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

Soumyanetra Chandra | Kavyashree Manjunath | Aparna Asok | Raghavan Varadarajan

1Molecular Biophysics Unit, Indian Institute of Science, Bangalore, India
2Institute for Stem Cell Science and Regenerative Medicine, Bangalore, India

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 rudimentary model was developed in an attempt to predict mutational effects on binding affinity of IDPs that form alpha-helical structures upon binding.

KEYWORDS
affinity prediction, protein–protein interaction, saturation mutagenesis, secondary structure prediction, single-site mutations, α-helix

1 | 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 & 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 & 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 (Datta et al., 2020; Doolan & Colby, 2015; Kowalsky et al., 2015; Mata-Fink et al., 2013; Najar et al., 2017; Van Blarcom et al., 2015) 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 & 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 & 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 & 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 (residues 1–39) and an intrinsically disordered C-terminal domain (residues 40–72). The CcdA N-terminal domain is responsible for dimerization (Madl et al., 2006; van Melder et al., 1996) and binds to the operator/promoter region of the ccdAB operon, playing an important role in autoregulation of ccdAB transcription (Aff et al., 2001; Dao-Thi et al., 2002; Vanderwelde 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 & 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, $\Delta G_{\text{bind}}^\circ$ 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 & Porter, 2022; Ruff & Pappu, 2021; Strodel, 2021). Based on the mutational tolerance scores for CcdA single variants, we also developed an interaction weighted mutational penalty-based model, IAPred (Intrinsically disordered Protein Affinity prediction), to predict mutational effects on binding affinity for IDPs involved in PPIs. While IAPred is a predictive model based solely on the DMS data of a single IDP, it showed encouraging results when applied to prediction of mutational effects on partner binding in a number of IDPs. IAPred can be
further improved once more DMS datasets that describe mutational effects on binding for IDRAs become available.

2 | RESULTS

2.1 | 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 Aga2p, 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 (Figure 1). We observed that single mutations in CcdA protein generally fail to completely abolish CcdB binding (Figure 2a, b). The available complex structure of

**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 normalization of the obtained raw data representing the mutant distribution across the affinity bins, a mutational tolerance score (MTSseq) is assigned to each mutant (see Section 4), which allows both ranking of mutants based on affinity as well as understanding residue specific contributions to the binding energetics.
the CcdA C-terminal domain bound to CcdB (PDB ID: 3G7Z) (de Jonge et al., 2009) (Figure 2c, d) shows an extensive interface surface area of \( \sim 2850 \, \text{Å}^2 \) between the CcdA C-terminus and the CcdB dimer, explaining the nominal effects of CcdA single mutants on binding energetics.

![Figure 2](image_url)

**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 2). (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).

seen in the case of CcdA (Figure 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 labor 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 (Figure 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

### 2.2 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; Reich et al., 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
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 (Figure S1a, b).

We recently used a similar approach to estimate mean fluorescence intensity, MFI of CcdB mutants using YSD in a previous study and showed that this simplified approach yields MFI values similar to more complicated Maximum Entropy based approaches (Ahmed et al., 2022). In current work, the MFI value of the frequency distribution across the Binding: Expression ratio axis was assigned as MFI\textsuperscript{ratio} for \sim 1290 single mutants of CcdA (94% of the total library diversity) (see Section 4 and Figure 1). The deep sequencing and MFI\textsuperscript{ratio} assignment were carried out in two biological replicates, which were highly correlated (Figure S1c). The MFI\textsuperscript{ratio} values were normalized with respect to the WT MFI\textsuperscript{ratio} and averaged across the replicates to obtain the deep sequencing derived Mutational Tolerance Scores (MTS\textsubscript{seq}) for CcdA mutants. The MTS\textsubscript{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) (Figure S1d). Expression scores obtained from deep sequencing for the CcdA mutants showed absence of significant mutational effects on surface expression (Figure S2), unlike effects seen in case of a notable fraction of yeast surface displayed globular protein variants (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 (Figure 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.

### 2.3 Experimental validation of BAMseq methodology

We experimentally determined the dissociation constant, \(K_d\) for 20 individual single mutants and WT CcdA (Figures 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 ± 0.05 nM for CcdB binding (Figure S3a and 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\textsubscript{2} interactions showed a \(K_d\) value of \sim 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 (\(\Delta G_{\text{bind}}\)) was calculated for each of the 20 CcdA mutants using the fitted \(K_d\) values (Table S1). These experimentally determined \(\Delta G_{\text{bind}}\) values correlate well with the DMS derived MTS\textsubscript{seq} scores (Pearson correlation, \(r = -0.84\)) (Figure 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\textsubscript{seq} scores and fit better to an inverse first order equation (Figure 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 Methods S1), where \(K_d\) = dissociation constant, \([C] = \text{CcdB (ligand)}\) concentration used in the FACS experiment and \([A] = \text{concentration of free CcdA}\), Simulation of theoretical curves
helps to determine the ideal ligand concentrations that can be used in BAMseq experiments for estimating binding affinities accurately (Figure S5).

2.4 | 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 MTSseq scores, since the MTSseq score only provides a relative measure of the CcdB binding affinities of CcdA mutants. The highly correlated MTSseq scores and experimental $\Delta G_{\text{bind}}$ values of 20 experimentally validated mutants provide the necessary standard curve (Figure S3b), facilitating extrapolation of apparent $\Delta G_{\text{bind}}$ and apparent $K_d$ for mutants from the MTSseq 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 $\Delta G_{\text{bind}}$ and $K_d$ values for all ~1290 CcdA single mutants that lack prior experimental binding energetic information. The apparent $\Delta G_{\text{bind}}$ 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 (Figure S3d).

