Received: 31 January 2022 | Revised: 4 April 2022 | Accepted: 11 May 2022

DOI: 10.1002/pro.4357

FULL-LENGTH PAPER

Structural and functional determinants inferred from deep mutational scans

Priyanka Bajaj[1] | Kavyashree Manjunath[2] | Raghavan Varadarajan[1]

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

[2]Centre for Chemical Biology and Therapeutics, Institute for Stem Cell Science and Regenerative Medicine, Bangalore, India

Correspondence
Raghavan Varadarajan, Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India.
Email: varadar@iisc.ac.in

Funding information
Department of Science and Technology and Department of Biotechnology, Grant/Award Numbers: BT/COE/34/SP15219/2015 DT. 20/11/2015, EMR/2017/004054 DT.15/12/2018

Review Editor: John Kuriyan

Abstract

Mutations that affect protein binding to a cognate partner primarily occur either at buried residues or at exposed residues directly involved in partner binding. Distinguishing between these two categories based solely on mutational phenotypes is challenging. The bacterial toxin CcdB kills cells by binding to DNA Gyrase. Cell death is prevented by binding to its cognate antitoxin CcdA, at an extended interface that partially overlaps with the GyrA binding site. Using the CcdAB toxin–antitoxin (TA) system as a model, a comprehensive site-saturation mutagenesis library of CcdB was generated in its native operonic context. The mutational sensitivity of each mutant was estimated by evaluating the relative abundance of each mutant in two strains, one resistant and the other sensitive to the toxic activity of the CcdB toxin, through deep sequencing. The ability to bind CcdA was inferred through a RelE reporter gene assay, since the CcdAB complex binds to its own promoter, repressing transcription. By analyzing mutant phenotypes in the CcdB-sensitive, CcdB-resistant, and RelE reporter strains, it was possible to assign residues to buried, CcdA interacting or GyrA interacting sites. A few mutants were individually constructed, expressed, and biophysically characterized to validate molecular mechanisms responsible for the observed phenotypes. Residues inferred to be important for antitoxin binding, are also likely to be important for rejuvenating CcdB from the CcdB–Gyrase complex. Therefore, even in the absence of structural information, when coupled to appropriate genetic screens, such high-throughput strategies can be deployed for predicting structural and functional determinants of proteins.

KEYWORDS

active-site, fitness, gene regulation, protein structure prediction, residue burial

1 | INTRODUCTION

The amino acid sequence dictates the tertiary structure of a protein which is closely tied to its activity. From a wealth of previous studies, it is known that mutations that affect function primarily occur either at exposed active site/ligand binding residues or at buried sites important for protein stability.[1,2] However, distinguishing them, purely from mutational phenotypes is challenging. Several computational approaches exploit sequence-structure relationship to predict functional patches on the protein surface through in silico modeling based on a query protein sequence, sequence conservation,[3] or structural homology with well-characterized proteins.[4] This is challenging for proteins with low sequence identity to those present in sequence databases.[5] Traditional methods are low-throughput and require purification and characterization of large numbers of individual variants to identify active site residues such as protein:protein and protein:ligand binding site residues.[6] X-ray crystallography, NMR spectroscopy, and cryoelectron microscopy of a protein complex can be used to identify residues important in binding to its interacting partner, a prerequisite is high yield and homogenous preparation of purified protein.[7–9] While these methods are useful in providing atomic level information but are labor intensive and not easily parallelizable. The advent of next generation sequencing has revolutionized biology,[1,10] resulting in the development of various approaches that can be used to predict structural features in the absence of a structure.[2,11] Alanine scanning and cysteine scanning mutagenesis approaches have been previously used to predict functional residues but have limitations. Alanine scanning mutagenesis is laborious as each alanine-substituted protein needs to be individually expressed and characterized.[12] Cysteine scanning mutagenesis requires an additional step of labeling the exposed residues, effects, and labeling of a buried cysteine may result in misfolding of the protein, thereby resulting in production of false positive results.[13] Generally, buried residues as well as active site residues are sensitive to mutations but differ in their mutational tolerance to aliphatic and charged substitutions.[2] Deep mutational scanning is a promising tool for mapping sequence–activity relationships in proteins for which an observable phenotypic readout is available. Mutations at the active site generally affect the specific activity of the protein as the native conformation remains intact, while buried site mutants affect the stability and folding of the protein, thereby affecting the total activity of the protein. However, distinguishing between these two classes of residues solely from phenotypic data is challenging.[14] The situation is even more complex for proteins with multiple binding partners.

The system used in the present study is a *ccd* operon. This is a Type II toxin–antitoxin (TA) system and both *ccdA* and *ccdB* genes encode proteins. CcdA acts as an antidote which neutralizes the toxicity of CcdB by forming a tight complex with it, and also rejuvenating Gyrase from its complex with CcdB.[15] The $(CcdA)_2$–$(CcdB)_2$ complex represses the operon at the transcriptional level by binding to the operator–promoter region of the operon to maintain the CcdB:CcdA ratio < 1.[16] Under stress conditions, CcdA is degraded and the CcdB:CcdA ratio is increased. This causes transient derepression of the operon and fresh synthesis of both CcdA and CcdB (Figure S1). The crystal structure, 3G7Z, shows that the C-terminal intrinsically disordered region of CcdA binds consecutively to two overlapping sites of CcdB with different affinities. Both sites are important for rejuvenation and autoregulation of expression of the *ccd* operon.[17,18] Promoter–operator binding of the CcdA–CcdB complex in vivo can be probed by co-expressing the complex and a reporter gene downstream of the *ccd* promoter within the cell. Dual selection reporter systems such as the tetA gene,[19] tetA-sacB cassette,[20] kill gene,[21] and RelE gene[22] have previously been used for genome recombineering studies. The RelE reporter system is reported to achieve higher selection stringency than previously reported negative selection systems, usable in *E. coli* strains.[22]

CcdAB is a convenient system to study mutational effects in an operonic context. In this report, we describe comprehensive single-site mutational scanning of CcdB in its operonic context, with the goal of determining multiple binding sites and, distinguishing active site residues from the buried site and exposed non-active-site residues. We attempt to address the following issues: (1) Is identifying active site residues of CcdB solely from mutational phenotypes possible? (2) Can we distinguish Gyrase binding site residues from buried residues from mutational phenotypes? (3) How well does the amount of accessible surface area buried upon complex formation with the interacting partner explain the mutational landscape of active site residues? (4) Is there any consistent pattern in substitution preferences at the active site residues? (5) Can we delineate molecular mechanisms behind the observed phenotypes in a high-throughput manner? (6) How can the inferred molecular mechanisms be validated? (7) Are the CcdA interacting residues impaired in CcdA binding also important for rejuvenating CcdB from the CcdB–Gyrase complex. Crystal structures of CcdB complexed to CcdA (PDBid: 3G7Z)[17] and to DNA Gyrase (PDBid: 1X75)[23] were used to rationalize mutant phenotypes obtained from deep sequencing.

2 | RESULTS

2.1 | Deep sequencing of CcdB SSM library in operonic context

A comprehensive site-saturation mutagenesis (SSM) library of CcdB was prepared in its native operon that contained the promoter, *ccdA* and *ccdB* genes in pUC57 plasmid (a high copy number plasmid), to get an amplified response required to distinguish the mutant phenotype from WT. The pooled mutant library of CcdB, transformed in two strains, one resistant (resistant strain) and the other sensitive (sensitive strain) to the toxic activity of CcdB (Figure S2a), was subjected to deep sequencing and each mutant analyzed was assigned a variant score in the form of "Relative Fitness^{CcdB}" (RF^{CcdB}) (Equation 6, see methods). Out of ~3,200 (100 positions*32 codons) mutants expected by NNK codon mutagenesis of the *ccdB* gene, reads for ~2,700 mutants were available in the resistant strain (unselected library). Deep sequencing data from the two biological replicates were compared using different read cut-offs (Figure S2b). The highest correlation of ~0.96 between the two was obtained when the read cut-off in the resistant strain was taken as 100 (Figure 1a). CcdB mutants were ranked based on their activity, for which the phenotypic readout is cell growth versus cell death. We first validated the deep sequencing data in a high-throughput manner by overlaying the RF^{CcdB} scores obtained by all CcdB mutants with synonymous CcdB mutants and non-functional CcdB mutants (Figure 1b). The synonymous mutant dataset represents an internal positive control with the median value of ~0.8 and the major fraction lie within a twofold range of the WT score (RF^{CcdB} = 1), that is, between 0.5 and 2, suggesting these mutants show a phenotype similar to WT. In contrast, stop codon mutant dataset represents an internal negative control with median value of 16 and most non-sense mutations exhibited inactive phenotypes (RF$^{CcdB} > 2$).

The RF^{CcdB} of the synonymous mutants mainly ranges from 0.5 to 2, a reason to consider mutants having RF$^{CcdB} < 0.5$ or RF$^{CcdB} > 2$ to be deviated from the WT phenotype. We classified the mutants into three classes that is, hyperactive, neutral, and inactive based on the variant score, that is, RF^{CcdB} less than 0.5 is "hyperactive," between 0.5 and 2 is "neutral," and more than 2 is "inactive." A similar cut-off for the hyperactive mutant class was also found using a *k*-means clustering algorithm (Figure S2c). The entire dataset was also divided into four structural categories (1) CcdA interacting residues, (2) only Gyrase binding residues (excludes the overlapping CcdA interacting residues), (3) buried site residues, and (4) exposed non-active-site residues (excludes both CcdA interacting and Gyrase binding site residues) to comprehend functional properties of the mutants in these regions (Figure 1c). The RF^{CcdB} scores for CcdA interacting mutants (Median = 0.5) were significantly different from all other classes (two-tailed *t*-test, $p < .001$). A significant difference exists between the mean RF^{CcdB} values for the CcdA interacting site mutants and the exposed CcdA non-interacting site mutants (median = 1.5) (two-tailed *t*-test, $p < .001$). However, RF^{CcdB} scores for buried site (median = 3) and only Gyrase binding site mutants (median = 2) lie in a similar range that is, primarily giving an inactive phenotype. This kind of screening distinguishes CcdA interacting residues from other classes but buried residues cannot be discriminated from Gyrase binding site residues.

