

# Design of novel cyanovirin-N variants by modulation of binding dynamics through distal mutations

I Can Kazan<sup>1,2†</sup>, Prerna Sharma<sup>2†</sup>, Mohammad Intiazur Rahman<sup>2†</sup>, Andrey Bobkov<sup>3</sup>, Raimund Fromme<sup>2</sup>, Giovanna Ghirlanda<sup>2\*</sup>, S Banu Ozkan<sup>1\*</sup>

<sup>1</sup>Center for Biological Physics and Department of Physics, Arizona State University, Tempe, United States; <sup>2</sup>School of Molecular Sciences, Arizona State University, Tempe, United States; <sup>3</sup>Sanford Burnham Prebys Medical Discovery Institute, La Jolla, United States

**Abstract** We develop integrated co-evolution and dynamic coupling (ICDC) approach to identify, mutate, and assess distal sites to modulate function. We validate the approach first by analyzing the existing mutational fitness data of TEM-1  $\beta$ -lactamase and show that allosteric positions co-evolved and dynamically coupled with the active site significantly modulate function. We further apply ICDC approach to identify positions and their mutations that can modulate binding affinity in a lectin, cyanovirin-N (CV-N), that selectively binds to dimannose, and predict binding energies of its variants through Adaptive BP-Dock. Computational and experimental analyses reveal that binding enhancing mutants identified by ICDC impact the dynamics of the binding pocket, and show that rigidification of the binding residues compensates for the entropic cost of binding. This work suggests a mechanism by which distal mutations modulate function through dynamic allostery and provides a blueprint to identify candidates for mutagenesis in order to optimize protein function.

**\*For correspondence:**

Giovanna.Ghirlanda@asu.edu (GG);  
banu.ozkan@asu.edu (SBO)

<sup>†</sup>These authors contributed equally to this work

**Competing interest:** The authors declare that no competing interests exist.

**Funding:** See page 17

**Preprinted:** 22 December 2020

**Received:** 12 February 2021

**Accepted:** 28 November 2022

**Published:** 06 December 2022

**Reviewing Editor:** José D

Faraldo-Gómez, National Institutes of Health, United States

© Copyright Kazan, Sharma, Rahman et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

## Editor's evaluation

A computational approach is proposed to identify mutations in enzymes that might impact their interactions with substrates. For one enzyme, in particular, the predictions are validated through experiments, using multiple techniques. Taken together, these data lead to non-trivial conclusions in regard to the nature of allosteric effects, albeit it remains unclear whether these conclusions will apply more broadly when other enzymes are examined.

## Introduction

The evolutionary history of a protein comprises the ensemble of mutations acquired during the course of its evolutionary trajectory across different species, and contains valuable information on which residue positions contribute the most to a given protein's 3D-fold and function based on their conservation (*Campbell et al., 2016; Rivoire et al., 2016; Yang et al., 2016*). Furthermore, the subset of positions that are co-evolved (i.e., correlated mutational sites) provide clues on specific, native-state interactions. Pairwise residue contacts inferred from co-evolved positions within a protein family can be used as distance restraints to accurately model 3D structures (*de Juan et al., 2013; Hopf et al., 2019; Kamisetty et al., 2013; Kim et al., 2014; Tripathi et al., 2015*). Recent revolutionary successes in accurate predictions of 3D protein structures combine these methods with machine learning strategies, that is, deep learning (*Jumper et al., 2021; Wang et al., 2016; Xu, 2019*). Co-evolved positions also embed information on protein function, for example, revealing how factors such as binding affinity

and specificity are modulated across evolutionary history and species (*Rivoire et al., 2016; Salinas and Ranganathan, 2018; Torgeson et al., 2022*). However, accessing, interpreting, and applying this information in a predictive manner is very challenging; mutations observed in the evolutionary history are often distal from the functional sites, implying that protein dynamics are responsible for their effects on function and that these sites act as distal allosteric regulators of function (*Campitelli et al., 2020a; Modi et al., 2021a; Romero and Arnold, 2009; Salinas and Ranganathan, 2018; Tokuriki et al., 2012; Torgeson et al., 2022; Wei et al., 2016*).

Molecular dynamics (MD) simulations can capture protein dynamics and reveal the impact of distal mutations on function (*Bowman and Geissler, 2012; Campbell et al., 2016; Campitelli et al., 2020a; Jiménez-Osés et al., 2014; Kolbaba-Kartchner et al., 2021; Modi et al., 2021a; Yang et al., 2016*). However, the computational cost of MD simulations of sufficient length can be prohibitively high; further, it's often far from straightforward to forge a clear connection to function. To bridge this gap, we developed a framework to quickly evaluate MD trajectories and identify the sensitivity of a given position to mutation based on its intrinsic flexibility, which we assess using our dynamic flexibility index (DFI) metric, and on its dynamic coupling with functionally critical positions assessed by dynamic coupling index (DCI) (*Campitelli et al., 2018; Gerek and Ozkan, 2011; Kumar et al., 2015b; Larimore et al., 2017*). DFI measures the resilience of a position by computing the total fluctuation response and thus captures the flexibility/rigidity of a given position. Applying DFI to several systems, we showed that rigid positions such as hinge sites contribute the most to equilibrium dynamics, and that mutations at hinge sites significantly impact function regardless of the distance from active sites (*Kim et al., 2015; Kolbaba-Kartchner et al., 2021; Modi et al., 2021b, Modi et al., 2018; Modi and Ozkan, 2018; Zou et al., 2021; Zou et al., 2015*). DCI measures the dynamic coupling between residue pairs and thus identifies positions most strongly coupled to active/binding sites; these positions point to possible allosteric regulation sites important for modulating function in adaptation and evolution (*Butler et al., 2015; Modi et al., 2021a, Campitelli et al., 2021; Kuriyan and Eisenberg, 2007; Lu and Liang, 2009; Modi and Ozkan, 2018; Ose et al., 2020; Risso et al., 2018; Wodak et al., 2019*).

In this paper, we present integrated co-evolution and dynamic coupling (ICDC) approach to identify distal allosteric sites, and to assess and predict the effects of mutations on these sites on function. We propose a system to classify residue positions in a binary fashion based on co-evolution (co-evolved, 1 or not, 0) and dynamic coupling by DFI and DCI (dynamically coupled 1, or not, 0) with the functionally critical sites. This classification captures the complementarity of dynamics-based and sequence-based methods. We hypothesize that positions belonging to category (1,1), that is, positions both co-evolved and dynamically coupled with the functional sites, will have the largest effect on function.

We validate our hypothesis first by analyzing the existing mutational fitness data for TEM-1  $\beta$ -lactamase, available for every position of the protein (*Stiffler et al., 2015*). In agreement with our hypothesis, we find that mutations on category (1,1) positions significantly modulate the function. A large fraction of mutations enhancing enzymatic activity correspond to category (1,1) irrespective of distance from the active site. Second, we apply our ICDC approach to blindly predict and experimentally validate mutations that allosterically modulate dimannose binding in a natural lectin, cyanovirin-N (CV-N). CV-N binds dimannose with nanomolar affinity and remarkable specificity (*Barrientos et al., 2003; Botos and Wlodawer, 2005; Botos and Wlodawer, 2003; Mori and Boyd, 2001; O'Keefe et al., 2003*). It is part of the CV-N family, found in a wide range of organisms including cyanobacterium, ascomycetous fungi, and fern (*Koharudin et al., 2008; Koharudin and Gronenborn, 2013; Patsalo et al., 2011; Percudani et al., 2005; Qi et al., 2009*). While the 3D folds is remarkably conserved in all experimentally characterized members, the affinity and specificity for different glycans and, in particular, to dimannose varies significantly (*Koharudin et al., 2009; Koharudin et al., 2008; Matei et al., 2016; Woodrum et al., 2013*). To design CV-N variants with improved binding affinities for dimannose based on distal allosteric coupling, we binned each position in one of the four categories based on computed DFI, DCI, and co-evolution rates. We explored mutations at these sites based on frequency in the sequence alignment. After obtaining the mutant models through MD simulations, we assessed the impact of each naturally observed mutation on binding affinity by docking dimannose to the mutant models via Adaptive BP-Dock (*Bolia et al., 2014a; Bolia et al., 2014b; Bolia and Ozkan, 2016*). We chose position I34, which belongs to category (1,1) and is 16 Å away from the binding pocket, for experimental validation. We found that mutations I34K/L/Y had a diverse effect on glycan

binding, either improving by twofold or abolishing completely. Through experimental and MD studies we show that the observed improvement in binding affinity is due to changes in the dynamics of residues in the binding pocket; mutation I34Y leads to rigidification of binding sites, thus compensating the entropic cost of binding (Breiten *et al.*, 2013; Chodera and Mobley, 2013; Cornish-Bowden, 2002; Fox *et al.*, 2018). Mutations at an additional position (A71T/S) from category (1,1) showed evidence of the same allosteric mechanism governing the modulation of binding dynamics. Overall, this study provides not only a new approach to identify distal sites whose mutations modulate binding affinity, but also sheds light into mechanistic insights on how distal mutations modulate binding affinity through dynamics allostery.