2.5 | 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 MTSseq values close to that of WT CcdA, except for a few mutants that

---

**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) (de Jonge et al., 2009). (c) The color key used depicts the ranges of MTSseq scores representing CcdB affinity. WT CcdA has MTSseq score of 1. The poor binders with lower tolerance scores are shown in dark colors, while lighter colors depict marginally better binders. Gray cells and blank cells (marked with X) indicate WT residue and mutants with no available data, respectively.
display modest impairment in CcdB binding (Figure 3a). Minor binding defects upon mutation occur primarily at buried positions (Figure 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 (Figure 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 Melder et al., 1996) (Figure 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) (Figure 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.

2.6 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

![Figure 4](image-url)
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 (Table S2). ΔASA is calculated by subtracting the accessible surface area (Lee & 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 & 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 (MTSseqavg) of each residue position correlates fairly well with the respective ΔASA, number of interactions and residue depth (Figure 4a–c).

Functionally important residues in proteins are generally found to be conserved across species in order to maintain optimal activity. Surprisingly, the MTSseq values 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 (Figure 4d and Table S2). This is consistent with the poor correlation between evolutionary conservation of residues and positional mutational tolerance observed in our previous study in M. tuberculosis antitoxin, MazE (Chandra et al., 2021). The current study shows that mutations at certain CcdA residues that are highly conserved, did not reduce binding affinity. This indicates that there might be other functions of the antitoxin molecule, besides binding and inhibiting the toxin; for example antitoxins need to be rapidly degraded in the absence of toxin. In addition, we have seen very dramatic phenotypic changes associated with synonymous mutations in the ccdA gene in its native operonic context (Chandra et al., 2022), these likely result from alterations in the relative amounts of antitoxin and toxin produced from the operon in such mutants. Additionally, for CcdA, residue evolutionary conservation does not correlate well with structural parameters such as ΔASA, number of interactions and changes in residue depth upon complex formation. In the current study, some point mutations that are poorly tolerated in E. coli CcdA are found in homologous sequences 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. Finally, in the case of antitoxins, the number of homologs is typically low- a feature which is also observed for IDPs (Brown et al., 2002; Brown et al., 2011) in general, this weakens inferences that can be made from evolutionary conservation analyses.

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 MTSseq values of $1.02 \pm 0.152$ and $0.49 \pm 0.165$, respectively. We categorize mutants having MTSseq lower than 0.9 ($\approx$mean$^{\text{class}2} + 2.5 \sigma^{\text{class}2}$) 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 with these structure based classes, the contributing and non-contributing classes of residues in CcdA identified solely based on a MTSseqavg cutoff of 0.9 yields reasonable sensitivity (89%), specificity (80%) and accuracy (76%) calculated as described previously (Bhasin & Varadarajan, 2021; Tripathi et al., 2016) (Figure 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 (Figure 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 72 W 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 (Figure 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.

### 2.7 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 (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 (Figure 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 (Figure 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 (Figure S7e), indicating that partner binding and activity of CcdA do not depend strongly on inherent helix forming capability of the disordered regions. We also observed that the mutational tolerance correlates poorly with the predicted helical content of CcdA C-terminal domain single mutants estimated using Agadir (Figure S7f). Agadir is an algorithm based on helix-coil transition theory that predicts the helical content of monomeric peptides (Lacroix et al., 1998; Muñoz & Serrano, 1994).

2.8 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 & 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\textsubscript{seq} patterns for each of the 20 substituted amino acids with the average residue MTS\textsubscript{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 (Figure S8a). Further, the single mutant MTS\textsubscript{seq} scores for Cys and charged mutants correlate well with the residue depth of CcdA residues in the complex (Figure S8b). While the charged substitutions possibly hinder 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 (Figure S8).

2.9 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\textsubscript{seq} (charged) (Figure 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\textsubscript{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 \(\alpha\)-helix containing 3.6 amino acids per turn (Figure 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 (Figure 5c). This indicates that the first 23 amino acids (region 40–62) form an \(\alpha\)-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 \text{MTS}_{\text{seq}} scores for individual charged substitutions also show a similar oscillating pattern to the averaged values, (Figure S9) and can be used to predict periodicity and structure.

### 2.10 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 (Figure 3b). This is because Glycine 63 has a non-canonical backbone conformation in the bound state \((\Phi, \Psi = 143, 144)\) (Figure 5d), which is uncharacteristic of an \(\alpha\)-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 (Figure 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 \((K_d = 13 \mu M)\) and the second, where the whole C-terminal 40–72 stretch binds with a high affinity \((K_d = 20 \mu M)\) (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.

### TABLE 1 Predominance of glycine residues with positive torsion angles in bacterial type II antitoxins

| Residue/Position | Antitoxin | Complex PDB ID | Chain | TA module | Organism | \(\Phi\) torsion angle \(^\circ\) |
|------------------|-----------|----------------|-------|-----------|----------|------------------|
| 1 Gly 25         | Maz E     | 6A6X           | C     | MazEF     | 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   |

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

*\(\Phi\) refers to the C-N-C backbone torsion angle for the Glycine residue calculated from toxin-antitoxin complex structures using PDB Goodies (Hussain et al., 2002).*
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 $\Phi$ 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.

### 2.11 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 $\Delta G^o_{\text{bind}}(\text{mut} - \Delta G^o_{\text{bind}})$ values for 20 single CcdA mutants (Figure S10a–c). However, these predicted mutational effects are significantly larger than observed experimentally, for CcdA. The BAMseq derived apparent $\Delta G^o_{\text{bind}}$ for all single mutants by BAMseq was compared with 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 (Figure S10d). Comparison of apparent $\Delta G^o_{\text{bind}}$ for all single CcdA mutants by BAMseq to the predicted $\Delta G^o_{\text{bind}}$ 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 (Figure S11). These observations highlight the drawbacks of using the currently available predictors of mutational effects on energetics of PPIs in case of IDPs.