To ensure that each mutant amino acid contributes equally to the overall average of the RF^{CcdB} levels for each position, we first averaged the RF^{CcdB} scores for all synonymous mutants of each mutant amino acid and then further averaged over all the mutant amino acids at each position (Figure 1d). Only CcdA interacting residues have avgRF$^{CcdB} < 0.5$, suggesting that mutations at the CcdA binding site result in the severest phenotypes. Mutational analysis indicates 52% (475/920) of the hyperactive mutants belong to the CcdA interacting site, 42% (389/920) to the exposed CcdA noninteracting class, and 6% (56/920) were buried. Among exposed CcdA non-interacting class, 30% (275/920) lie proximal (within 8 Å) to CcdA residues, whereas 12% (114/920) lie distal to it. *K*-means clustering, that divided the dataset into 2 clusters also point toward CcdA interacting mutant enrichment in cluster 2 (Figure S2c).

2.2 | In vivo activity and in vivo solubility assays mirror deep sequencing data

Phenotypes of selected mutants inferred from deep sequencing data were validated by individual transformations of the point mutants and plating in both CcdB-resistant and -sensitive strains (Figure 2a). Plasmids containing the WT operon and an operon with non-functional toxin were used as controls for inferring the phenotype of the mutant relative to the WT, and for accounting for transformation efficiency differences between the two strains, respectively. The WT construct used in this study has a mutation in the putative SD sequence of CcdA because of a restriction site introduced to facilitate cloning of the *ccdA* coding region of the operon (Figures S3a and 2a). The growth in the sensitive strain (Top 10) of the WT construct used in this study was compared with a construct without any mutations in the promoter and identical to that present in F plasmid (WTF′ in Figures S3b and 2a). We found that the construct without any mutations in the promoter grew similar to a construct with a stop codon mutation in the toxin gene (Y6_TAA in Figure 2a), whereas the present WT construct grew more poorly compared to the construct with a non-functional toxin. This indicates that modification of the putative SD sequence of CcdA, likely reduces the expression of CcdA, thereby showing reduced growth and increased toxicity relative to the operon present in F plasmid. The construct exhibiting higher toxicity is used as the WT because it enables to screen both inactive as well as hyperactive mutants (Figures 2a and S3a). Deep sequencing results were validated by the growth phenotypes shown by 25 individual CcdB mutants in the

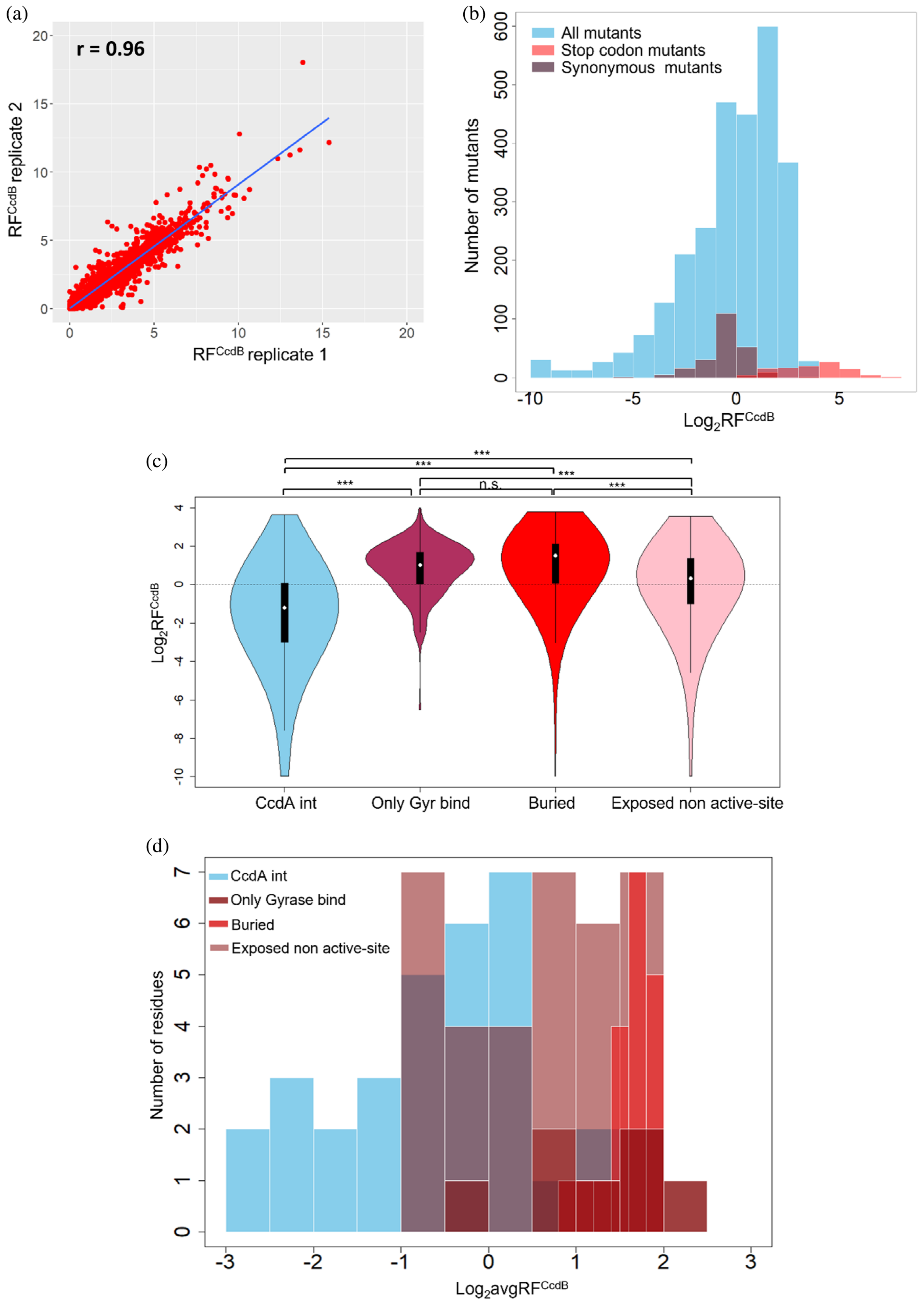

FIGURE 1 Legend on next page.

sensitive strain, which were in concordance to their inferred RF^{CcdB} scores (Figure 2a). CcdB mutants displaying a hyperactive phenotype did not grow in the sensitive strain suggesting these mutants show higher toxicity than the WT. P72L which has a RF^{CcdB} score of 0.55 shows no growth in the sensitive strain, whereas S47V with a score of 0.6 grows weakly in the sensitive strain. This supports use of RF^{CcdB} cut-off score of 0.5 that was chosen to classify mutants as hyperactive.

An in vivo solubility assay was conducted for a subset of CcdB mutants (Figure 2b). In this assay, CcdB was expressed in the absence of CcdA, under control of the pBAD24 promoter.[24] The percentage solubility of CcdA interacting mutants (S12G, R13G, D26A, and P28A) was comparable to the WT CcdB protein. L16, a synonymous mutant, also showed in vivo solubility comparable to the WT. V18R a buried mutant, likely due to its misfolded nature, mainly was found in inclusion bodies. Although Q21S is a buried mutant, it showed a hyperactive phenotype both in the deep sequencing data as well as in an individual spotting assay, and showed in vivo solubility comparable to WT. Percentage solubility of G57M and E87P that had shown inactive phenotypes in the deep sequencing data, decreased, with 20% and 50% of the induced fractions for G57M and E87P, respectively, targeted to the pellet fraction (Figure 2b). Thus, the in vivo solubility assay is consistent with deep sequencing results. For G57M, a substantial fraction is present in the soluble fraction. However, this is likely to be misfolded, given the positive phi value of G57 (also see SPR data below). While there is some variation in absolute levels of expression for different mutants in Figure 2b, this does not affect the primary result that is, determination of the relative fraction of protein targeted to inclusion bodies for each mutant, which is likely to be a consequence of mutant induced destabilization and reduced folding rate.[2]