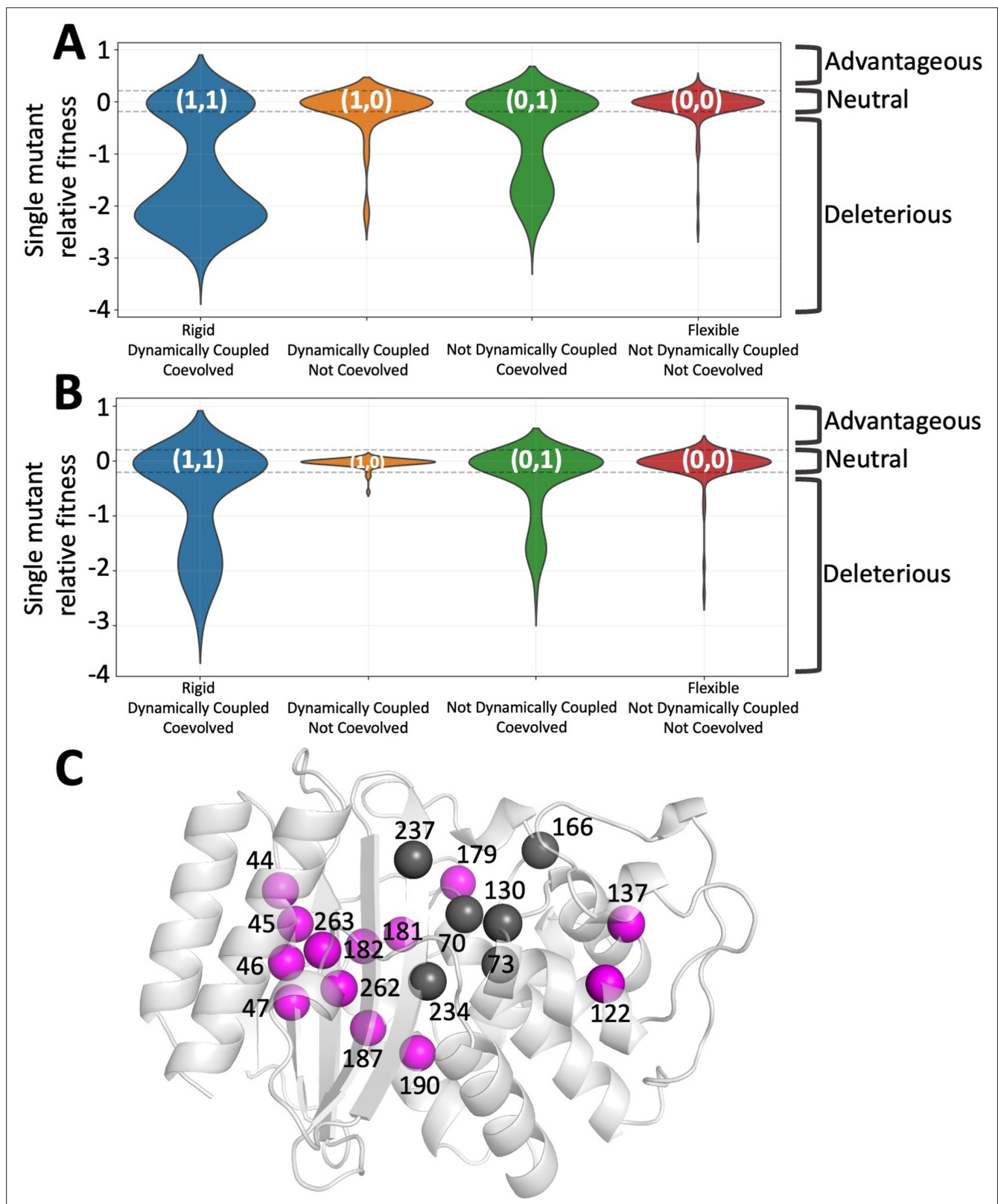
## Results and discussion

### Combining long-range dynamic coupling analysis with co-evolution allows to identify distal sites that contribute to functional activity

With our ICDC approach, we aim to explore the role of dynamics versus evolutionary coupling (EC) as well as the role of rigidity versus flexibility in allosterically modulating active/binding site dynamics. To this extent, we created four unique categories that classify residue positions based on residue DFI score, DCI score, and co-evolutionary score: category (1,1) is dynamically and co-evolutionarily coupled rigid sites (exhibiting %DFI values 0.2 or lower, showing 0.7 or higher %DCI with the binding site, and showing 0.6 or higher co-evolution scores with the binding site); category (1,0) is dynamically coupled but co-evolutionarily not coupled sites; category (0,1) is dynamically not coupled but co-evolutionarily coupled sites; category (0,0) is dynamically not coupled, and co-evolutionarily not coupled flexible sites (exhibiting %DFI values 0.7 or higher) (Supplementary file 1 and Supplementary file 2); importantly, this classification is based on two independent statistical approaches thus compensate the noise of individual approaches. Based on our evolutionary analysis (Campitelli *et al.*, 2020a; Modi *et al.*, 2021b; Modi and Ozkan, 2018), we hypothesize that category (1,1) would impact protein activity or binding affinity the most.

To test our hypothesis, we first analyzed the deep mutational scanning data available for the TEM-1  $\beta$ -lactamase, correlating changes in ampicillin degradation activity (e.g., MIC values) with mutations to all possible amino acids at each position (Stiffler *et al.*, 2015). The experimental results showed that amino acid substitutions at the catalytic site residues of TEM-1 negatively impacted activity. Mutations at other positions also affected activity; while most mutations were deleterious, surprisingly, others resulted in increased activity. The impact of mutations on dynamics and function of TEM-1 have been heavily explored but the distal mutational effects are still poorly understood (Kolbaba-Kartchner *et al.*, 2021; Modi *et al.*, 2021b; Modi and Ozkan, 2018; Salverda *et al.*, 2010; Schneider *et al.*, 2021; Stiffler *et al.*, 2015; Thomas *et al.*, 2010; Zimmerman *et al.*, 2017; Zou *et al.*, 2015). We applied our approach by obtaining DFI, DCI, and co-evolution scores for every position of TEM-1 and binning residue positions into each ICDC category (Supplementary file 1 and Supplementary file 5). We constructed fitness distributions for each category using the experimentally measured single mutant relative fitness values for all mutations per position provided in the dataset (Figure 1).

We found that category (1,1) positions show the highest impact, both significantly enhancing and reducing ampicillin degradation by TEM-1 (Figure 1A&C). In addition, category (0,0) residue mutations (i.e., the exact opposite of category (1,1)) lie within the neutral-like activity range defined by Stiffler *et al.*, 2015, suggesting that mutations on positions that neither co-evolve nor dynamically couple to active site do not affect the function significantly. Category (1,0) residues enhance activity more than those in the neutral category (0,0). Mutations in category (0,1) positions also modulate function in both positive and negative direction, albeit not as strongly as those in category (1,1). However, mutations that negatively impact activity are conspicuously under-represented in the multiple sequence alignment (MSA) of native sequences (Figure 1B), particularly in category (1,1). This finding implies nature mostly allows mutations that don't compromise fold and function: Negative selection (i.e., elimination of amino acid types that are detrimental to the folding) is a major force in shaping the mutational landscape (Jana *et al.*, 2014; Modi *et al.*, 2021a; Morcos, 2020; Morcos *et al.*, 2014; Morcos *et al.*, 2013). Thus, the use of conservation information from MSA is a useful tool in eliminating deleterious amino acid substitutions in protein design.