### 2.12 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 of mutational tolerance in CcdA on residue burial and physico-chemistry of substitutions, we have designed and tested a predictive model IPApred (Intrinsically disordered Protein Affinity prediction) (Figure 6) to explore the utility of DMS data in functional predictions. This method allows prediction of the functional effects of single mutations on the binding affinity, in terms of an arbitrary Mutational Tolerance Score (MTS$\text{pred}$) for IDPs forming extended interfaces with partners. For each mutant, the MTS$\text{pred}$ 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. We have observed that residue depth in CcdA correlates better with positional mutational tolerance relative to $\Delta$ASA (Figure 4). Thus, we have introduced residue depth in the model to weigh mutational effects at each position. Using residue depth 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$\text{pred}$ correlated well ($r = 0.65$) with the experimental MTS$\text{seq}$ for the CcdA mutants test set (Figure 7a). Based on the good correlation between the $\Delta$MTS$\text{pred}$ (MTS$\text{pred}$ $-$ MTS$\text{seq}$) values from IPApred and the experimentally determined $\Delta G^o_{\text{bind}}$ values from IPApred and the experimentally determined $\Delta G^o_{\text{bind}}$ in the set of validated 20 CcdA mutants (Figure 7b), we further developed our approach to predict $\Delta G^o_{\text{bind}}$ both for CcdA and other systems. Our predictions for CcdA also correlated moderately with predictions by BeAtMuSiC server (Figure 7c).

Validation 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 (Nuclear Coactivator Binding Domain) domain of CBP (CREB binding protein) (Dogan et al., 2013) and Hif 1α (hypoxia inducible factor 1α)
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 ΔΔG°/C14 binding of 20 individual mutants as a function of ΔMTSpred (MTSpred (mut) – MTSpred(WT)) is used to further estimate the predicted ΔΔG°/C14 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 (Equation 9) as used for CcdA (a–c). The ΔΔG°/C14 values are then predicted using these calculated MTSpred scores (Equation 10).
(Lindström et al., 2018). All these IDRs fold to predominantly helical structures. Our predictions showed better correlation to the literature derived experimental data, relative to predictions of BeAtMuSiC (Figure 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 (Figure 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 (Figure 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 (Figure 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 (Figure 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 (Figure 7g, h).

However, CcdA DMS data driven IPApred is unable 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 ΔΔGbind 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 ΔΔGbind values may be smaller 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 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 with 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 (Figure S13), indicating applicability of IPApred for prediction of binding energetic effects for all possible single-site mutants of IDPs.

3 | DISCUSSION

Deep mutational scanning of a yeast surface displayed library of single-site CcdA variants reveal that mutations in its N-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 mutations (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,

**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 (MTSpred) and the experimental mutational tolerance scores (MTSexp) obtained from the DMS study for the test set of CcdA C-terminal domain mutants. (b) Correlation of ΔMTSpred (MTSpredmutant - MTSpredWT) with ΔΔGbind Values determined experimentally for 20 single-siteCcdA mutants. The fitted straight line (where y is ΔΔGbind and x is ΔMTSpred) is further used as an internal standard to derive predicted ΔΔGbind values from MTSpred scores in other systems. (c) Correlation of the ΔΔGbind values predicted by IPApred for CcdA C-terminal domain mutants in the test set with the corresponding ΔΔGbind 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: 1KBIH), 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 ΔΔGbind Values predicted by IPApred for disordered protein domains, with available experimentally determined ΔΔGbind values for corresponding single mutants. (f, h, j, and l) Correlation of the ΔΔGbind values for disordered protein domains predicted by BeAtMuSiC with available experimentally determined ΔΔGbind 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 and 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 ΔΔGbind Values after removal of a single outlier in each case (indicated in Figure S12) is shown (i–l).
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 ∆ASA (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) (Figure S5), we found that CcdB (ligand) concentrations in the range of 2–10 nM are suitable for the single ligand concentration approach used in BAMseq 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 (MTSseq) calculated for each single mutant describes the mutational effects on IDP-target interactions accurately and elucidates sequence-function relationships in IDPs. Further, these scores also allow parallel and high throughput quantification of binding energetics of all protein variants using an internal standard. While wild type CcdA shows a $ΔG^\text{bind}_{\text{obs}}$ value of $−13.08 \pm 0.05$ kcal/mole, $ΔG^\text{bind}_{\text{obs}}$ 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 MTSseq 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 residues, positions), including the CcdB non-interacting sites (Figure 3). Proline is known to disrupt helices except when located at the first three N-terminal positions (Bajaj et al., 2007; Kim & Kang, 1999). This is owing to its lack of the amide hydrogen that forms an essential intramolecular 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 $\Phi$ 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 $\Phi$ torsional angle and putative backbone bending roles, widespread in several type II antitoxins (Table 1). We suspect that our understanding of such interface shaping residues in proteins and IDPs will improve considerably upon availability of more mutational landscapes of IDR-partner pairs for which the complex structure is known.

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 case of CcdA, NMR reveals that the C-terminal domain is unstructured in the free state (Madl et al., 2006) and is also predicted to have low helical content (only 3.14%) by Agadir. Additionally, we know that the structure attained upon binding in CcdA is not fully alpha helical and forms a bent, elongated interface (de Jonge et al., 2009). It is therefore not easy to make detailed structural predictions using solely a helical wheel model or aromatic residues predicted to be potential binding sites. Instead, we looked at the periodicity in wave-like pattern of mutational effects in CcdA to determine local structural features. 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, reminiscent of an $\alpha$-helix with 3.6 residues per turn. 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 (Figure 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 (Figures 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 simple 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. 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$_{\text{pred}}$) is a useful approach. The MTS$_{\text{pred}}$ scores can be converted to $\Delta \Delta G^\circ$ values by using a small experimental dataset of binding affinity measurements of single mutants (~10–15 variants) of the IDP system of interest. This low-throughput experimental data can act as an accurate internal standard to obtain a mutational prediction on binding energetics for any single-site mutant, as well as to determine the extent of mutational effects on binding affinity in various IDPs with interfaces of various dimensions. Using CcdA experimental affinity values for the same purpose may produce inaccurate quantitative estimates of $\Delta \Delta G^\circ$ values in cases of some IDPs.