2.3 | RF^{RelE} and RF^{CcdB} phenotypic scores provide structural insights

We engineered a construct with the *ccd* promoter upstream of the reporter gene that is, RelE, a toxin of the RelBE TA operon.[25] We standardized the reporter assay by manipulating the putative Shine Dalgarno sequence to modulate the expression of RelE toxin (see methods in Supporting Information) (Figure S4a,b). A significant difference between the growth levels of Top10Gyr co-transformed with WT *ccdAB* operon and consensus RelE, relative to co-transformation with a mutant *ccdAB* operon containing a stop codon CcdB mutant and consensus RelE was observed at 37°C, in both LB media (two fold difference) and minimal media (five fold difference) (Figure S4c). At 42 and 45°C, the reporter was less sensitive possibly because heat shock induced chaperones might help in folding of CcdA, enhancing its binding to the promoter/operator region even in the absence of CcdB, thus decreasing the difference in the growth between the WT and CcdB stop codon mutant. Since the CcdB mutants should not affect the amount of RelB antitoxin produced in the cells, we transformed the CcdB NNK library in the background of the RelE reporter gene in Top10Gyr strain, to avoid any toxic effects resulting directly from the CcdB toxin. In this system, RelE expression and toxicity are regulated by the level of binding of the CcdAB complex to the operator upstream of the RelE reporter. The reporter thus provides a measure of the amount of CcdAB complex within the cell (Figure 3a). Each CcdB variant was assigned a variant score defined as "Relative $Fitness^{RelE}$" (RF^{RelE}) (Equation 7, see methods). A high correlation ($r = \sim.94$) was found between the two biological replicates of the CcdB NNK libraries prepared in the RelE reporter strain using a read cut-off of 100 in the resistant strain, indicating that sequencing errors have largely been removed from the analysis (Figure 3b). The dynamic range of RF^{RelE} scores is much lower than RF^{CcdB} scores and the RF^{RelE} score of WT is 1. Thus, a variant with $RF^{RelE} < 1$ was considered to have increased RelE toxicity, suggesting that the variant has a lower amount of CcdAB complex relative to cells expressing WT CcdAB. A variant with a variant score of more than 1 was considered to have similar to higher levels of CcdAB, relative to cells expressing WT CcdAB. Additionally, a similar cut-off was obtained when the entire dataset was divided into two clusters by k-means clustering algorithm (Figure S5).

FIGURE 1 Reproducibility and distribution of Relative $Fitness^{CcdB}$ (RF^{CcdB}) values of the entire dataset. (a) Correlation between RF^{CcdB} values for the two biological replicates for mutants with read cut-off greater than 100 in the resistant strain Top10Gyr. (b) Histogram of RF^{CcdB} for all the mutants in the entire dataset (blue) with synonymous mutants (gray) and stop-codon mutants (pink). Overlapping region between the two classes is shown in light purple. (c) Violin plot with width proportional to the number of mutants at a given RF^{CcdB} value found in each class (mentioned on the *x*-axis). White dot in the middle of the box plot represents the median, box indicates the interquartile range. Black dashed line represents the WT value. *** denotes *p*-value between the two datasets is statistically significant (two-tailed *t*-test, $p < .001$). If the difference is not significant, it is represented by n.s. The null hypothesis states that the hypothesized mean difference between the two datasets for which the *p*-value is calculated is zero, implying the RF^{CcdB} values for the two datasets are similar. (d) Frequency distribution of the residue averaged Relative $Fitness^{CcdB}$ scores.

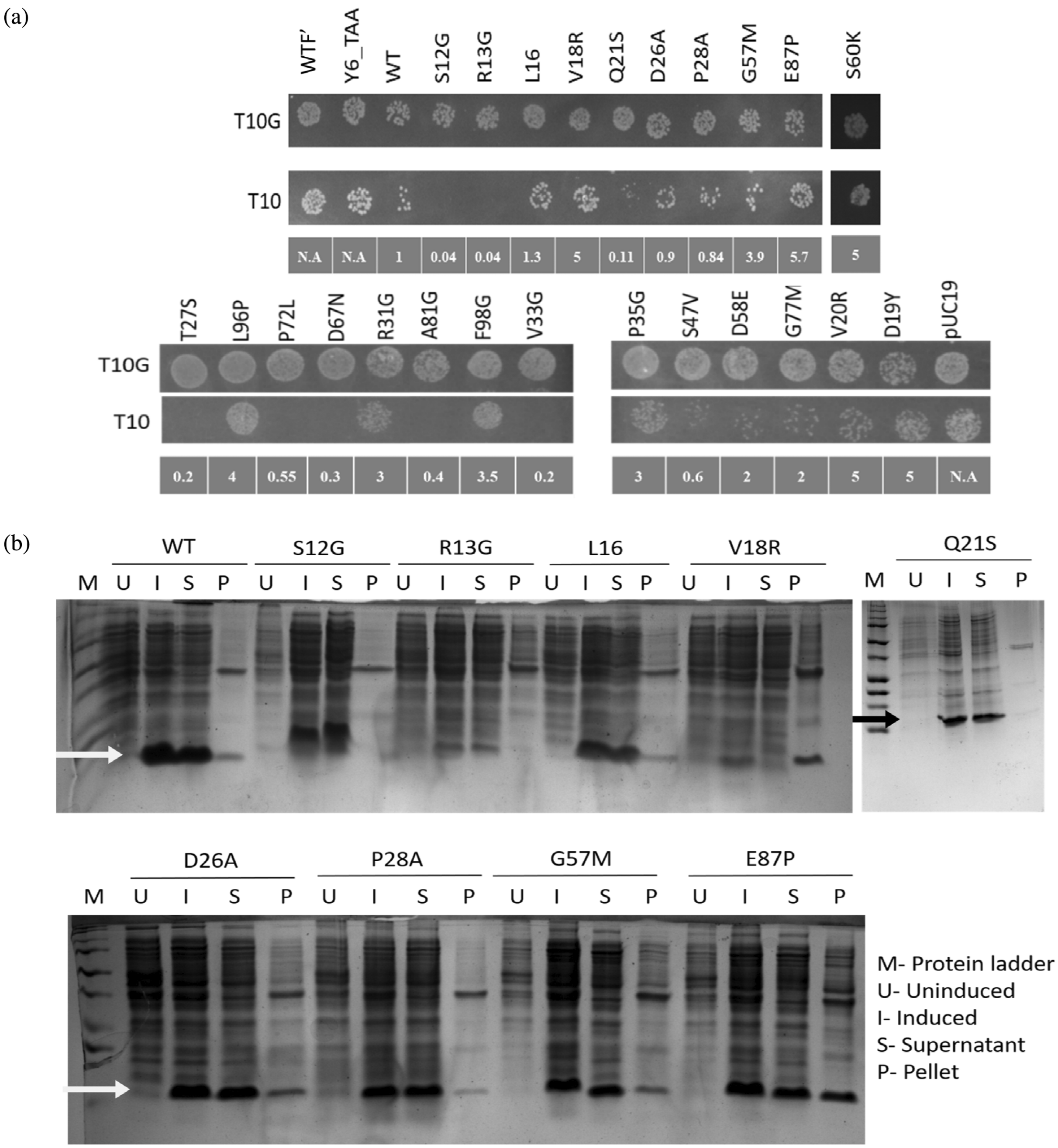

FIGURE 2 In vivo activity and in vivo solubility of CcdB mutants validate deep sequencing results. (a) CcdB mutants with different RF^{CcdB} values are selected for individually spotting on LBamp plates. RF^{CcdB} for each mutant spotted has been indicated in the panel below each Top10 panel. N.A is mentioned for controls. WTF' are Top10 cells containing the F-plasmid, which naturally possesses a WT ccd operon. Y6_TAA represents Top 10 cells transformed with a control plasmid in which residue Y6 of *ccdB* is replaced with a TAA stop codon. (b) In vivo solubility is estimated from the relative fractions of protein in supernatant and pellet, determined by densitometric analysis, following 15% SDS–PAGE. The arrow indicates the protein of interest (CcdB). The relative estimates of protein present in the soluble fraction and inclusion bodies for all mutants are shown in Table S3. These experiments were performed in duplicates.

There are 101 residues in CcdB. The first six residues were eliminated because they were used for cloning the previously generated site-saturation mutagenesis library[26] in the operonic context. Here, we classify the remaining residues based on four structural categories, that is, 33 CcdA interacting, 7 only Gyrase binding (positions that interact exclusively with GyrA), 19 buried and 36 exposed non-active-site residues. The mutants were further classified into four mutational categories based on the RF^{CcdB} and the RF^{RelE} levels defined as (1) hyperactive and derepressing ($RF^{CcdB} < 0.5$ and $RF^{RelE} < 1$), (2) inactive and repressing ($RF^{CcdB} > 2$ and $RF^{RelE} > 1$),