**Figure 1.** Integrated co-evolution and dynamic coupling (ICDC) categories based on the dynamics and co-evolutionary analyses applied on TEM-1  $\beta$ -lactamase. (A) The distributions in the form of violin plots are obtained for each ICDC category using all available experimental mutational data (Stiffler et al., 2015). (B) Violin plots showing the fitness values for amino acid substitutions observed in the natural sequences. (C) The category (1,1) positions are mapped on 3D structure. The catalytic site residues are shown in dark gray whereas category (1,1) positions are shown in magenta color. The function altering category (1,1) positions are widely distributed over the 3D structure.

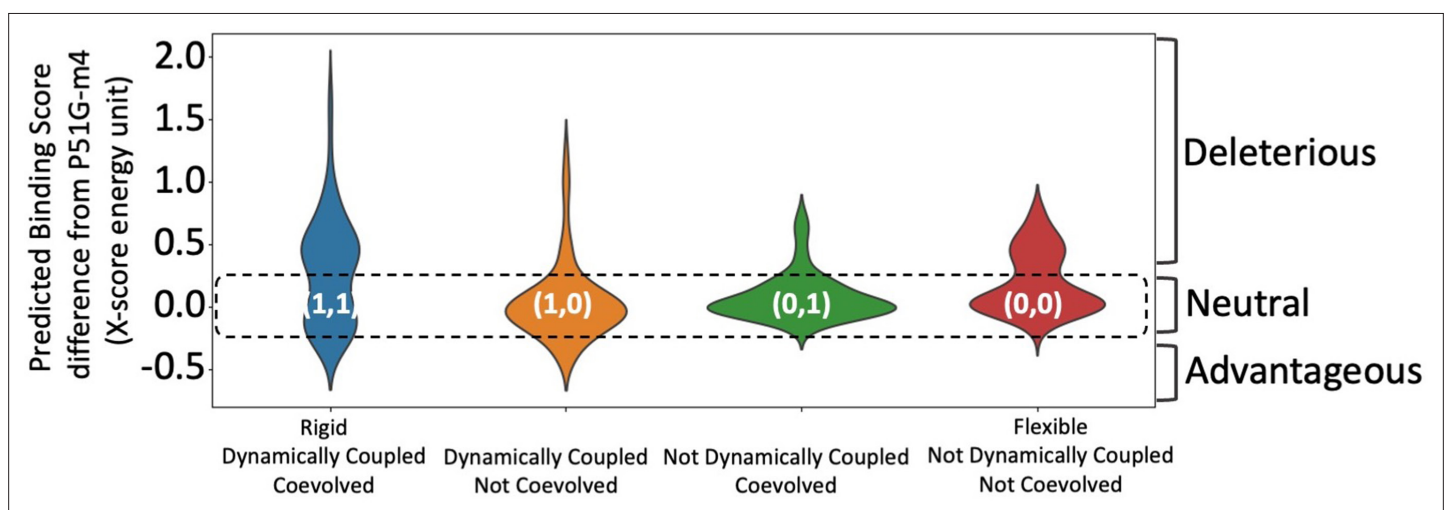
Our ICDC selection criteria effectively identifies residue positions and their amino acid substitutions that could fine-tune function without leading to a functional loss; and category (1,1) residues have the largest impact on function irrespective of their distance from active site (Figure 1C).

## Application of ICDC approach to modulate CV-N binding affinity through distal mutations

CV-N is a small (11 kDa) natural lectin isolated from cyanobacterium *Nostoc elliposporum* which comprises two quasi-symmetric domains, A (residues 1–38/90–101) and B (residues 39–89 respectively), that are connected to each other by a short helical linker. Despite almost having identical structures, the domains show relatively low sequence homology (28% sequence identity and 52% similarity). Functionally, they both bind dimannose, yet the affinity is quite different, with domain B having tighter binding affinity ( $K_d = 15.3 \mu\text{M}$ ), and domain A showing weak affinity ( $K_d = 400 \mu\text{M}$ ) (Balzarini, 2007; Bolmstedt et al., 2001; Li et al., 2015).

To simplify our analyses, we used a designed CV-N variant, P51G-m4, that contains a single high-affinity dimannose binding site (domain B), folds exclusively as a monomer in physiological conditions, and is more stable to thermal denaturation than wild type (Fromme et al., 2008; Fromme et al., 2007). The binding pocket of domain B of CV-N has been subjected to intense scrutiny to glean information on the origin of its binding specificity for dimannose (Bewley, 2001; Bolia et al., 2014b; Botos and Wlodawer, 2003; Li et al., 2015; Vorontsov and Miyashita, 2009). Previous mutational studies on the binding pocket residues have shown their importance in modulating interaction with dimannose (Barrientos et al., 2006; Bolia et al., 2014b; Chang and Bewley, 2002; Matei et al., 2008). All known substitutions of the binding residues led to decreased binding affinity for dimannose on domain B (Bolia et al., 2014b; Fujimoto and Green, 2012; Kelley et al., 2002; Matei et al., 2011; Ramadugu et al., 2014). Evolutionary analyses shows that the majority of the binding site residues are conserved in CV-N glycan interactions, suggesting that affinity is already optimized at the binding site (Koharudin et al., 2008; Percudani et al., 2005). We hypothesized that amino acid substitutions at distal positions could enhance the dimannose affinity of CV-N by rigidification of the binding site and applied our ICDC approach to CV-N to identify positions in each category (Supplementary file 2).

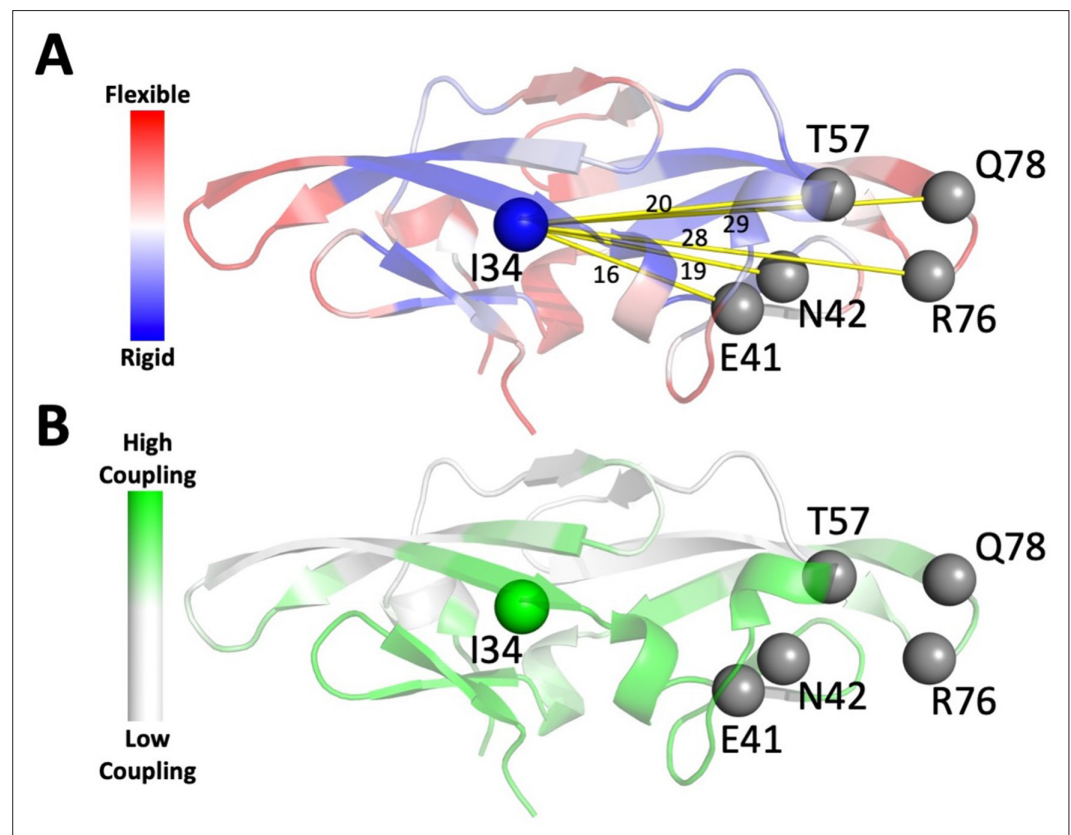
We generated models of CV-N variants in each ICDC category by mutating these positions to amino acid types observed in the MSA of CV-N family members, choosing the subset of sequences that have binding sites with identical or similar amino acid composition to P51G-m4 CV-N. As discussed above, this approach allows us to identify amino acid substitutions with the least impact on fold. All the substitutions identified (104 variants in total) were modeled using the crystal structure of P51G-m4 CV-N (Fromme et al., 2008) and subjected to MD simulations (Abraham et al., 2015; Van Der Spoel et al., 2005). The best conformation sampled for each variant obtained from