We applied IPApred to a number of IDP domains with available experimental binding affinity data. These test set IDPs happen to fold to predominantly $\alpha$-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 all cases of these IDRs, except in certain
residue stretches of one IDR that fail to attain helical structure upon target binding. This is likely because IPApred was trained solely on the only available IDP DMS data set of CcdA C-terminal domain that forms a structure with ~70% helical content upon partner binding. We believe predictions can be significantly improved by incorporating high throughput mutational data on binding effects in other IDPs attaining diverse structures upon binding to protein partners, but such data are currently unavailable. Such DMS studies will also help verify the generalizability of the mutational effects observed in CcdA. We also require experimentally determined ΔΔG values for small sets of mutants (from various classes of substitutions) of multiple IDPs to understand the extent of mutational effects on binding affinity in an IDP and its dependence on interface area. While carrying out DMS for other such IDPs is beyond the scope of the current study, this will be done in the future. Despite the above shortcomings, IPApred showcases the utility of DMS data for calibrating and subsequently predicting mutational effects on IDP binding energetics.

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 on 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. Structure prediction in proteins with few homologs, a common situation for 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. This makes prediction of bound structure of IDPs a challenge. With the availability of more DMS data, in combination with deep learning based predictions of protein–protein complex structure (Jumper et al., 2021; Mirdita et al., 2022) should lead to improvements in our ability to predict functional effects of mutations in intrinsically disordered functional domains in proteins where structural and/or functional information is often largely unavailable.

4 | MATERIALS AND METHODS

Brief descriptions are provided below; more details are available in the Supplemental Methods S1.

4.1 | 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 & 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.

4.2 | 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 μM CcdB was incubated at 4°C for 2 h 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.

4.3 | 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 (2 nM for sorting and varying concentrations between 0.01 pM and 200 nM 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).
4.4 Determination of mutational tolerance score (MTSseq) for each mutant

NGS data for the CcdA mutant library was processed using an in-house developed pipeline (see Methods S1). 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 Methods S1 for details on normalization).

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

\[
\text{MFI}_x = \frac{\sum_i (D^i_x) \times F_i}{\sum_i F_i}
\]

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 Methods S1).

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

\[
\text{MTS}_{\text{seq}} = \frac{\text{MFI}_x}{\text{MFI}_{\text{WT}}}
\]

The MTS\textsubscript{seq} values depicted in the heatmaps are the MTS\textsubscript{seq} values averaged over replicates 1 and 2 to maximize the information obtainable from the dataset. The MTS\textsubscript{seq} values used in correlation studies are MTS\textsubscript{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\textsubscript{seq} values of Replicate 2 and showed comparable Pearson correlation coefficients.

Using the same approach, we also calculated the Expression Scores (ES\textsubscript{seq}) for all single mutants of CcdA from the raw mutant reads obtained in the gates P11-P15 (Figure S1d).

4.5 Determination of \(K_d\) and \(\Delta G^0\text{bind}\) for surface displayed WT CcdA and single mutants

WT and single mutants of CcdA were cloned in the pETcon vector in \textit{S. cerevisiae} EBY100 strain. The CcdA molecules were expressed, displayed and probed for CcdB binding with a range of concentrations (0.01 pM to 200 nM) 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:

\[
y = x B_{\text{max}} / (x + K_d)
\]

where \(B_{\text{max}}\) is the saturated fluorescence Intensity and \(K_d\) is the dissociation constant of binding (Rathore et al., 2018). The Gibbs free energy of binding (\(\Delta G^0\text{bind}\)) was calculated using the relationship \(\Delta G^0\text{bind} = R T \ln K_d\), where \(R\) (ideal gas constant) = 1.9872 * 10\textsuperscript{-3} kcal K\textsuperscript{-1} mole\textsuperscript{-1} and \(T\) (temperature) = 298.15 K. The YSD titration estimated \(K_d\) and \(\Delta G^0\text{bind}\) values of CcdB binding for WT and mutant CcdA molecules are denoted as experimentally derived binding energetic parameters (\(K_d\)\textsubscript{exp} and \(\Delta G_{\text{bind}}\text{exp}\)) in this work.

4.6 Calculation of apparent \(\Delta G^0\text{bind}\) using MTS\textsubscript{seq} values

Based on the observed high correlation \((r = -0.84)\) between the experimentally determined \(\Delta G^0\text{bind}\) values of CcdA molecules (WT and single mutants) and the corresponding deep sequencing derived MTS\textsubscript{seq} values, we acquired a standard curve describing this relationship. The data points of \(\Delta G^0\text{bind}_{\text{exp}}\) versus MTS\textsubscript{seq} axis plot were fit to a straight line equation. This was used as an internal standard to calculate the apparent \(\Delta G^0\text{bind}_{\text{seq}}\) from obtained from deep sequencing for mutants (with unknown experimental \(K_d\) value) using the measured MTS\textsubscript{seq} values as follows:

\[
\Delta G^0\text{bind}_{\text{seq}} = -2.04 \times \text{MTS}_{\text{seq}} - 11.0035
\]

4.7 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 (Figure 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 Methods S1). 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 MTSpred) from the residue depth parameter obtained from the protein-partner complex structure.

Positional MTSpred = 1.315 – 0.0934 residue depth \( (5) \)

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 \( j \):

\[
 nRD^j = \frac{2RD^j}{RD^{\text{max}}} \tag{6}
\]

where \( RD^j \) is residue depth of a particular residue \( j \) in the IDP, and \( RD^{\text{max}} \) 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 \((a \rightarrow b)\) category of mutation at position \( j \), which was finally used in the current work is as follows:

\[
 \text{Predicted score}^j \! \! (a \rightarrow b) = \text{positional MTSpred}^j \! \! + nRD^j \times \text{Category Penalty}^{(a \rightarrow b)} \tag{7}
\]

where

\[
 \text{Category penalty}^{(a \rightarrow b)} = \text{category averaged MTSeq}^{(a \rightarrow b)} - \text{WT MTSeq} \tag{8}
\]

\[
 = \text{category averaged MTSeq}^{(a \rightarrow b)} - 1
\]

category averaged MTSeq is the averaged MTSeq value for all substitutions of \( a \rightarrow b \) category in the training set.