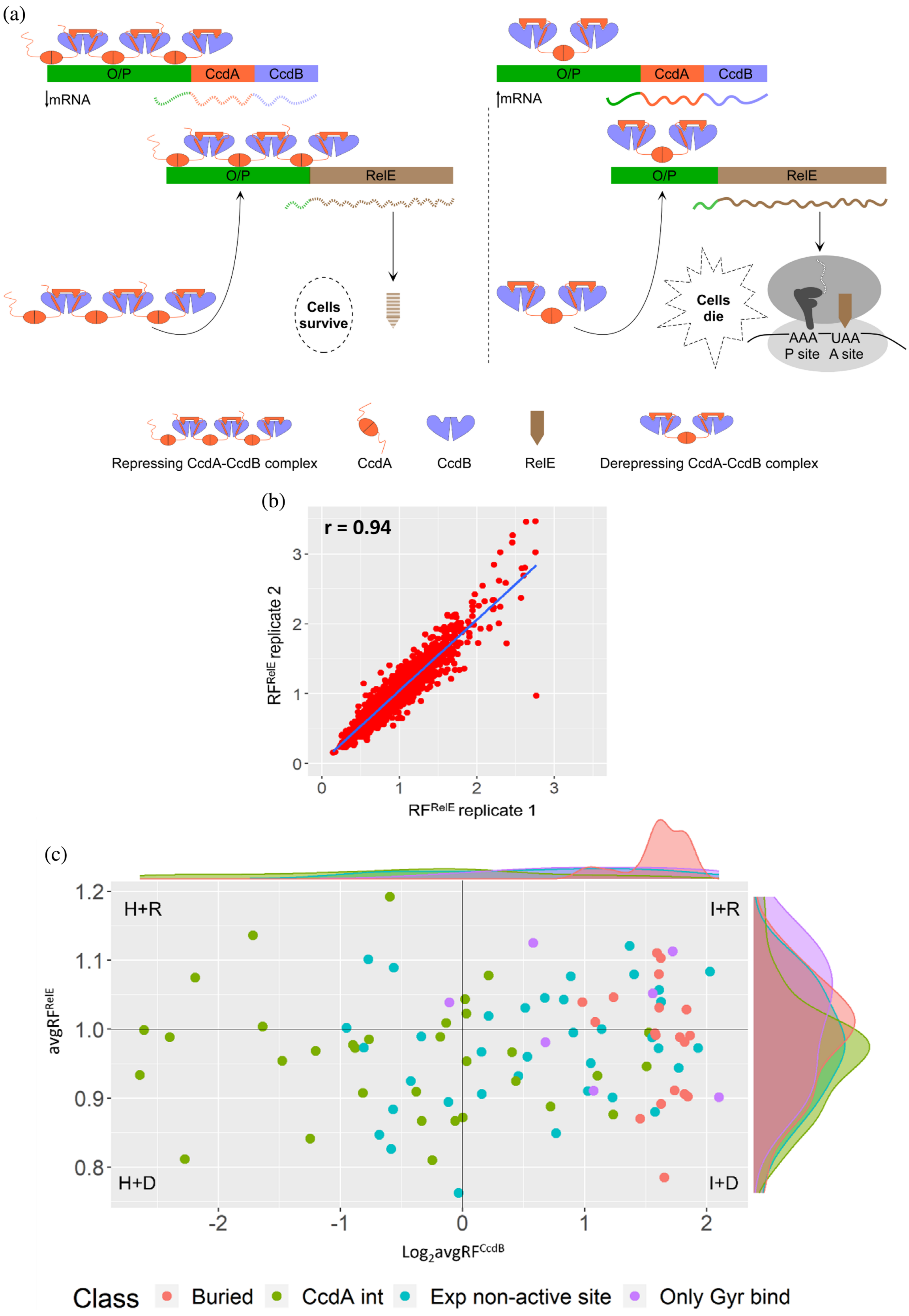

FIGURE 3 Legend on next page.

(3) inactive and derepressing ($RF^{CcdB} > 2$ and $RF^{RelE} < 1$), and (4) hyperactive and repressing ($RF^{CcdB} < 0.5$ and $RF^{RelE} > 1$) (Table S1). 60% (223/369) of the mutants belonging to the "hyperactive and derepressing" ($RF^{CcdB} < 0.5$ and $RF^{RelE} < 1$) class fall at the CcdA binding interface. 31% (114/369) of the mutants belonging to this class are exposed CcdA non-interacting mutants, of which 19% (69/369) mutants are proximal to the CcdA chains whereas 12% (45/369) are distal to it. The remaining 7% (26/369) are buried mutants. Enrichment of CcdA interacting mutants and mutants proximal to CcdA chains in this class indicate that these mutants are impaired in binding to CcdA. Thirty out of a total of 33 residues at the CcdA binding site are found in this class. We expected that residue positions which have high values of both RF^{CcdB} and RF^{RelE} would be enriched at residues exclusively involved in Gyrase binding, since mutations at such sites would not impair CcdA binding but should result in decreased toxicity of any free CcdB that might be present. Surprisingly, Gyrase binding site mutants (14%) were not predominantly enriched in this "inactive and repressing" class (Table S1). The buried site mutants (36%) dominate over other structural categories in the "inactive and derepressing" class (Table S1), implying that several mutations at buried residues result in misfolding of the protein which in turn hampers its binding to GyrA as well as CcdA. Interestingly, CcdA interacting residues (62%) are also enriched in the "hyperactive and repressing" class (Table S1).

To reduce noise in the data, we further averaged all the RF^{CcdB} and RF^{RelE} variant scores for each position to yield position averaged scores $avgRF^{CcdB}$ and $avgRF^{RelE}$, respectively (Figure 3c). Only CcdA interacting residues (7/7) are enriched in the class with $avgRF^{CcdB} < 0.5$ and $avgRF^{RelE} < 1$ (Figure 3c), suggesting that the majority of mutations at CcdA interacting sites impede CcdB binding to CcdA, resulting in an enhanced amount of free CcdB in vivo, thereby causing cell death. 44% (11/25) of the residues found in the class with $avgRF^{CcdB} > 2$ and $avgRF^{RelE} < 1$ are buried site residues (Figure 3c). A large number of buried site mutations were well tolerated in the operonic context (Figure 3c), likely because CcdA was able to relieve the folding defect of the protein. However, further investigation is required. 14% (2/14) of residues belonging to the inactive and repressing class with $avgRF^{CcdB} > 2$ and $avgRF^{RelE} > 1$, are Gyrase binding site residues (Figure 3c). However, although mutations at both the Gyrase binding site and buried site display an inactive phenotype in Top10, we can still discriminate several Gyrase binding site residues from buried residues as most Gyrase binding site mutants of CcdB repress RelE expression more efficiently than most buried site mutants, presumably because buried site mutants reduce the amount of properly folded CcdB that can complex with CcdA.

2.4 | Structural and mutational data help delineate residue specific contributions to binding

Examination of the phenotypes displayed by CcdB mutants in the sensitive strain revealed that CcdA interacting mutants primarily displayed a hyperactive phenotype (Figure S6a, CcdA interacting residues highlighted with an * and Figure 4a). Based on the cell growth phenotype analyzed for the same class of mutants in the RelE reporter strain, the mutants were classified into two categories, repressing ($RF^{RelE} > 1$) and derepressing ($RF^{RelE} < 1$) the RelE reporter gene expression. Mutants with $RF^{RelE} < 1$ were likely defective in binding to CcdA, whereas the mutants with $RF^{RelE} > 1$ are associated with larger amounts of the CcdAB complex relative to WT (Figure S6b). Mutations in the 8–14, 23–30, 41–46 and 64–72 residue stretches are enriched in the "hyperactive and derepressing" class (Figure S6). The enrichment is clearer for the RF^{CcdB} data, because of the larger dynamic range of this parameter. However the overall enrichment of CcdA interacting residues in this category is also apparent in Figure 3c. Mutants at residues 12, 13, 28, 30, 42, 43, 46, and 66, CcdA interacting positions are most enriched in this class (>10) (Figure 4a,b). Residues

FIGURE 3 RelE reporter assay to quantitate the relative amounts of CcdA:CcdB complex for different mutants. (a) Schematic of the RelE reporter system. When CcdA is in excess of CcdB (left panel), a repressing complex is formed that binds to the ccd O/P, repressing transcription of the RelE reporter. Decreased mRNA levels and subsequent decrease in RelE is depicted by dashed lines. When CcdA is not in excess of CcdB (right panel), a derepressing complex is formed, resulting in increased transcription of the RelE toxin, causing cell death. Increased mRNA levels and subsequent increase in RelE is depicted by solid lines. (b) Correlation between the two biological replicates for each CcdB mutant for RF^{RelE} values. (c) Corresponding values of Relative $Fitness^{CcdB}$ and Relative $Fitness^{RelE}$ scores averaged over all mutants for each CcdB position belonging to four different classes that is, Buried site residues (Buried), CcdA interacting site residues (CcdA int), exposed non-active-site residues (Exp non-active-site) and only Gyrase binding residues (only Gyr bind). The top horizontal panel is the density spread for different classes based on the $avgRF^{CcdB}$ scores. The right vertical panel depicts the residue averaged density spread for different classes based on the $avgRF^{RelE}$ scores. The black lines represent WT values and divide the graph into four quadrants.