**Figure 2.** Predicted binding energies for each integrated co-evolution and dynamic coupling (ICDC) category. Mutations in category (1,1) positions comprise the highest number of binding energy enhancing mutations as well as deleterious mutations. Mutations in category (0,0) positions are mostly near neutral (category (1,1) and (0,0) p value <0.3).

equilibrated production trajectories was used as a model for dimannose docking analysis. We evaluated the variants using Adaptive BP-Dock (Bolia and Ozkan, 2016), a computational docking tool that incorporates both ligand and receptor flexibility to accurately sample binding-induced conformations, and ranks them using X-scores binding energy units (XEU) (Figure 4—figure supplement 1). In previous work on CV-N this method yielded good correlations with experimentally measured binding affinities ( $K_d$ ), and established  $-6.0$  XEU as a good threshold to differentiate variants that bind dimannose from ‘non-binders’ (Bolia et al., 2014b; Li et al., 2015; Woodrum et al., 2013). Here, we applied Adaptive BP-Dock initially on wild-type CV-N and its variants, P51G-m4 and mutDB (a mutant in which binding by domain B has been obliterated) and the results recapitulate the success of previous studies (Supplementary file 3). This result shows that Adaptive BP-Dock can correctly assess the dimannose binding of CV-N and its variants, thus, we applied it on new P51G-m4 CV-N variants to predict the impact of mutations on dimannose binding. Figure 2 shows the distribution of changes in predicted binding energy scores relative to the P51G-m4 energy scores for mutations belonging to each binary category: a positive change in binding score represents an unfavorable effect on binding, and, conversely, a negative change in the score indicates an enhancement in binding.

The substitutions on positions in category (1,1) (Figure 2) yield a wide range of change in binding energy scores: the tail of the distribution on the positive side reaches nearly a binding score change of 2.0 XEU and on the negative side values below  $-0.5$  XEU. Strikingly, the positions in category (1,1) yield the most binding enhancing energy scores compared to all other categories, mirroring TEM-1 results. Additionally, the substitutions applied in category (1,0) also result in more favorable binding energy scores for dimannose. Mutations in both category (1,1) and (1,0) present favorable binding energy scores. However, the number of mutations predicted to be enhancing binding in category (1,1) is more than those in category (1,0) (26% of category (1,1) compared to 14% of category (1,0)). Interestingly, the mutations in category (1,0) that disrupt the binding energy scores is not as strong



**Figure 3.** DFI and DCI analyses on CV-N. (A) Dynamic flexibility index (DFI) profile mapped onto cyanovirin-N (CV-N) structure: red corresponds to high DFI (very flexible sites), and blue to low DFI values (rigid sites). Position I34 (low DFI score) is highlighted. (B) Dynamic coupling index (DCI) profile projected on CV-N structure with green corresponding to sites exhibiting high coupling with binding site residues.

**Table 1.** Predicted binding affinities of domain B, experimental ITC data, and chemical denaturation experiments for P51G-m4 and its I34 variants.

| Protein      | Predicted binding score (X-score energy unit) | ITC dimannose $K_d$ ( $\mu$ M) | ITC dimannose $\Delta H$ (kcal/mol) | ITC dimannose $T\Delta S$ (kcal/mol) (T=298 K) | ITC dimannose $\Delta G$ (kcal/mol) | $\Delta G_{H_2O}$ (kcal/mol) | $C_m$ (M)        |
|--------------|---|--------------------------------|-------------------------------------|--|-------------------------------------|------------------------------|------------------|
| P51G-m4      | -6.62   | 117 $\pm$ 3                    | -12.3 $\pm$ 0.3                     | -7.00 $\pm$ 0.3                                | -5.30 $\pm$ 0.3                     | 3.01 $\pm$ 0.047             | 1.46 $\pm$ 0.019 |
| P51G-m4-I34K | -5.85   | No binding                     | No binding                          | No binding                                     | No binding                          | 2.40 $\pm$ 0.124             | 0.68 $\pm$ 0.015 |
| P51G-m4-I34L | -6.19   | 148 $\pm$ 2                    | -9.60 $\pm$ 0.1                     | -4.40 $\pm$ 0.1                                | -5.20 $\pm$ 0.1                     | 2.95 $\pm$ 0.077             | 1.39 $\pm$ 0.009 |
| P51G-m4-I34Y | -6.75   | 64 $\pm$ 5                     | -4.35 $\pm$ 0.1                     | 1.32 $\pm$ 0.2                                 | -5.67 $\pm$ 0.2                     | 2.91 $\pm$ 0.157             | 1.13 $\pm$ 0.017 |

as category (1,1), but similar to category (0,1) and (0,0). The observed mostly neutral behavior with category (0,0) agrees with the same trend obtained with TEM-1 analyses.

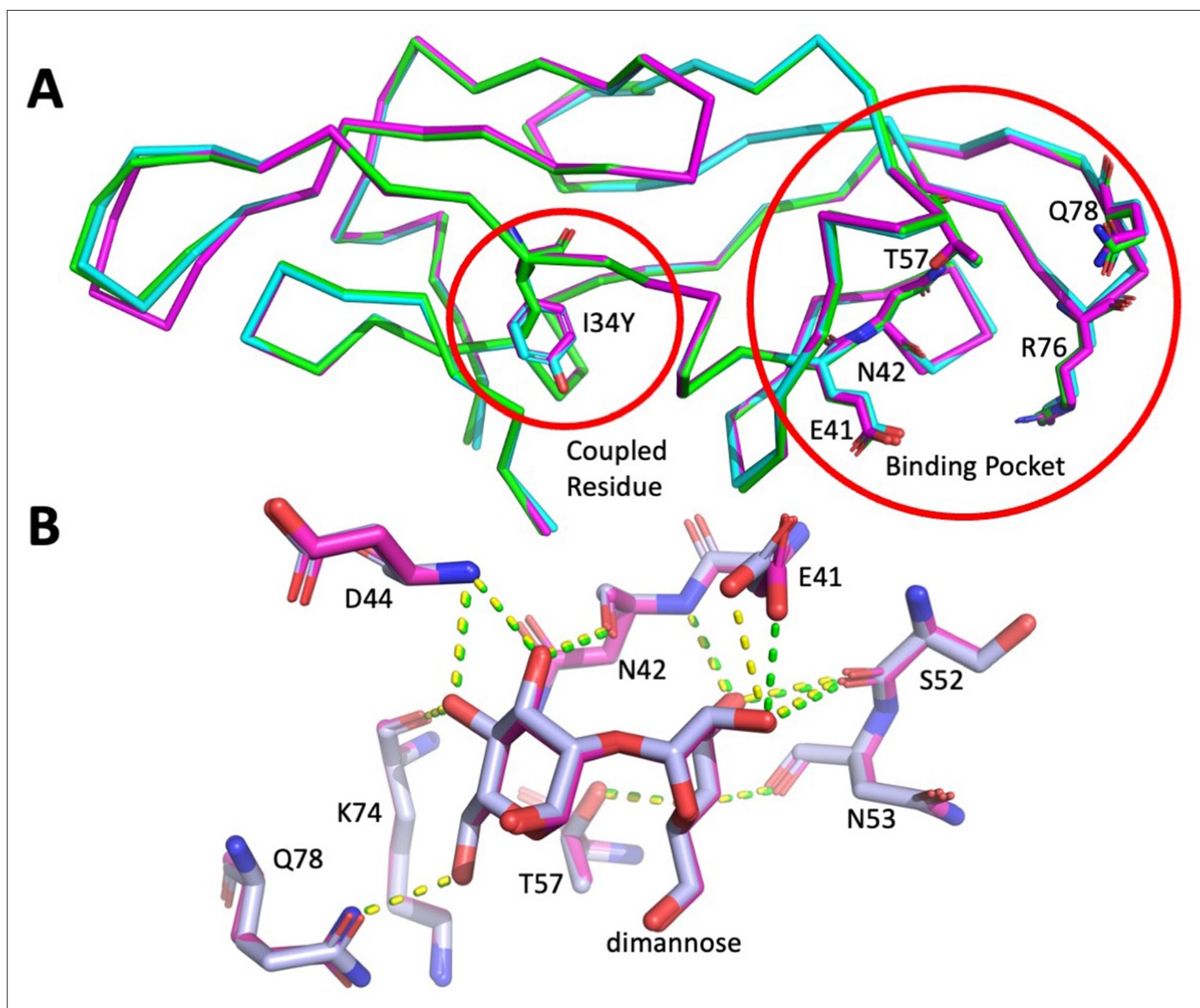
Overall, the distribution of computational binding scores of dimannose binding to CV-N in each category aligns with the distribution of experimentally characterized TEM-1 fitness results of the same category. However, there are some discrepancies, for example, there are beneficial mutations in category (0,1) in TEM-1, but we don't observe the same trend in CV-N. This is due to the initial challenge faced in constructing the MSA of CV-N homologous proteins. There is limited sequence information, and most of the proteins in the CV-N family exhibits binding specificity to a different glycan (Fujimoto and Green, 2012; Koharudin et al., 2009). In contrast,  $\beta$ -lactamase family proteins exhibit highest activity toward penicillin, and they have been subjected to strong natural selection leading to conservation in both fold and function (Salverda et al., 2010; Zou et al., 2021). Hence, the less noise in evolutionary analysis in case of  $\beta$ -lactamase family of proteins allows us to correctly filter deleterious type of substitutions based on the MSA. Regardless, however, in both cases, as hypothesized, substitutions on category (1,1) residues impact the function most.