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

\[
 \text{MTSpred}^{j,(a \rightarrow b)} = 0.45 \times \text{Predicted score}^{j,(a \rightarrow b)} + 0.62 \tag{9}
\]

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

4.8 Prediction of \( \Delta \Delta G_{\text{bind}} \) values from predicted mutational tolerance scores

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

\[
 \Delta \Delta G_{\text{pred}}^{j,(a \rightarrow b)} = -1.31 \Delta \text{MTSpred}^{j,(a \rightarrow b)} - 0.01 \tag{10}
\]
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.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT
All materials generated in this study are available from the lead contact (varadar@iisc.ac.in) without restriction. The data relevant to the figures in the paper have been made available within the article and in the Supplementary Information section or submitted to GitHub (https://github.com/rvaradarajanlab/CcdA-SSM-Library-Yeast-surface-display). The deep sequencing data are submitted in NCBI SRA database and to be released upon acceptance of the manuscript (Bioproject ID: PRJNA813037; Reviewers’ link: https://dataview.ncbi.nlm.nih.gov/object/PRJNA813037?reviewer=569rl1llc305tdp4toafc676).

ORCID
Raghavan Varadarajan https://orcid.org/0000-0002-0823-7577

REFERENCES
Adams RM, Mora T, Walczak AM, Kinney JB, Supe N. Measuring the sequence-affinity landscape of antibodies with massively parallel titration curves. Elife. 2016;5:1–27. https://doi.org/10.7554/eLife.23156
Adkar BV, Tripathi A, Sahoo A, Bajaj K, Goswami D, Chakrabarti P, et al. Protein model discrimination using mutational sensitivity derived from deep sequencing. Structure (London, England: 1993). 2012;20(2):371–81. https://doi.org/10.1016/j.str.2011.11.021
Affé H, Allali N, Couturier M, van Melderen L. The ratio between CcdA and CcdB modulates the transcriptional repression of the ccd poison-antidote system. Mol Microbiol. 2001;41(1):73–82. https://doi.org/10.1046/j.1365-2958.2001.02492.x
Aghera NK, Prabha J, Tandon H, Chattopadhyay G, Vishwanath S, Srinivasan N, et al. Mechanism of CcdA-mediated rejuvenation of DNA gyrase. Structure (London, England: 1993). 2020;28(5):562–572.e4. https://doi.org/10.1016/j.str.2020.03.006
Ahmed S, Bhasin M, Manjunath K, Varadarajan R. Prediction of residue-specific contributions to binding and thermal stability using yeast surface display. Front Mol Biosci. 2022;8:800819. https://doi.org/10.3389/fmolb.2021.800819
Bacon K, Blain A, Burroughs M, McArthur N, Rao BM, Menegatti S. Isolation of chemically cyclized peptide binders using yeast surface display. ACS Comb Sci. 2020;22(10):519–32. https://doi.org/10.1021/acscombsci.0c00076
Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. Science (New York, N.Y.). 2021;373(6557):871–6. https://doi.org/10.1126/science.abj8754
Bajaj K, Madhusudhan MS, Adkar BV, Chakrabarti P, Ramakrishnan C, Sali A, et al. Stereoechemical criteria for prediction of the effects of proline mutations on protein stability. PLoS Comput Biol. 2007;3(12):e241. https://doi.org/10.1371/journal.pcbi.0030241
Bernard P, Couturier M. 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. 1991;226(1):297–304. https://doi.org/10.1007/BF00273616
Bernard P, Couturier M. Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. J Mol Biol. 1992;226(3):735–45. https://doi.org/10.1006/jmbi.1997.1357
Bhasin M, Varadarajan R. Prediction of function determining and buried residues through analysis of saturation mutagenesis datasets. Front Mol Biosci. 2021;8:58. https://doi.org/10.3389/fmolb.2021.635425
Boder ET, Wittrup KD. Yeast surface display for screening combinatorial polypeptide libraries. Nat Biotechnol. 1997;15(6):553–7. https://doi.org/10.1038/nbt0697-553
Brown CJ, Johnson AK, Dunker AK, Daughdrill GW. Evolution and disorder. Curr Opin Struct Biol. 2011;21(3):441–6. https://doi.org/10.1016/j.sbi.2011.02.005
Brown CJ, Takayama S, Campen AM, Vise P, Marshall TW, Oldfield CJ, et al. Evolutionary rate heterogeneity in proteins with long disordered regions. J Mol Evol. 2002;55(1):104–10. https://doi.org/10.1007/s00239-001-2309-6
Chakravarty D, Porter LL. AlphaFold2 fails to predict protein fold switching. Protein Sci. 2022;31(6):e4353. https://doi.org/10.1002/prot.4353
Chakravarty S, Varadarajan R. Residue depth: a novel parameter for the analysis of protein structure and stability. Structure (London, England: 1993). 1999;7(7):723–32. https://doi.org/10.1016/S0969-2126(99)80097-5
Chandra S, Chattopadhyay G, Varadarajan R. Rapid identification of secondary structure and binding site residues in an intrinsically disordered protein segment. Front Genet. 2021;12:2173. https://doi.org/10.3389/fgene.2021.755292
Chandra S, Gupta K, Khare S, Kohli P, Asok A, Mohan SV, et al. The high mutational sensitivity of ccdA antitoxin is linked to codon optimality. Mol Biol Evol. 2022;39(10):msac187. https://doi.org/10.1093/molbev/msac187
Chao G, Lau WL, Hackel BJ, Sazinsky SL, Lippow SM, Wittrup KD. Isolating and engineering human antibodies using yeast surface display. Nat Protoc. 2006;1(2):755–68. https://doi.org/10.1038/nprot.2006.94
Combet C, Blanchet C, Geourjon C, Deléage G. NPS@: network protein sequence analysis. Trends Biochem Sci. 2000;25(3):147–50. https://doi.org/10.1016/S0968-0004(99)01540-6
Critchlow SE, O’Dea MH, Howells AJ, Couturier M, Gellert M, Maxwell A. The interaction of the F plasmid killer protein, CcdB, with DNA gyrase: induction of DNA cleavage and blocking of transcription. J Mol Biol. 1992;245(3):297–304. https://doi.org/10.1006/jmbi.1997.1357
Cunningham BC, Wells JA. High-resolution epitope mapping of the hGH-receptor interactions by alanine-scanning mutagenesis. Science (New York, N.Y.). 1989;244(4908):1081–5. https://doi.org/10.1126/science.2471267
Dao-Thi M-H, Charlier D, Loris R, Maes D, Messens J, Wyns L, et al. Intricate interactions within the ccd plasmid addiction
system. J Biol Chem. 2002;277(5):3733–42. https://doi.org/10.1074/jbc.M105050200