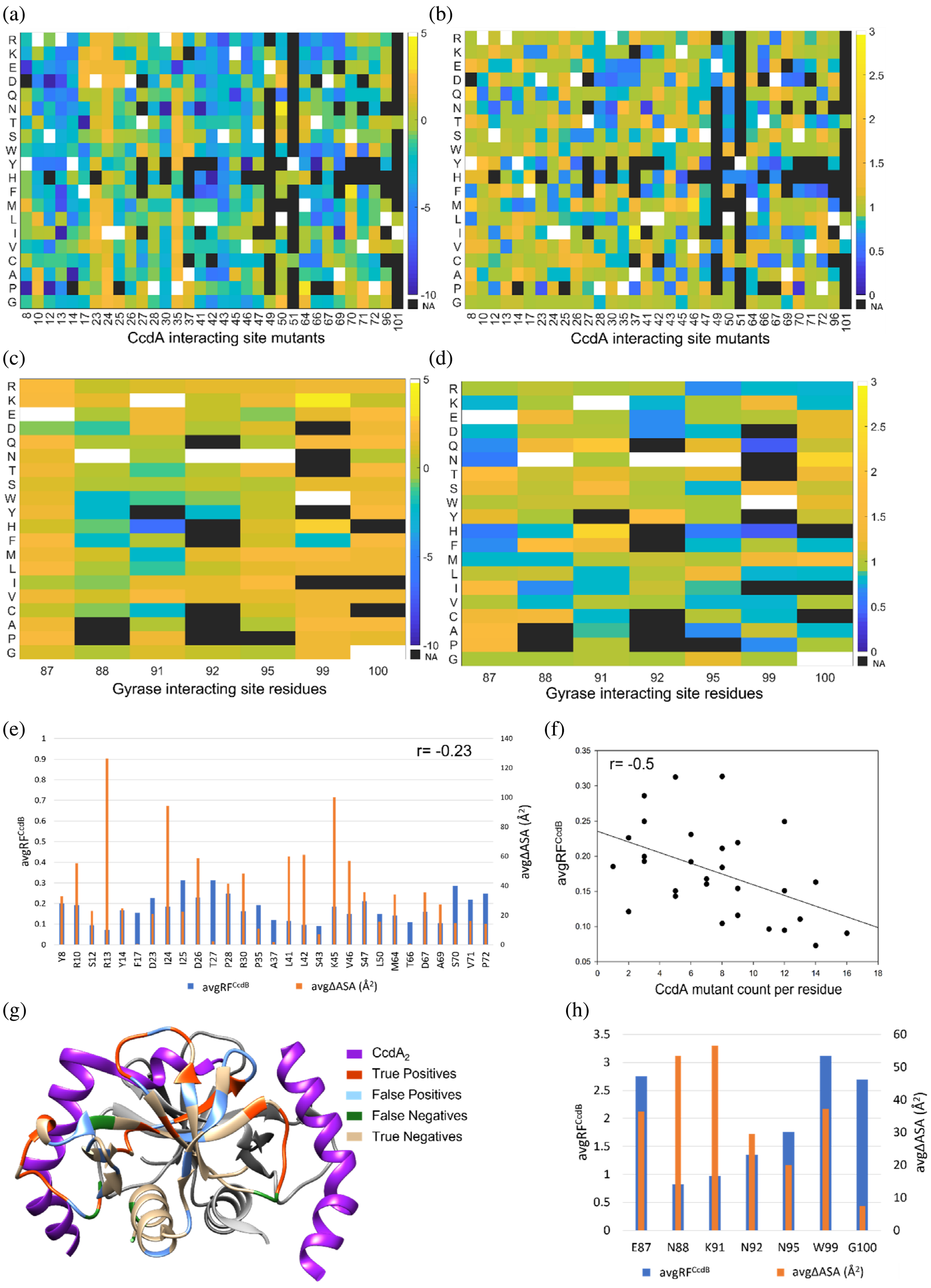

FIGURE 4 Legend on next page.

24, 25, 26, 96, and 101 are part of both the CcdA and Gyrase binding sites. Among these positions, mutants at position 25 and 26 are found in this class whereas most of the mutants at residues 24 and 96 display an inactive phenotype (Figure 4a). Most mutations at residue 101 are missing in the current dataset (Figure 4a). The data suggest that I24 and L96 make more critical contacts with Gyrase rather than CcdA. Mutants at residues 24, 87, 88, 91, 92, 95, 96, 99, and 100 of CcdB, which predominantly interact with residues at the Gyrase binding interface do not show much impairment in CcdA binding (Figure 4a–d). This suggests that residues important in CcdA and Gyrase binding are largely independent from each other.

We averaged the accessible surface area of the CcdB surface buried upon complex formation across the two chains of CcdA and also averaged the RF^{CcdB} scores over mutants for each CcdA interacting residue belonging to the "hyperactive and derepressing" class. Although a weak negative correlation was obtained between the two parameters for each CcdA interacting residue ($r = -.23$), we observed that residues that bury more than $30Å^2$ surface area upon CcdA binding have an average RF^{CcdB} score less than 0.2, that is, ~8-fold higher toxicity than the WT (Figure 4e). We also calculated the number of mutants belonging to this class for a given CcdA interacting residue and defined it as mutant count enrichment for a CcdA interacting residue. Average RF^{CcdB} score and the mutant count enrichment for CcdA interacting residues is negatively correlated (~−0.5), suggesting the residues that have a large number of mutants falling in this class and simultaneously display low average RF^{CcdB} score are the most important residues for CcdA binding (Figure 4f). These analyses identify S12, R13, F17, R30, L41, L42, S43, V46, and T66 as critical residues for CcdA binding (Table S2). From the crystal structure (PDB ID 3G7Z), it is apparent that all the residues in the 41–43 stretch interact with E54 of CcdA that is, part of the 52–55 helical stretch of CcdA,[17] implying E54 in CcdA is likely an important residue for binding to CcdB. CcdA interacting residues get highlighted when mutants belonging to the "hyperactive and derepressing" class with an additional constraint of ≥3 mutants/residue, are mapped onto the 3G7Z crystal structure[17] (Figure 4g).

The relative fitness scores based on CcdB and RelE toxicity also help to delineate the most important residues for Gyrase binding. Based on the RF^{CcdB} and $avgRF^{CcdB}$ scores, most of the mutants at these residues are inactive while a few are partially active (Figure 4c,h). As estimated from the RF^{RelE} scores, most Gyrase binding site mutations repress RelE expression except for mutations at residue 99 which is buried in free CcdB (Figure 4d). Gyrase binding induces conformational changes in CcdB.[18,27] E87, and G100 were identified as the most important residues for Gyrase binding because most mutants at these positions repress RelE expression while showing an inactive phenotype (Figure 4c,d,h). Surprisingly, N88 and K91 that show significant contact with Gyrase ($\Delta ASA > 50Å^2$) have $avgRF^{CcdB}$ scores close to 1, implying that these interactions are not critical for Gyrase binding (Figure 4h).

2.5 | Substitution preferences at functional residues versus buried residues

To further analyze substitution preferences, we grouped amino acids into the following categories aliphatic (A, C, I, L, M, V), aromatic (F, W, Y, H), polar (N, Q, S, T), and charged (D, E, K, R) with G and P into separate categories.[2] Mutations at most residues for the CcdA binding site are found in the "hyperactive and derepressing" class (Figure S7a). For the Gyrase binding site, mutations to M, S, and T are enriched in the "inactive and repressing" class (Figure S7b). We observed mutations to charged (R, K, E, D), and glycine residues are primarily enriched for buried site residues belonging to the "inactive and derepressing" category (Figure S7c). Mutations to hydrophobic residues at buried positions are expectedly not enriched in this class. Therefore, by combining

FIGURE 4 RF^{CcdB} and RF^{RelE} heatmaps and structural correlates. (a–d) Heat maps of Relative $Fitness^{CcdB}$ and Relative $Fitness^{RelE}$ scores. Blue to yellow color gradation represents increasing RF^{CcdB} and RF^{RelE} values. RF^{CcdB} scores are shown in log scale (Log_2RF^{CcdB}) whereas RF^{RelE} scores are shown in linear scale. No data are shown by black color, that is, NA (Not Applicable). WT residue at each position is indicated in white. (a) Heat map of RF^{CcdB} scores for CcdA interacting site mutants. (b) Heat map of RF^{RelE} scores for CcdA interacting site mutants. (c) Heat map of RF^{CcdB} scores for Gyrase binding site mutants. (d) Heat map of RF^{RelE} scores for Gyrase binding site mutants. (e) Comparing the $avgRF^{CcdB}$ for each CcdA interacting residue which has $RF^{CcdB} < 0.5$ and $RF^{RelE} < 1$ with the accessible surface area buried upon complex formation with CcdA (PDB ID 3G7Z). (f) Correlation between $avgRF^{CcdB}$ scores and mutant count enrichment at the CcdA interacting site residues with $RF^{CcdB} < 0.5$ and $RF^{RelE} < 1$. (g) Residues with $RF^{CcdB} < 0.5$, $RF^{RelE} < 1$ and ≥ 3 such mutants/residue are mapped onto the crystal structure of CcdB complexed with CcdA peptide (PDB ID 3G7Z).[17] One monomer of CcdB is shown in light gray while the residues identified as true positives (orange), false positives (blue), false negatives (green), and true negatives (tan) from the mutational phenotypes are mapped on the other monomer. Both chains of CcdA are shown in purple. The majority of the False Positives lie proximal to CcdA chains (<8 Å from the CcdA chains). (h) Comparing $avgRF^{CcdB}$ values with ΔASA for residues that bind only to DNA Gyrase (PDB ID 1X75).

phenotypic scores and careful examination of the mutational pattern, discrimination of Gyrase binding residues from buried residues is possible from mutational data alone.

2.6 | Evaluating performance

To avoid bias present at the codon level, we averaged RF^{CcdB} and RF^{RelE} scores over synonymous mutations and then averaged over all mutants found for a given position. Simply examining the distribution of positions belonging to different structural categories into the four mutational categories based on these two averaged RF^{CcdB} and RF^{RelE} scores per residue was not sufficient in discriminating between CcdA interacting, only Gyrase binding and buried residues. We therefore used the absolute values of RF^{CcdB} and RF^{RelE} and impose the additional constraint that a minimum number of mutants (3 or 5) per residue should have the appropriate RF^{CcdB} and RF^{RelE} values. The above criteria (summarized in Table 1) provided relatively accurate assignment of residues to the different structural categories. Results of statistical tests applied to evaluate the performance of this method for predicting CcdA interacting site, Gyrase binding site and buried site residues are mentioned in Table 1. As we enhance the stringency by increasing the number of mutants per residue in the same category (Table 1), the sensitivity decreases whereas the specificity increases.