To further investigate the mechanism of functional modulation of category (1,1) mutations, we chose the position with highest binding enhancing docking scores, I34, from category (1,1). I34 exhibits %DFI values lower than 0.2 (Figure 3A), is at least 16 Å away from binding residues (distal), dynamically coupled (Figure 3B) and co-evolved with the binding pocket (Supplementary file 2 and Supplementary file 6). Moreover, docking scores of I34 variants suggest that the mutations can modulate binding in a wide range: I34Y variant leads to an increase in binding affinity (beneficial), I34K decreases the binding affinity (deleterious), and I34L yields no change (neutral) (Table 1).

To verify the predictions of I34 variants, we first assessed the folding and thermal stability of these mutants by circular dichroism (CD) spectroscopy. Far-UV CD spectroscopy showed that all mutants are well folded and adopt a fold similar to the parent protein, characterized by spectra with a single negative band centered at 216 nm. We determined the stability of the mutants by CD monitored thermal denaturation; the thermal denaturation curves were analyzed to obtain apparent melting temperature ( $T_m$ ) values. We found that the conservative mutation I34L is as stable as P51G-m4, with apparent  $T_m$  of 57.8°C and 58°C, respectively. In contrast, I34Y and I34K were less thermostable than P51G-m4 as shown by apparent  $T_m$  values of 54.7°C and 47°C, respectively. Not surprisingly, substituting a hydrophobic residue with a basic aliphatic amino acid (lysine) has a large destabilizing effect, while aromatic and polar tyrosine is better tolerated. The trend of thermostability is P51G-m4-I34 L > I34 Y > I34 K (Figure 4—figure supplement 2).

Chemical denaturation experiments were used to extract thermodynamic values, after ensuring complete equilibration at each concentration of guanidinium hydrochloride by incubating the samples for 72 hr (Patsalo et al., 2011). The  $\Delta G_{H_2O}$  values and  $C_m$  values of P51G-m4, I34L, I34Y, and I34K are found as 3.0, 2.94, 2.91, and 2.38 kcal/mol and of 1.45, 1.39, 1.13, and 0.68 M respectively (Table 1). The results align with the thermal denaturation results: P51G-m4 is the most stable to denaturant, followed by I34L, I34Y, and I34K (Figure 4—figure supplement 3).

Next, we evaluated the impact of the mutations on the dimannose binding affinity by isothermal titration calorimetry (ITC) (Figure 4—figure supplement 4); data were analyzed to extract  $K_d$  values listed in Table 1. We found that I34Y binds dimannose with tightest affinity ( $K_d$ : 64  $\mu$ M) of all the mutants tested, a twofold improvement over P51G-m4 ( $K_d$ : 117  $\mu$ M). Binding by I34L is slightly weaker



**Figure 4.** The comparison of the crystal structures of P51G-m4 and I34Y. **(A)** The crystal structures of I34Y (bound in magenta and unbound in cyan) and its template protein P51G-m4 (green) are superimposed. **(B)** Overlay of bound structures of I34Y (magenta) and P51G-m4 (gray) (RMSD 0.15 Å); dashed lines depict polar interactions with dimannose.

The online version of this article includes the following figure supplement(s) for figure 4:

**Figure supplement 1.** The flowchart of the Adaptive BP-Dock Scheme.

**Figure supplement 2.** Fits for thermal melts of the cyanovirin-N (CV-N) **(A)** I34 variants and **(B)** A71 variants.

**Figure supplement 3.** Fits for the chemical denaturation experiments of the variants.

**Figure supplement 4.** Binding isotherms of cyanovirin-N (CV-N) mutants upon titration with dimannose: **(A)** I34Y, **(B)** P51G-m4, and **(C)** A71T.

**Figure supplement 5.** Comparison of experimentally solved I34Y structure with docked pose from Adaptive BP-Dock algorithm.

with a  $K_d$  of 148  $\mu\text{M}$ . No binding was observed for I34K in these conditions. Thermodynamic values extracted from ITC experiments (*Table 1*), suggesting that entropy changes play an important role in the observed changes in binding affinity: surprisingly, entropy is positive for I34Y, indicating an increase in disorder upon binding.

To glean more information on the mode of binding by I34Y, we determined the X-ray structure of the unbound and dimannose-bound form and compared it with the template protein P51G-m4. The



fold is highly conserved (**Figure 4**) as shown by main chain RMSD of 0.16 and 0.20 Å with bound and unbound I34Y, respectively, and tyrosine is well tolerated at position I34. The binding pocket region is also structurally conserved compared to P51G-m4. Analysis of the polar contacts between dimannose and P51G-m4 and I34Y (**Figure 4B**) shows an identical number of hydrogen bonds (11) with the ligand, indicating a conserved binding pose. We compared the docked pose of I34Y acquired from Adaptive BP-Dock with the bound X-ray structure. The ligand shows an RMSD value of 0.75 Å (**Figure 4—figure supplement 5**). These observations suggest that the increase in binding affinity of I34Y toward dimannose might be mediated by equilibrium dynamics, which are not captured by the crystal structure. This hypothesis is supported by the changes in entropy compensation measured experimentally (ITC) in dimannose binding by P51G-m4 (negative  $\Delta S$ ) and I34Y (positive  $\Delta S$ ).