Dai-Thi MH, van Melderen L, de Genst E, Alfö H, Buts L, Wyns L, et al. Molecular basis of gyrase poisoning by the addiction toxin CcdB. J Mol Biol. 2005;348(5):1091–102. https://doi.org/10.1016/j.jmb.2005.03.049

Das R, Datta R, Varadarajan R. Probing the structure of the HIV-1 envelope trimmer using aspartate scanning mutagenesis. J Virol. 2020;94(21):e01426. https://doi.org/10.1128/JVI.01426-20

Datta R, Roy Chowdhury R, Manjunath K, Hanna LE, Varadarajan R. A facile method of mapping HIV-1 neutralizing epitopes using chemically masked cysteines and deep sequencing. Proc Natl Acad Sci U S A. 2020;117(47):29584–94. https://doi.org/10.1073/pnas.201256117

de Jonge N, García-Pino A, Buts L, Haesaerts S, Charlier D, Zanger K, et al. Rejuvenation of CcdB-poisoned gyrase by an intrinsically disordered protein domain. Mol Cell. 2009;35(2):154–63. https://doi.org/10.1016/j.molcel.2009.05.025

DeBartolo J, Dutta S, Reich L, Keating AE. Predictive Bcl-2 family binding models rooted in experiment or structure. J Mol Biol. 2012;422(1):124–44. https://doi.org/10.1016/j.jmb.2012.05.022

Dehouck Y, Kwasiqroch JM, Rooman M, Gilis D. BeATMuSiC: Prediction of changes in protein-protein binding affinity on mutations. Nucleic Acids Res. 2013;41(W333–9. https://doi.org/10.1093/nar/gkt450

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

Doolan KM, Colby DW. Conformation-dependent epitopes recognized by prion protein antibodies probed using mutational scanning and deep sequencing. J Mol Biol. 2015;427(2):328–40. https://doi.org/10.1016/j.jmb.2014.10.024

Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol. 2005;6(3):197–208. https://doi.org/10.1038/nrm1589

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

Ernst A, Gfeller D, Kan Z, Seshagiri S, Kim PM, Bader GD, et al. Coevolution of PDZ domain-ligand interactions analyzed by high-throughput phage display and deep sequencing. Mol Biostem. 2010;6(10):1782–90. https://doi.org/10.1093/cmb/mbq0061b

Forsyth CM, Juan V, Akamatsu Y, DuBridge RB, Doan M, Ivanov AV, et al. Deep mutational scanning of an antibody against epidermal growth factor receptor using mammalian cell display and massively parallel pyrosequencing. mAbs. 2013;5(4):523–32. https://doi.org/10.4161/mabs.24979

Fowler DM, Araya CL, Fleishman SJ, Kellogg EH, Stephany JJ, Baker D, et al. High-resolution mapping of protein-scene-function relationships. Nat Methods. 2010;7(9):741–6. https://doi.org/10.1038/nmeth.1492

Fowler DM, Fields S. Deep mutational scanning: a new style of protein science. Nat Methods. 2014;11(8):801–7. https://doi.org/10.1038/nmeth.3027

Fujino Y, Fujita R, Wada K, Fujishige K, Kanamori T, Hunt L, et al. Robust in vitro affinity maturation strategy based on interface-focused high-throughput mutational scanning. Biochem Biophys Res Commun. 2012;428(3):395–400. https://doi.org/10.1016/j.bbrc.2012.10.066

Gietz RD, Schiestl RH. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat Protoc. 2007;2(1):31–4. https://doi.org/10.1038/nprot.2007.13

Giri R, Morrone A, Toto A, Brunori M, Gianni S. Structure of the transition state for the binding of c-Myc and KIX highlights an unexpected order for a disordered system. Proc Natl Acad Sci. 2013;110(37):14942–947. https://doi.org/10.1073/pnas.1307337110

Gray VE, Hause RJ, Fowler DM. Using large-scale mutagenesis to guide single amino acid scanning experiments. bioRxiv. 2017;1:140707. https://doi.org/10.1101/140707

Heyne M, Pape N, Shifman JM. Generating quantitative binding landscapes through fractional binding selections combined with deep sequencing and data normalization. Nat Commun. 2020;11(1):297. https://doi.org/10.1038/s41467-019-13895-8

Hietpas RT, Jensen JD, Bolon DNA. Experimental illumination of a fitness landscape. Proc Natl Acad Sci. 2011;108(19):7896–901. https://doi.org/10.1073/pnas.1016024108

Hussain ASZ, Shanthi V, Sheik SS, Jeyakanthan J, Selvarani P, Sekar K. PDB Goodies: a web-based GUI to manipulate the Protein Data Bank file. Acta Crystallogr D Biol Crystallogr. 2002;58 (Pt 8):1385–6. https://doi.org/10.1107/s090744490200985x

Jenson JM, Xue V, Stretz L, Mandal T, Reich LL, Keating AE. Peptide design by optimization on a data-parameterized protein interaction landscape. Proc Natl Acad Sci U S A. 2018;115(44): E10342–51. https://doi.org/10.1073/pnas.1812939115