2.7 | Characterization of selected mutants validate their inferred molecular mechanism

A subset of CcdB mutants were individually purified and identities were confirmed by measuring their mass by ESI mass spectrometry (Figure S8). Equal concentrations of all the CcdB proteins were used to measure the thermal stability of CcdB variants in the absence and presence of CcdA using nanoDSF (Figure 5a,b). The thermal shift in T_m upon CcdA binding observed for WT is 12°C. Thermal shifts observed for D26A, G57M, and E87P are 15, 14.5, and 14°C, respectively that is, slightly higher than the WT. These mutants also show decreased RelE reporter expression in vivo ($RF^{RelE} > 1$) (Figure 5c). In contrast, R13G, Q21S, and P28A show a thermal shift of 2, 8, and 10°C, respectively (Figure 5a,b), that is, less than the WT and also conferred increased RelE expression relative to WT ($RF^{RelE} < 1$) (Figure 5c). These analyses confirm that mutants that display weaker binding to CcdA relative to WT derepress RelE expression, whereas mutants that display stronger binding to CcdA than WT repress RelE expression relative to WT. Out of the three CcdA interacting-site mutants characterized, the largest decrease in binding to CcdA was observed for R13G. Reduced binding of the CcdB variant with CcdA was further confirmed by Microscale Thermophoresis experiments in which binding of the CcdB variant R13G was monitored with respect to CcdB WT protein. Since CcdB has two interaction sites with CcdA, one low affinity site in the micromolar range[17] and the other high affinity site in the picomolar range,[28] the CcdB protein was titrated in both the concentration ranges with a fixed concentration of fluorescently labeled CcdA peptide. R13G mutant did not bind to CcdA whereas WT protein bound to lower affinity (μM) binding site of CcdA (Figure S9). Binding affinity of CcdB with CcdA could not be determined for the high affinity site due to extremely tight binding of CcdB with CcdA. Gyrase binding studies were carried out for these mutants using SPR. P28A and Q21S mutants that showed a hyperactive phenotype in the deep sequencing data had higher k_{on} rate than the WT protein. R13G showed an anomalous behavior by weakly binding to GyrA14 (GyraseA14) even though it is not involved in direct contact with GyrA14 (Figure 5d).

G57M, an exposed CcdA non-interacting mutant was inferred to be inactive from the deep sequencing data.

TABLE 1 Prediction of active-site and buried positions solely from mutational data, based on RF^{CcdB} and RF^{RelE} scores

Structural category	Parameter values	Sensitivity[c] (%)	Specificity[d] (%)	Accuracy[e] (%)	MCC[f]
CcdA interacting positions[a]	$RF^{CcdB} < 0.5$, $RF^{RelE} < 1$	79	68	72	0.44
Buried positions[a]	$RF^{CcdB} > 2$, $RF^{RelE} < 1$	63	95	88	0.62
Gyrase binding site positions[a]	$RF^{CcdB} > 2$, $RF^{RelE} > 1$	86	53	56	0.20
CcdA interacting positions[b]	$RF^{CcdB} < 0.5$, $RF^{RelE} < 1$	61	82	75	0.44
Buried positions[b]	$RF^{CcdB} > 2$, $RF^{RelE} < 1$	74	74	74	0.40
Gyrase binding site positions[b]	$RF^{CcdB} > 2$, $RF^{RelE} > 1$	43	75	73	0.11

[a] In these cases, an additional criterion of mutants/residue ≥3 was used.

[b] In these cases, an additional criterion of mutants/residue ≥5 was used.

[c-f] Refer to Equations S1[c], S2[d], S3[e], and S4[f] in Supporting Information for the definition of these statistical parameters.

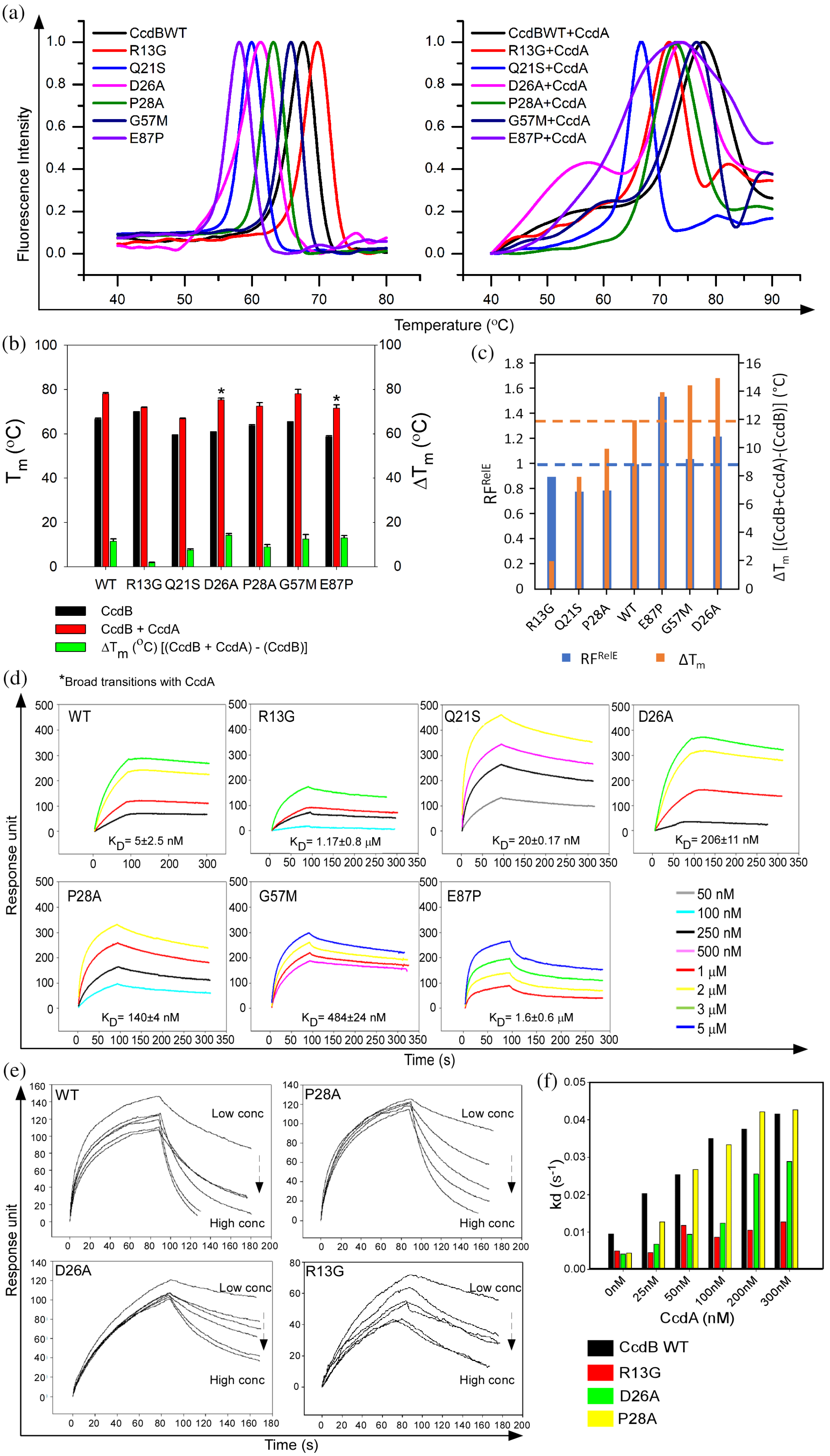

FIGURE 5 Legend on next page.

This was validated by SPR studies (Figure 5d) and several other mutants at the same position also display inactive phenotypes (Figure S6a). The misfolded fraction also increased relative to the WT (Figure 2b). Phi and psi values of 57th residue for chain A is 53.5 and −128, and for chain B is 78.6 and −106.5, respectively, indicating they lie in the disallowed regions which is only tolerated by Glycine. Hence non-Gly mutations likely cause misfolding of the protein. The 87th position was inferred to be important for Gyrase binding from deep sequencing results. SPR studies show E87P binds to GyrA14 very weakly, further validating the deep sequencing data (Figure 5d). The characterized data for these mutants is summarized in Table S3.

2.8 | CcdA binding residues are also crucial in rejuvenating CcdB from CcdB–Gyrase complex

GyrA14 was immobilized on a CM5 chip and SPR studies were carried out by passing a fixed concentration of CcdB followed by increasing concentrations of CcdA at 25°C. CcdA peptide, residues 45–72 was used for this experiment as it rejuvenates CcdB from the CcdB–Gyrase complex in a fashion similar to full length CcdA protein.[18] When this CcdA peptide was passed over the CcdB–GyrA14 complex, dissociation of CcdB was observed, and the rate of dissociation increased with increasing CcdA concentration (Figure 5e). SPR studies showed that the dissociation rate of CcdB from GyrA14 in presence of $CcdA_{45\text{-}72}$ peptide is approximately fourfold lower for the R13G variant even at the highest concentration, implying that R13 is a key residue involved in forming a complex with CcdA during or after rejuvenation of CcdB from GyrA. At 25 nM CcdA, the dissociation rate of D26A and P28A CcdB mutants is three- and twofold lower than WT CcdB, respectively (Figure 5f). This indicates that WT CcdA dissociates WT CcdB from its complex with GyrA14 more efficiently than CcdB with mutations at CcdA binding site residues. Residue 26 interacts with both Gyrase and CcdA. Along with identification of CcdA interacting residues within the CcdB toxin, this strategy can also be used to identify residues important in rejuvenation of CcdB from the CcdB-Gyrase complex.

3 | DISCUSSION

We have previously employed saturation mutagenesis coupled to deep mutational scanning to infer mutational phenotypes of ∼1,600 CcdB mutants, when heterogeneously expressed under the pBAD promoter.[1,2,26] In the present study, a comprehensive site-saturation mutagenesis library of the bacterial toxin CcdB, part of a TypeII CcdA–CcdB TA system was generated in its native operon to study mutational effects on organismal fitness. Furthermore, we developed a RelE reporter system that made use of the ability of the CcdB mutants to regulate RelE expression in order to infer the molecular mechanism behind the observed phenotypes. The two phenotypic readouts, one based on CcdB toxicity and the other based on RelE toxicity enabled discrimination between binding site residues to two different ligands, CcdA and DNA Gyrase, as well as between binding site and buried residues.