## Molecular mechanism governing the binding dynamics in I34 variants

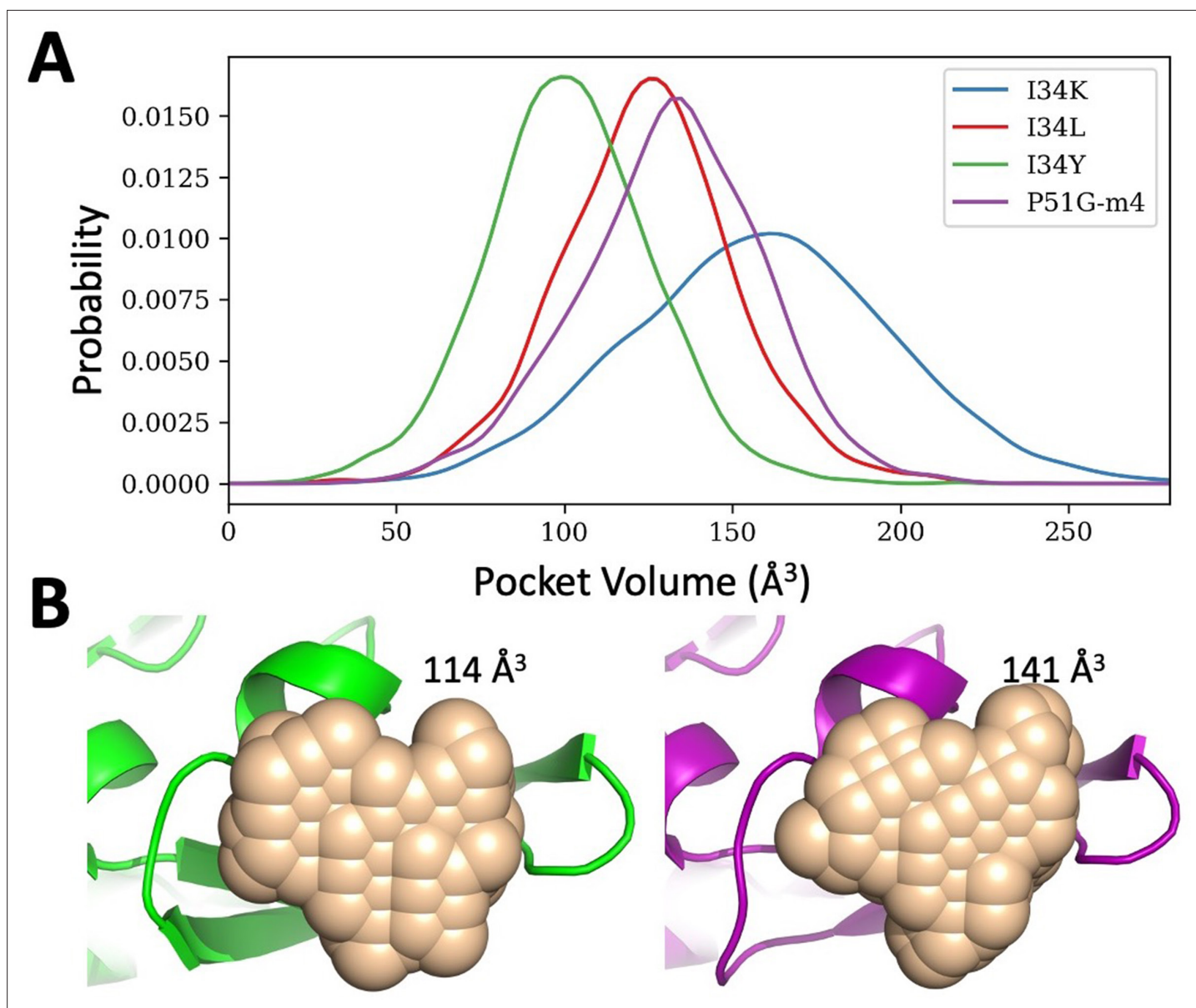
It is interesting to observe that a distal site can modulate binding affinity to a wide range based on amino acid substitutions. This finding has also been observed for allosterically regulated enzymes such as LacI, for which different amino acid substitutions on non-conserved sites lead to gradual changes in function, acting like a rheostatic switch to modulate function through conformational dynamics (**Campitelli et al., 2021; Campitelli et al., 2020b; Meinhardt et al., 2013; Miller et al., 2017; Swint-Kruse et al., 1998**). To gather atomic level detail on how the substitutions on I34 dynamically modulate the binding affinity, we employed MD simulations in both bound and unbound forms (see Methods for details of the simulations). The unbound trajectories were analyzed for acquiring binding pocket hydrogen bond distances and pocket volume. Later, to learn about the ligand-induced conformational dynamic changes, the bound trajectories were utilized to estimate computational binding free energies (**Deng and Roux, 2009; Okazaki et al., 2006**).

Previous computational work in our lab had linked binding affinity in the CV-N family to the accessibility of the binding pocket: A hydrogen bond between the amide hydrogen of N42 and carbonyl oxygen of N53 forms a closed pocket, hindering glycan accessibility, whereas the loss of this hydrogen bond leads to an open pocket (**Li et al., 2015**). Using the formation of this hydrogen bond in the trajectories of unbound WT and I34Y as metric for assessing open and closed conformations, we found that I34Y variant samples the open binding pocket more often than P51G-m4 (**Figure 5—figure supplement 1**).

Another compelling evidence differentiating I34 variants from P51G-m4 is the change in their binding pocket volumes estimated by POVME pocket volume calculation tool (**Wagner et al., 2017**). The calculated pocket volumes for I34Y, I34K, and P51G-m4 were converted into frequencies to obtain probability distributions (**Figure 5A**), revealing that I34Y variant samples a more compact pocket volume compared to P51G-m4. If the pocket is too small or too large, dimannose cannot maximize its interaction with the protein, and a compact conformation enables dimannose to easily make the necessary hydrogen bond interactions with the protein. This optimum pocket volume sampled by I34Y may also explain the different binding energetics observed by ITC, in which a positive entropy change upon binding compensates for the loss in enthalpy compared to P51G-m4 (**Table 1; Breiten et al., 2013; Cornish-Bowden, 2002**). Pocket volume analysis reveals a larger value for I34K compared to P51G-m4, suggesting that this mutant cannot accommodate the necessary interactions with the dimannose resulting in loss of binding. We applied the same pocket volume calculation to the X-ray structures of P51G-m4 and I34Y variant, and we found volumes of 141 and 114 Å<sup>3</sup> for P51G-m4 and I34Y, respectively, in the unbound forms (**Figure 5B**). These volumes correlate well with the mean volumes from MD trajectories, suggesting that the variants modulate the conformational dynamics of binding pocket.

Overall, the conformational dynamics analysis of the unbound conformations indicates a shift of the native ensemble toward a smaller pocket volume upon I34Y mutation. This could explain the decrease in the entropic cost of binding observed in ITC results. We also analyzed the binding energetics by carrying out dimannose docking with 2000 different conformations sampled from the binding pocket volume distributions. We found that the small volume restrict accessibility to the side-chain conformations of binding residue R76 in the I34Y variant, yielding different hydrogen bond patterns with the dimannose (**Figure 5—figure supplement 2**) and suggesting a loss in enthalpic contribution.

The bound simulation trajectories were subjected to the MM-PBSA approach to estimate computational binding free energies and related enthalpic and entropic contributions (**He et al., 2020**;