Jiang L, Mishra P, Hietpas RT, Zeldovich KB, Bolon DNA. Latent effects of Hsp90 mutants revealed at reduced expression levels. PLoS Genet. 2013;9(6):e1003600. https://doi.org/10.1371/journal.pgen.1003600

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):383–9. https://doi.org/10.1038/s41586-021-03819-2

Kanaya E, Kanaya S, Kikuchi M. Introduction of a non-native disulfide bridge to human lysozyme by cysteine scanning mutagenesis. Biochem Biophys Res Commun. 1990;173(3):1194–9. https://doi.org/10.1016/S0006-291X(05)80912-x

Kim MK, Kang YK. Positional preference of proline in alpha helices. Protein Sci. 1999;8(7):1492–9. https://doi.org/10.1101/p.s.8.7.1492

Kowalsky CA, Faber MS, Nath A, Dann HE, Kelly VW, Liu L, et al. Rapid fine conformational epitope mapping using comprehensive mutagenesis and deep sequencing. J Biol Chem. 2015;290(44):26457–70. https://doi.org/10.1074/jbc.M115.676635

Lacroix E, Viguera AR, Serrano L. Elucidating the folding problem rationally designed antibody platform for sequencing information of static accessibility. J Mol Biol. 1971;55(3):379–400. https://doi.org/10.1016/0022-2836(71)90324-x

Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. J Mol Biol. 1971;55(3):379–400. https://doi.org/10.1016/0022-2836(71)90324-x
protein-protein interactions. Nucleic Acids Res. 2016;44(W1): W494–501. https://doi.org/10.1093/nar/gkw374

Lindström I, Andersson E, Dogan J. The transition state structure for binding between TAIZ1 of CBP and the disordered Hif-1α CAD. Sci. Rep. 2018;8(1):7872. https://doi.org/10.1038/s41598-018-26213-x

Lipowsk D, Lippow SM, Hackel BJ, Gregson MW, Cheng P, Kapila A, et al. 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 camellia and shark antibodies. J Mol Biol. 2007; 368(4):1024–41. https://doi.org/10.1016/j.jmb.2007.02.029

Madl T, Van ML, Mine N, Respondek M, Oberer M, Keller W, et al. Structural basis for nucleic acid and toxin recognition of the bacterial anti toxin CcdA. J Mol Biol. 2006;364(2):170–85. https://doi.org/10.1016/j.jmb.2006.08.082

Maki S, Takiguchi S, Horiiuchi T, Sekimizu K, Miki T. Partner switching mechanisms in inactivation and rejuvenation of Escherichia coli DNA gyrase by F plasmid proteins LetD (CcdB) and LetA (CcdA). J Mol Biol. 1996;256(3):473–82. https://doi.org/10.1006/jmbi.1996.0102

Mata-Fink J, Kriegsman B, Yu HX, Zhu H, Hanson MC, Irvine DJ, et al. Rapid conformational epitope mapping of anti-gp120 antibodies with a designed mutant panel displayed on yeast. J Mol Biol. 2013;425(2):444–56. https://doi.org/10.1016/j.jmb.2012.11.010

Mateos B, Conrad-Billroth C, Schiavi M, Beier A, Kontaxis G, Konrat R, et al. The ambivalent role of proline residues in an intrinsically disordered protein: from disorder promoters to compaction facilitators. J Mol Biol. 2020;432(9):3093–41. https://doi.org/10.1016/j.jmb.2019.11.015

McLaughlin RNJ, Poelwijk FJ, Raman A, Gosal WS, Ranganathan R. The spatial architecture of protein function and adaptation. Nature. 2012;491(7422):138–42. https://doi.org/10.1038/nature11500

Melamed D, Young DL, Gamble CE, Miller CR, Fields S. Deep mutational scanning of an RRm domain of the Saccharomyces cerevisiae poly(A)-binding protein. RNA (N.Y., N.Y.). 2013;19(11):1537–51. https://doi.org/10.1261/rna.040709.113

Mészáros B, Erdős G, Dosztányi Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res. 2018;46(W1):W329–37. https://doi.org/10.1093/nar/gky384

Miki T, Yoshioka K, Horiiuchi T. 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. 1984;174(4):605–25. https://doi.org/10.1016/0022-2836(84)90086-X

Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColaFold: making protein folding accessible to all. Nat Methods. 2022;19(6):679. https://doi.org/10.1038/s41999-022-01488-1

Mittag T, Kay LE, Forman-Kay JD. Protein dynamics and conformational disorder in molecular recognition. J Mol Recog. 2010;23(2):105–16. https://doi.org/10.1002/jmr.961

Muñoz V, Serrano L. Elucidating the folding problem of helical peptides using empirical parameters. Nat Struct Biol. 1994;1(6):399–409. https://doi.org/10.1038/nsb0694-399

Najar TA, Khare S, Pandey R, Gupta SK, Varadarajan R. Mapping protein binding sites and conformational epitopes using cysteine labeling and yeast surface display. Structure. 2017;25(3):395–406. https://doi.org/10.1016/j.str.2016.12.016

Newberry RW, Leong JT, Chow ED, Kampmann M, DeGrado WF. Deep mutational scanning reveals the structural basis for α-synuclein activity. Nat Chem Biol. 2020;16(6):653–9. https://doi.org/10.1016/j.chembiol.2020.02.0480-6

Pál G, Fong S-Y, Kossiakoff AA, Sidhu SS. Alternative views of functional protein binding epitopes obtained by combinatorial shotgun scanning mutagenesis. Protein Sci. 2005;14(9):2405–13. https://doi.org/10.1110/ps.051519805

Park S, Xu Y, Stowell XF, Gai F, Saven JG, Boder ET. Limitations of yeast surface display in engineering proteins of high thermostability. Protein Eng Des. Select. 2006;19(5):211–7. https://doi.org/10.1093/protein/gz003