Mutations can affect activity by either altering the specific activity or total activity, through altering the fraction of the natively folded protein in vivo, or by a combination of both.[29] Determining which of the two is the main contributor is a nontrivial task. Computational approaches have been deployed in which sequence-based predicted accessibility scores are combined with mutational sensitivity data to help discriminate between interface residues and buried residues.[14] However, the predictions were only moderately accurate and were limited to predicting contacts for only one interacting partner. In the present study, solely on the basis of the

FIGURE 5 Biophysical characterization of selected CcdB mutants. (a) Thermal stability of CcdB mutants measured in the absence (left panel) and presence of CcdA (right panel) using nanoDSF. Fluorescence intensity has been normalised between 0 and 1. (b) Bar graph (below) represents the T_m data. ΔT_m is the difference in the T_m of the CcdB mutant relative to its complex with CcdA. Error bar signifies standard error between two biological duplicates. (c) Validation of the RelE reporter assay. Mutants with $RF^{RelE} > WT$ have $\Delta T_m > WT$ while mutants with $RF^{RelE} < WT$ have $\Delta T_m < WT$. (d) Gyrase binding studies of these mutants. For comparative analysis between the mutants, the scale on the y-axis is kept constant. The same concentration for all the mutants is shown by the same color. The K_D values are mentioned below each plot. SPR studies were performed in biological duplicates and the standard error for the two duplicates is mentioned. (e,f) $CcdA_{45\text{-}72}$-mediated rejuvenation of WT and mutant CcdB from their complex with GyrA14. (e) Representative binding sensorgrams of WT and CcdA interacting site mutants. Overlays show the dissociation of WT CcdB and mutants bound to GyrA14 with $CcdA_{45\text{-}72}$ peptide, concentration increasing from the top to bottom (0 nM, 25 nM, 50 nM, 100 nM, 200 nM, 300 nM). (f) k_d (s^{-1}) obtained for each mutant as a function of $CcdA_{45\text{-}72}$ concentrations. The apparent dissociation rate constants (k_d) mediated by $CcdA_{45\text{-}72}$ are approximately fourfold lower for the R13G–CcdB–GyrA14 complex than for WT–CcdB–GyrA14.

mutational sensitivity displayed by the CcdB mutants and without having to measure the expression levels of individual mutants, we were able to overcome these limitations and predict residues involved in binding to both its cognate partners and discriminate them from buried and exposed noninteracting residues. Mutations of CcdA interacting site residues majorly showed $RF^{CcdB} < 0.5$ and $RF^{RelE} < 1$ as they were most likely defective in binding to CcdA. Mutational data were in concordance with structural information, and indicated residues 8–14, 41–46, and 62–74 as most important for CcdA binding.[17] CcdB residues in the loop region 8–14 do not contact Gyrase and are important in CcdA binding.[17,18,30] Buried site residues were enriched in the class of mutants with $RF^{CcdB} > 2$ and $RF^{RelE} < 1$ as these mutations often result in misfolding of the protein,[2,31,32] thereby hampering their binding to Gyrase as well as CcdA. Most Gyrase binding site residues either showed a neutral phenotype or were enriched in the class of mutants with $RF^{CcdB} > 2$ and $RF^{RelE} > 1$ as these mutants are well folded[2] and thus form a complex with CcdA more efficiently than the buried mutants. Exposed non-active-site residues are largely insensitive to mutations ($0.5 < RF^{CcdB} < 2$), a general inference also drawn by other studies.[1,2,14] The ones that are mutationally sensitive are either active-site proximal or have RF^{CcdB} values close to the cut-offs. Previous analyses from the laboratory have successfully analyzed data from the CcdB mutant library to identify GyrA binding site residues.[1,2] However, identification of CcdA binding residues was not attempted, as a growth-based phenotypic screen had not been developed. In general, utility of mutational scanning datasets increases when combined with appropriate screens to infer molecular mechanisms responsible for the observed phenotypes.

These inferences were further validated via both in vivo and in vitro studies of the WT and mutant proteins. These studies confirmed that as expected, a defect in binding to CcdA, resulted in an increase in levels of CcdB in the cell which was sufficient to kill cells more efficiently than the WT. In the context of heterologous expression, overexpression can compensate for binding defects.[1,2] However, in the present experiments where all mutants are expressed under control of the same native promoter, this is not possible. The current approaches that help identify important CcdB residues involved in CcdA binding, will also help identify residues important in rejuvenation of CcdB from CcdB-Gyrase complex. Since, CcdA is an intrinsically disordered protein, identification of these residues can help in understanding the role of intrinsic disorder and allostery in rejuvenation.

We demonstrate that in an operonic context we are able to use straightforward genetic screens to pull out information on binding site residues to two different binding partners, one of which is an intrinsically disordered protein. We are also able to distinguish between mutational effects on stability and binding, the former occur primarily at buried residues and the latter at exposed residues. This discrimination has previously been difficult to accomplish in the absence of additional structural information. While most substitutions that resulted in increased toxicity (low RF^{CcdB}) occurred at residues in contact with CcdA, the actual amount of buried surface upon complex formation is not a good predictor of mutational sensitivity. More data on other systems is required to ascertain if there are consistently predictable phenotypic effects depending on the nature of substitution at contact residues. Preliminary biophysical characterization of residues involved in interaction with CcdA also suggests that mutations at these residues can affect rejuvenation. There are very few prior studies that have carried out deep mutational scans in an operonic context with a native promoter. Similar approaches can be applied both to other TA systems,[33–35] as well as more generally to other multiprotein systems and should therefore be of general interest. These are especially useful when there are limited sequence homologs available, and where recently developed deep learning approaches[36] to elucidate protein complex structures do not work well.

4 | MATERIALS AND METHODS

4.1 | Generation of a CcdB site saturation mutagenesis library in its native operon

A previously generated CcdB SSM library[26] was cloned in pUC57 vector via Gibson Assembly according to the manufacturer's protocol.[37] The Gibson product was transformed into high efficiency (10^9 CFU/μg of pUC57 plasmid DNA) electrocompetent *E. coli* Top10Gyr cells.[38] The cells were plated on LB agar plates containing 100 μg/mL ampicillin, for selection of transformants and incubated for 12 hr at 37°C. Pooled plasmid was purified using a Qiagen plasmid maxiprep kit as per the manufacturer's instructions.

4.2 | Preparation and isolation of barcoded PCR products for multiplexed deep sequencing

The master library was purified from Top10Gyr (resistant strain). The library was then transformed and subjected to selection in both Top10 (sensitive strain) and Top10-Gyr harboring the RelE reporter gene (RelE reporter strain). Pooled, purified plasmid samples from each condition were PCR amplified with primers containing a six-base long Multiplex Identifier (MID) tag. 370-bp long PCR products containing the full *ccdB* gene, were pooled, gel-band purified, and sequenced using Illumina Sequencing, on the Hi-seq 2500 platform at Macrogen, Korea.

4.3 | Normalization

Read numbers for all mutants at all 101 positions (1-101) in CcdB were analyzed. Mutants having less than 100 reads in the resistant strain were not considered for analysis.

$$F(x_i) = \frac{x_i}{\sum x_i + x_{WT}} \tag{1}$$

$$\mathrm{F}(y_i) = \frac{y_i}{\sum y_i + y_{WT}} \tag{2}$$

$$F(z_i) = \frac{z_i}{\sum z_i + z_{WT}} \tag{3}$$

Here, a given mutant is represented by "*i*" whereas WT is represented by "WT." Number of reads in Top10Gyr resistant strain, Top10 sensitive strain, and RelE reporter strain is represented by "*x*," "*y*," and "*z*," respectively. *F* (x_i), $F(y_i)$, and $F(z_i)$ are the fraction representation of a mutant in resistant, sensitive, and RelE reporter strain, respectively.

$$\text{Deepseq ratio}_{\text{sen}_i} = \frac{F(y_i)}{F(x_i)} \tag{4}$$

$$\text{Deepseq ratio}_{\text{RelE}_i} = \frac{F(z_i)}{F(x_i)} \tag{5}$$

$$\text{RF}_i^{\text{CcdB}} = \frac{\text{Deepseq ratio}_{\text{sen}_i}}{\text{Deepseq ratio}_{\text{sen}_{\text{WT}}}} \tag{6}$$

$$\text{RF}_i^{\text{RelE}} = \frac{\text{Deepseq ratio}_{\text{RelE}_i}}{\text{Deepseq ratio}_{\text{RelE}_{\text{WT}}}} \tag{7}$$

For simplicity, these mutational scores are represented as RF^{CcdB} and RF^{RelE} throughout the text. RF^{CcdB} is based on the CcdB toxicity readout in the Top10 strain while RF^{RelE} is based on the RelE toxicity readout in Top10Gyr strain harboring the RelE reporter gene. Variant scores of RF^{RelE} were rounded up to 1 decimal place for data analysis. An average of the mutational scores of the two biological replicates for each variant is taken. The two variant scores are generally indicated in linear scale throughout the text.