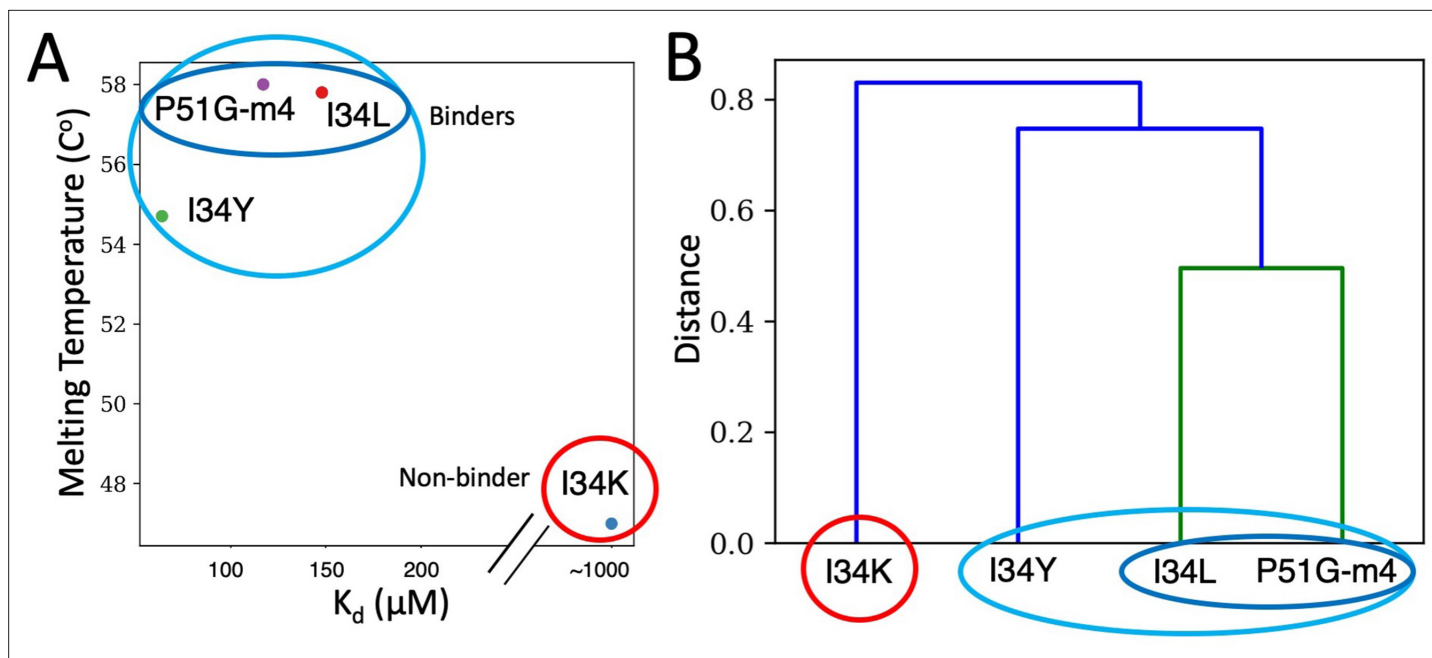
**Figure 5.** Binding pocket volume estimations for P51G-m4 and its variants. **(A)** Probability distribution of the pocket volume analyses obtained from molecular dynamics (MD) simulation trajectories. I34Y populates a conformation with an optimum volume more than others. P51G-m4 and I34L variant samples similar pocket volumes, but I34K variant has a larger pocket volume compared to others. **(B)** Pocket volume comparison of the domain B of solved structures for P51G-m4 (purple) and I34Y variant (green).

The online version of this article includes the following figure supplement(s) for figure 5:

**Figure supplement 1.** The difference in accessibility of the binding pocket for P51G-m4 and I34Y.

**Figure supplement 2.** We sampled 2000 different conformations from molecular dynamics (MD) simulations for P51G-m4 cyanovirin-N (CV-N) and I34Y mutant and performed dimannose docking to obtain docked poses and then analyzed hydrogen bond patterns.

*Rastelli et al., 2010*). The results are tabulated on **Supplementary file 4**. The computed binding free energies capture the trend of experimental binding affinities ( $R=0.87$ ). The I34Y variant displays a more favorable binding with dimannose compared to wild type. Interestingly, both experimental and computational results show I34Y compensating the enthalpic loss with entropic gain. While I34L variant enthalpic loss is greater than I34Y in computational approach, the overall binding free energy mirrors the ITC results. Additionally, loss of binding of I34K variant overlaps with the ITC data.



**Figure 6.** Clustering of CV-N variants using DFI profiles and biophysical properties.

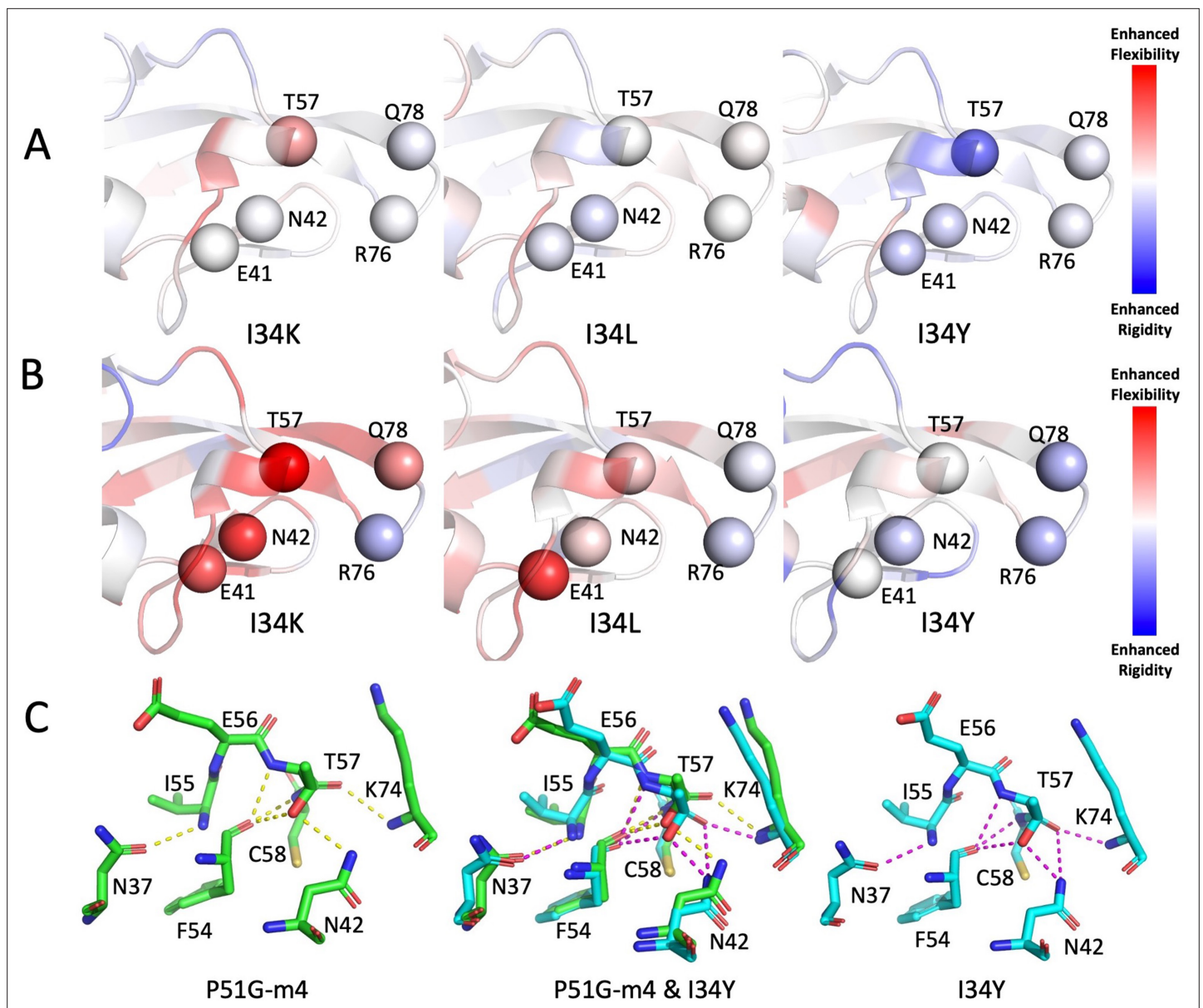
(A) 2D map of  $K_d$  and melting temperature of P51G-m4 and its variants. (B) Principal component analyses (PCA) clustering on the first two principal components of the dynamic flexibility index (DFI) profiles as a dendrogram.

### Substitutions of I34 modulates the conformational ensemble leading to change in dimannose binding affinity

Proteins adapt to a new environment by modulating the native state ensemble through mutations of different positions while keeping the 3D structure conserved (Campitelli et al., 2020a; Kuriyan and Eisenberg, 2007; Li et al., 2015; Liu and Nussinov, 2017; Modi and Ozkan, 2018; Risso et al., 2018; Tripathi et al., 2015; Woodrum et al., 2013). As we also observed a similar pattern of conservation of structure yet change in function in our designed CV-N I34 variants, we further analyzed the flexibility profiles of I34 variants. The DFI profiles clustered using principal component analyses match the 2D map of melting temperature and  $K_d$  as reaction coordinates, suggesting a correlation between changes in dynamics and changes in function (Figure 6). The 2D map shows I34L, P51G-m4, and I34Y under the same cluster, with I34L and P51G-m4 close, while I34K is markedly different (Figure 6A). The dendrogram constructed based on the DFI profiles captures this clustering (Figure 6B) with P51G-m4 and I34L variant under the same branch, suggesting their dynamics are very similar; I34Y is under the same main cluster albeit in a different branch. I34K is under a separate branch, indicating different dynamics. This is in agreement with our previous studies, where substitutions on DARC spots modulate binding dynamics reflected in their flexibility profiles to adapt to a new environment (Campitelli et al., 2021; Kumar et al., 2015b; Modi et al., 2021a).