Rathore U, Purwar M, Vignesh VS, Das R, Kumar AA, Bhattacharyya S, et al. Bacterially expressed HIV-1 gp120 outer-domain fragment immunogens with improved stability and affinity for CD4-binding site neutralizing antibodies. J Biol Chem. 2018;293(39):15002–20. https://doi.org/10.1074/jbc.RA118.005006

Ravn U, Dídelot G, Venet S, Ng K-T, Gueneau F, Rousseau F, et al. Deep sequencing of phage display libraries to support antibody discovery. Methods (San Diego, CA). 2013;60(1):99–110. https://doi.org/10.1016/j.ymeth.2013.03.001

Reich LL, Dutta S, Keating AE. SORTCERY: A high-throughput method to affinity rank peptide ligands. J Mol Biol. 2015;427(11):2135–50. https://doi.org/10.1016/j.jmb.2014.09.025

Reich LL, Dutta S, Keating AE. Generating high-accuracy peptide-binding data in high throughput with yeast surface display and SORTCERY. Methods in molecular biology. Vol 1414. Clifton, N.J.: Humana Press; 2016. p. 233–47. https://doi.org/10.1007/978-1-4939-3569-7_14

Rodrigues CHM, Myung Y, Pires DE V, Ascher DB. mCSM-PP12: predicting the effects of mutations on protein–protein interactions. Nucleic Acids Res. 2019;47(W1):W338–44. https://doi.org/10.1093/nar/gky283

Rogers JM, Wong CT, Clarke J. Coupled folding and binding of the disordered protein PUMA does not require particular residual structure. J Am Chem Soc. 2014;136(14):5197–200. https://doi.org/10.1021/ja4125065

Roscoe BP, Thayer KM, Zeldovich KB, Fushman D, Bolon DNA. Analyses of the effects of all ubiquitin point mutants on yeast growth rate. J Mol Biol. 2013;425(8):1363–77. https://doi.org/10.1016/j.jmb.2013.01.032

Ruff KM, Pappu RV. AlphaFold and implications for intrinsically disordered proteins. J Mol Biol. 2021;433(20):167208. https://doi.org/10.1016/j.jmb.2021.167208

Staller MV, Holehouse AS, Swain-Lenz D, Das RK, Pappu RV, Cohen BA. A high-throughput mutational scan of an intrinsically disordered acidic transcriptional activation domain. Cell Syst. 2018;6(4):444–455.e6. https://doi.org/10.1016/j.cels.2018.01.015

Starita LM, Pruneda JN, Lo RS, Fowler DM, Kim HJ, Hiatt JB, et al. Activity-enhancing mutations in an E3 ubiquitin ligase identified by high-throughput mutagenesis. Proc Natl Acad Sci U S A. 2013;110(14):E1263–72. https://doi.org/10.1073/pnas.1303309110

Strobel B. Energy landscapes of protein aggregation and conformation switching in intrinsically disordered proteins. J Mol Biol. 2021;433(20):167182. https://doi.org/10.1016/j.jmb.2021.167182
Swers JS, Kellogg BA, Wittrup KD. Shuffled antibody libraries created by in vivo homologous recombination and yeast surface display. Nucleic Acids Res. 2004;32(3):e36. https://doi.org/10.1093/nar/gnh030

Tan KP, Varadarajan R, Madhusudhan MS. DEPTH: a web server to compute depth and predict small-molecule binding cavities in proteins. Nucleic Acids Res. 2011;39(suppl. 2):1–7. https://doi.org/10.1093/nar/gkq1235

Tripathi A, Gupta K, Khare S, Jain PC, Patel S, Kumar P, et al. Molecular determinants of mutant phenotypes, inferred from saturation mutagenesis data. Mol Biol Evol. 2016;33(11):2960–75. https://doi.org/10.1093/molbev/msw182

Uchański T, Zögg T, Yin J, Yuan D, Wohlkönig A, Fischer B, et al. An improved yeast surface display platform for the screening of nanobody immune libraries. Sci Rep. 2019;9(1):382. https://doi.org/10.1038/s41598-018-37212-3

Uversky VN. The triple power of $D^3$: protein intrinsic disorder in degenerative diseases. Front Biosci. 2014;19:181–258. https://doi.org/10.2741/4204

Uversky VN, Dunker AK. Understanding protein non-folding. Biochim Biophys Acta. 2010;1804(6):1231–64. https://doi.org/10.1016/j.bbapap.2010.01.017

Uversky VN, Oldfield CJ, Dunker AK. Intrinsically disordered proteins in human diseases: introducing the D2 concept. Annu Rev Biophys. 2008;37:215–46. https://doi.org/10.1146/annurev.biophys.37.032807.125924

Van Blarcom T, Rossi A, Foletti D, Sundar P, Pitts S, Bee C, et al. Precise and efficient antibody epitope determination through library design, yeast display and next-generation sequencing. J Mol Biol. 2015;427(6):1513–34. https://doi.org/10.1016/j.jmb.2014.09.020

van Melderen L, Thi MH, Lecchi P, Gottesman S, Couturier M, Maurizi MR. ATP-dependent degradation of CcdA by Lon protease. Effects of secondary structure and heterologous subunit interactions. J Biol Chem. 1996;271(44):27730–8. https://doi.org/10.1074/jbc.271.44.27730

Vandervelde A, Drobnak I, Hadˇ S, Sterckx YG, Welte T, De GH, et al. NAR breakthrough article molecular mechanism governing ratio-dependent transcription regulation in the ccdAB operon. Nucleic Acids Res. 2017;45(6):2937–50. https://doi.org/10.1093/nar/gkx108

Whitehead TA, Chevalier A, Song Y, Dreyfus C, Fleishman SJ, de Mattos C, et al. Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. Nat Biotechnol. 2012;30(6):543–8. https://doi.org/10.1038/nbt.2214

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Chandra S, Manjunath K, Asok A, Varadarajan R. Mutational scan inferred binding energetics and structure in intrinsically disordered protein CcdA. Protein Science. 2023;32(3):e4580. https://doi.org/10.1002/pro.4580