4.4 | In vivo activity, expression, and purification of CcdB mutant proteins

Growth assay was carried out for selected CcdB mutants in Top10Gyr versus Top10 *E. coli* strains. Expression and solubility of a subset of CcdB mutants heterologously expressed from pBAD24 vector in Top10Gyr strain was estimated as described previously.[2] These mutants were purified via CcdA affinity chromatography.[39] Protein mass was confirmed using ESI mass spectrometry. All CcdB concentrations reported here are in monomeric units. Purified proteins were further used for nanoDSF, MST, and SPR experiments. Detailed description of procedures is mentioned in Supporting Information.

AUTHOR CONTRIBUTIONS

Priyanka Bajaj: Conceptualization (supporting); data curation (lead); formal analysis (lead); investigation (lead); methodology (equal); software (equal); validation (lead); visualization (lead); writing – original draft (lead); writing – review and editing (equal). **Kavyashree Manjunath:** Software (equal). **Raghavan Varadarajan:** Conceptualization (lead); formal analysis (supporting); funding acquisition (lead); methodology (equal); resources (lead); supervision (lead); writing – review and editing (equal).

ACKNOWLEDGMENTS

This work was funded by grants to RV from the Department of Science and Technology, grant number-EMR/2017/004054, DT. December 15, 2018) and Department of Biotechnology, grant no. BT/COE/34/SP15219/2015 DT. November 20, 2015, Government of India. The authors also acknowledge funding for infrastructural support from the following programs of the Government of India: DST FIST, UGC Centre for Advanced study, Ministry of Human Resource Development (MHRD), and the DBT IISc Partnership Program. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication. PB acknowledges University Grants Commission, Government of India, for her fellowship. RV is a J. C. Bose Fellow of DST. Nonavinakere Seetharam Srilatha is acknowledged for assisting with the SPR experiments. We thank all members of the RV Lab for their valuable suggestions.

CONFLICT OF INTEREST

The authors claim no conflict of interest.

DATA AVAILABILITY STATEMENT

The raw deep sequencing data used in the present study has been deposited in NCBI's Sequence Read Archive (accession no. SRR17982061).

ORCID

Priyanka Bajaj https://orcid.org/0000-0001-8474-6149

REFERENCES

1. Adkar BV, Tripathi A, Sahoo A, et al. Protein model discrimination using mutational sensitivity derived from deep sequencing. Structure. 2012;20:371–381.
2. Tripathi A, Gupta K, Khare S, et al. Molecular determinants of mutant phenotypes, inferred from saturation mutagenesis data. Mol Biol Evol. 2016;33:2960–2975.
3. Berezin C, Glaser F, Rosenberg J, et al. ConSeq: The identification of functionally and structurally important residues in protein sequences. Bioinformatics. 2004;20:1322–1324.
4. Gherardini PF, Helmer-Citterich M. Structure-based function prediction: Approaches and applications. Briefings Funct Genomics Proteomics. 2008;7:291–302.
5. Verma R, Mitchell-Koch K. In silico studies of small molecule interactions with enzymes reveal aspects of catalytic function. Catalysts. 2017;7:212. https://www.mdpi.com/2073-4344/7/7/212
6. Steele DL, El-Kabbani O, Dunten P, et al. Expression, characterization and structure determination of an active site mutant (Glu202-Gln) of mini-stromelysin-1. Protein Eng. 2000;13:397–405.
7. Lai J, Niks D, Wang Y, et al. X-ray and NMR crystallography in an enzyme active site: The indoline quinonoid intermediate in tryptophan synthase. J Am Chem Soc. 2011;133:4–7.
8. Kaur G, Dutta D, Thakur KG. Crystal structure of mycobacterium tuberculosis CarD, an essential RNA polymerase binding protein, reveals a quasidomain-swapped dimeric structural architecture. Proteins Struct Funct Bioinforma. 2014;82:879–884.
9. Schureck MA, Maehigashi T, Miles SJ, et al. Structure of the Proteus vulgaris HigB-(HigA) 2-HigB toxin-antitoxin complex. J Biol Chem. 2014;289:1060–1070.
10. Bajaj P, Arya PC. Evolution and spread of SARS-CoV-2 likely to be affected by climate. Clim Change Ecol. 2021;1:1–10.
11. Fowler DM, Araya CL, Fleishman SJ, et al. High-resolution mapping of protein sequence-function relationships. Nat Methods. 2010;7:741–746.
12. Lefèvre F, Rémy MH, Masson JM. Alanine-stretch scanning mutagenesis: A simple and efficient method to probe protein structure and function. Nucleic Acids Res. 1997;25:446–448.
13. 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;25:395–406.
14. Bhasin M, Varadarajan R. Prediction of function determining and buried residues through analysis of saturation mutagenesis datasets. Front Mol Biosci. 2021;281:37942–37951.
15. Tam JE, Kline BC. The F plasmid ccd autorepressor is a complex of CcdA and CcdB proteins. Mol Gen Genet MGG. 1989;219:26–32.
16. Magnuson R, Yarmolinsky MB. Corepression of the P1 addiction operon by Phd and doc. J Bacteriol. 1998;180:6342–6351.
17. De Jonge N, Garcia-Pino A, Buts L, et al. Rejuvenation of CcdB-poisoned gyrase by an intrinsically disordered protein domain. Mol Cell. 2009;35:154–163.
18. Aghera NK, Prabha J, Tandon H, et al. Mechanism of CcdA-mediated rejuvenation of DNA gyrase. Structure. 2020;28:562–572.e4.
19. Ryu YS, Chandran SP, Kim K, Lee SK. Oligo- and dsDNA-mediated genome editing using a tetA dual selection system in Escherichia coli. PLoS One. 2017;12:e0181501.
20. Li XT, Thomason LC, Sawitzke JA, Costantino N, Court DL. Positive and negative selection using the tetA-sacB cassette: Recombineering and P1 transduction in *Escherichia coli*. Nucleic Acids Res. 2013;41:e204.
21. Chen W, Li Y, Wu G, et al. Simple and efficient genome recombineering using kil counter-selection in *Escherichia coli*. J Biotechnol. 2019;294:58–66.
22. Khetrapal V, Mehershahi K, Rafee S, Chen S, Lim CL, Chen SL. A set of powerful negative selection systems for unmodified Enterobacteriaceae. Nucleic Acids Res. 2015;43:e83.
23. Dao-Thi MH, Van Melderen L, De Genst E, et al. Molecular basis of gyrase poisoning by the addiction toxin CcdB. J Mol Biol. 2005;348:1091–1102.
24. Chakshusmathi G, Mondal K, Lakshmi GS, et al. Design of temperature-sensitive mutants solely from amino acid sequence. Proc Natl Acad Sci. 2004;101:7925–7930.
25. Overgaard M, Borch J, Gerdes K. RelB and RelE of *Escherichia coli* form a tight complex that represses transcription via the ribbon-helix-helix motif in RelB. J Mol Biol. 2009;394:183–196.
26. Jain PC, Varadarajan R. A rapid, efficient, and economical inverse polymerase chain reaction-based method for generating a site saturation mutant library. Anal Biochem. 2014;449:90–98.
27. Smith AB, Maxwell A. A strand-passage conformation of DNA gyrase is required to allow the bacterial toxin, CcdB, to access its binding site. Nucleic Acids Res. 2006;34:4667–4676.
28. Hadži S, Mernik A, Podlipnik Č, Loris R, Lah J. The thermodynamic basis of the fuzzy interaction of an intrinsically disordered protein. Angew Chemie Int Ed. 2017;56:14494–14497.
29. Stiffler MA, Hekstra DR, Ranganathan R. Evolvability as a function of purifying selection in TEM-1 β-lactamase. Cell. 2015;160:882–892.
30. Loris R, Garcia-Pino A. Disorder- and dynamics-based regulatory mechanisms in toxin-antitoxin modules. Chem. Rev. 2014;13:6933–6947.
31. Moreno-Gonzalez I, Soto C. Misfolded protein aggregates: Mechanisms, structures and potential for disease transmission. Seminars in cell & developmental biology. 2011;22:482–487.
32. Sahoo A, Khare S, Devanarayanan S, Jain PC, Varadarajan R. Residue proximity information and protein model discrimination using saturation-suppressor mutagenesis. eLife. 2015;4:e09532.
33. Xue L, Yue J, Ke J, et al. Distinct oligomeric structures of the YoeB–YefM complex provide insights into the conditional

cooperativity of type II toxin–antitoxin system. Nucleic Acids Res. 2020;48:10527–10541.

34. Kim D-H, Kang S-M, Park SJ, Jin C, Yoon H-J, Lee B-J. Functional insights into the Streptococcus pneumoniae HicBA toxin-antitoxin system based on a structural study. Nucleic Acids Res. 2018;46:6371–6386.
35. Guillet V, Bordes P, Bon C, et al. Structural insights into chaperone addiction of toxin-antitoxin systems. Nat Commun. 2019; 10:1–15.
36. Torrisi M, Pollastri G, Le Q. Deep learning methods in protein structure prediction. Comput Struct Biotechnol J. 2020;18: 1301–1310.
37. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009;6:343–345.
38. Chandra S, Gupta K, Khare S, et al. Codon optimality is the primary contributor to the exceptional mutational sensitivity of CcdA antitoxin in its operonic context. bioRxiv. 2022. doi:10.1101/2022.01.15.476443
39. Chattopadhyay G, Bhowmick J, Manjunath K, Ahmed S, Goyal P, Varadarajan R. Mechanistic insights into global suppressors of protein folding defects. bioRxiv. 2021; doi:10.1101/2021.11.18.469098

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Bajaj P, Manjunath K, Varadarajan R. Structural and functional determinants inferred from deep mutational scans. Protein Science. 2022;31(7):e4357. https://doi.org/10.1002/pro.4357