods *60*, 99–110. https://doi.org/10.1016/j.ymeth.2013.03.001.

Reich, L.L., Dutta, S., and Keating, A.E. (2015). SORTCERY-A High-Throughput Method to Affinity Rank Peptide Ligands. J. Mol. Biol. *427*, 2135–2150. https://doi.org/10.1016/j.jmb.2014.09.025.

Reich, L.L., Dutta, S., and Keating, A.E. (2016). Generating High-Accuracy Peptide-Binding Data in High Throughput with Yeast Surface Display and SORTCERY. Methods Mol. Biol. *1414*, 233–247. https://doi.org/10.1007/978-1-4939-3569-7_14.

Rodrigues, C.H.M., Myung, Y., Pires, D.E. V, and Ascher, D.B. (2019). mCSM-PPI2: predicting the effects of mutations on protein–protein interactions. Nucleic Acids Res. *47*, W338–W344. https://doi.org/10.1093/nar/gkz383.

Rogers, J.M., Wong, C.T., and Clarke, J. (2014). Coupled Folding and Binding of the Disordered Protein PUMA Does Not Require Particular Residual Structure. J. Am. Chem. Soc. *136*, 5197–5200. https://doi.org/10.1021/ja4125065.

Roscoe, B.P., Thayer, K.M., Zeldovich, K.B., Fushman, D., and Bolon, D.N.A. (2013). Analyses of the effects of all ubiquitin point mutants on yeast growth rate. J. Mol. Biol. *425*, 1363–1377. https://doi.org/10.1016/j.jmb.2013.01.032.

Ruff, K.M., and Pappu, R. V (2021). AlphaFold and Implications for Intrinsically Disordered Proteins. J. Mol. Biol. *433*, 167208. https://doi.org/10.1016/j.jmb.2021.167208.

Staller, M. V, Holehouse, A.S., Swain-Lenz, D., Das, R.K., Pappu, R. V, and Cohen, B.A. (2018). A High-Throughput Mutational Scan of an Intrinsically Disordered Acidic

Transcriptional Activation Domain. Cell Syst. *6*, 444-455.e6. https://doi.org/https://doi.org/10.1016/j.cels.2018.01.015.

Starita, L.M., Pruneda, J.N., Lo, R.S., Fowler, D.M., Kim, H.J., Hiatt, J.B., Shendure, J., Brzovic, P.S., Fields, S., and Klevit, R.E. (2013). Activity-enhancing mutations in an E3 ubiquitin ligase identified by high-throughput mutagenesis. Proc. Natl. Acad. Sci. U. S. A. *110*, E1263-72. https://doi.org/10.1073/pnas.1303309110.

Strodel, B. (2021). Energy Landscapes of Protein Aggregation and Conformation Switching in Intrinsically Disordered Proteins. J. Mol. Biol. *433*, 167182. https://doi.org/10.1016/j.jmb.2021.167182.

Swers, J.S., Kellogg, B.A., and Wittrup, K.D. (2004). Shuffled antibody libraries created by in vivo homologous recombination and yeast surface display. Nucleic Acids Res. *32*, e36–e36. https://doi.org/10.1093/nar/gnh030.

Tan, K.P., Varadarajan, R., and Madhusudhan, M.S. (2011). DEPTH: A web server to compute depth and predict small-molecule binding cavities in proteins. Nucleic Acids Res. *39*, 1–7. https://doi.org/10.1093/nar/gkr356.

Tan, K.P., Nguyen, T.B., Patel, S., Varadarajan, R., and Madhusudhan, M.S. (2013). Depth: a web server to compute depth, cavity sizes, detect potential small-molecule ligand-binding cavities and predict the pKa of ionizable residues in proteins. Nucleic Acids Res. *41*, W314–W321. https://doi.org/10.1093/nar/gkt503.

Tripathi, A., Gupta, K., Khare, S., Jain, P.C., Patel, S., Kumar, P., Pulianmackal, A.J., Aghera, N., and Varadarajan, R. (2016). Molecular Determinants of Mutant Phenotypes, Inferred from Saturation Mutagenesis Data. Mol. Biol. Evol. *33*, 2960–2975. https://doi.org/10.1093/molbev/msw182.

Uchański, T., Zögg, T., Yin, J., Yuan, D., Wohlkönig, A., Fischer, B., Rosenbaum, D.M., Kobilka, B.K., Pardon, E., and Steyaert, J. (2019). An improved yeast surface display platform for the screening of nanobody immune libraries. Sci. Rep. *9*, 382. https://doi.org/10.1038/s41598-018-37212-3.

Uversky, V.N. (2014). The triple power of $D^3$: protein intrinsic disorder in degenerative diseases. Front. Biosci. (Landmark Ed. *19*, 181–258. https://doi.org/10.2741/4204.

Uversky, V.N., and Dunker, A.K. (2010). Understanding protein non-folding. Biochim. Biophys. Acta *1804*, 1231–1264. https://doi.org/10.1016/j.bbapap.2010.01.017.

Uversky, V.N., Oldfield, C.J., and Dunker, A.K. (2008). Intrinsically disordered proteins in human diseases: introducing the D2 concept. Annu. Rev. Biophys. *37*, 215–246. https://doi.org/10.1146/annurev.biophys.37.032807.125924.

Vandervelde, A., Drobnak, I., Hadˇ, S., Sterckx, Y.G., Welte, T., Greve, H. De, Charlier, D., Efremov, R., Loris, R., and Lah, J. (2017). NAR Breakthrough Article Molecular mechanism governing ratio-dependent transcription regulation in the ccdAB operon. *45*, 2937–2950. https://doi.org/10.1093/nar/gkx108.

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., et al. (2012). Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. Nat. Biotechnol. *30*, 543–548. https://doi.org/10.1038/nbt.2214.