We further gleaned a molecular view of the role of flexibility in binding by comparing changes in DFI profiles of the binding site residues with P51G-m4 for each mutant, in the unbound and bound form (Figure 7A and B). We found that flexibility at position T57 is highly dependent on the amino acid at position I34: flexibility increases in I34K, suggesting a higher entropic penalty for binding interactions. It is unchanged in I34L, which has similar binding affinity. In contrast, T57 becomes much more rigid in I34Y mutant. This indicates that the rigidification leading to a decrease in the entropic cost can contribute to the binding affinity enhancement of this mutant, which is also in agreement with the ITC results.

Comparison of the flexibility profiles of the bound form with those of the unbound form reveals that residue I34 in WT drastically gets rigidified upon binding, whereas I34Y variant does not. The decreased flexibility of T57 in the unbound form of I34Y accommodates the interactions with dimannose, contributing to the entropic compensation. In addition to the binding site residues of domain B, the flexibility of the rest of the residues also contributes to the total change in binding free energies.



**Figure 7.** Changes in flexibility of the binding site residues upon mutations in bound and unbound forms. **(A)** Change in flexibility of I34K, I34L, and I34Y relative to P51G-m4 in unbound form is shown. Residues E41, N42, and T57 rigidifies on I34Y compared to P51G-m4. **(B)** Change in flexibility of I34K, I34L, and I34Y relative to P51G-m4 in bound form is projected on structure. **(C)** Hydrogen bonding interactions of residues I55, E56, T57, and C58 are shown for P51G-m4 and I34Y variant.

The online version of this article includes the following figure supplement(s) for figure 7:

**Figure supplement 1.** Correlation between change in DFI profiles and change in  $\Delta G$  of binding.

**Figure supplement 2.** Network of hydrogen bond interactions connecting residue location 34 to T57 is investigated in I34Y variant and P51G-m4 cyanovirin-N (CV-N).

Therefore, we analyzed the correlation between (i) the sum of total change in flexibility of the binding site residues, (ii) the binding site residues and the residues exhibiting highly coupling with the binding pocket, with the experimentally measured binding affinity change. We observe a strong correlation between change in flexibility and change in affinity, as expected I34Y exhibiting tighter binding also gets more rigidified upon binding compared to P51G-m4. Moreover, inclusion of the highly coupled residues in addition to the domain B binding sites in computing the total sum of DFI scores yields a higher correlation with the experimental binding affinity change (**Figure 7—figure supplement 1A**).

On the other hand, the correlation between the flexibility change of the randomly selected residues and experimental binding affinities yields poor correlation coefficient (**Figure 7—figure supplement 1B**). These results strongly support the role of dynamic allostery in modulating binding affinity.

The rigidification of T57 in I34Y variant is compelling evidence that the distal mutation is allosterically controlling the binding site dynamics. We further computed the network of interactions that connects the residue position 34–57 and investigated whether distinct pathways emerge after I34Y mutation. We analyzed the hydrogen bond networks, particularly computed the possible network of hydrogen bonds creating pathways from 34 to 57 using the sampled snapshots from the MD trajectories (**Figure 7—figure supplement 2**). This analysis presents a unique pathway from 34 to 57 by first forming a new hydrogen bond between the side chain oxygen of the Tyrosine 34 and the nitrogen of the Tyrosine 100 in I34Y variant. Furthermore, a second pathway is also found which is sampled much more frequently in I34Y variant strengthening the communication between positions 34 and 57. Thus, both pathways may contribute to the rigidification of T57. We also analyzed the conformations from MD clustered with highest percentage based on alpha carbon RMSD for I34Y and P51G-m4 and compared the hydrogen bond interactions of T57 and its neighboring residues. The closest neighbors of T57; positions I55, E56, and C58 conserved their hydrogen bond interactions with their surrounding residues between P51G-m4 and I34Y. On the other hand, T57 makes an additional hydrogen bond interaction in I34Y compared to P51G-m4 (**Figure 7C**), suggesting that enhancement in hydrogen bond networking of T57 in I34Y leads to rigidification of this position in equilibrium dynamics.

To gain more insight on distal dynamic modulation of binding pocket particularly the decrease of binding site flexibility through distal coupling, we computationally and experimentally characterized another residue, A71, belonging to category (1,1) and its mutations: T, S. The docking scores and DFI profiles of A71 variants show high similarity to position I34 ones. The variant A71T is predicted as binding enhancing by our docking scheme displaying a similar binding score as I34Y (A71T predicted binding score:  $-6.81$  XEU), whereas variant A71S is predicted as analogous to I34L variant (A71S predicted binding score:  $-6.20$  XEU). This position is next to residue E72, which is within the hydrogen bond pathway (Pathway 2) (**Figure 7—figure supplement 2**) identified previously connecting I34 and binding residue T57. Furthermore, the computed binding free energies by MM-PBSA is found to be correlating with position I34 results. The A71T variant shows a binding free energy near I34Y (A71T  $\Delta G$ :  $-13.70$  kcal/mol with  $\Delta H$ :  $-29.33$  kcal/mol and  $T\Delta S$ :  $-15.63$  kcal/mol), and A71S close to I34L (A71S  $\Delta G$ :  $-9.98$  kcal/mol with  $\Delta H$ :  $-29.00$  kcal/mol and  $T\Delta S$ :  $-19.02$  kcal/mol). All computational analyses suggested that A71 can modulate binding affinity through distal dynamic coupling similar to I34, hence we experimentally characterized these two variants.

The experimental binding affinity by ITC correlates with in silico predictions. When the change in total DFI score upon binding is compared to change in free energy of binding from ITC experiments (**Figure 7—figure supplement 1**), A71T ( $\Delta G$ :  $-5.70$  kcal/mol with  $\Delta H$ :  $-6.00$  kcal/mol and  $T\Delta S$ :  $-0.30$ ) features both a change in total DFI and  $\Delta G$  closer to I34Y, and A71S (A71S  $\Delta G$ :  $-5.10$  kcal/mol with  $\Delta H$ :  $-9.10$  kcal/mol and  $T\Delta S$ :  $-4.00$ ) shows a score identical to I34L. The entropy of A71T shows a similar change as I34Y experimentally (A71T  $T\Delta S$ :  $-0.30$ ) indicating that the same compensation mechanism is utilized by another category (1,1) residue. A71S is closer to I34L (A71S  $T\Delta S$ :  $-4.00$ ). Similar to I34Y, the melting temperature of A71T is lower than P51G-m4 (**Figure 4—figure supplement 2**). Results of A71 variants further establish the potential of ICDC and category (1,1) residues in diversely tuning the binding affinity of domain B of CV-N through playing enthalpy-entropy compensation of binding process.

Our new ICDC approach suggests that it is possible to identify and incorporate distal mutations into protein design bringing together evolutionary inferences with long-range dynamic communications within the 3D network of interactions.

## Methods

### Adaptive BP-Dock

Adaptive backbone perturbation docking, Adaptive BP-Dock in short, allows us to model the interaction between CV-N and glycans in silico (**Bolia and Ozkan, 2016**). Adaptive BP-Dock combines the complex simulation of backbone flexibility of a protein into Rosetta's ligand docking application (**Davis and Baker, 2009**). The common restriction in docking is the implementation of flexibility

of receptor and ligand (Davis et al., 2009; Davis and Baker, 2009; DeLuca et al., 2015; Meiler and Baker, 2006). Rosetta included the flexibility of ligand in their Monte Carlo sampling approach but lacking full receptor flexibility. This high-order challenge is overcome by utilizing perturbation response scanning (PRS) to compute backbone changes during docking (Atilgan and Atilgan, 2009; Bolia et al., 2014b; Ikeguchi et al., 2005). This procedure also allows the modeling of transition from an unbound state to a bound state (Bolia and Ozkan, 2016). The computational cost of sampling is reduced by using a coarse-grained approach employing elastic network model (ENM) leading to an efficient way of computing backbone perturbations, mimicking the ligand interacting with receptor (Atilgan et al., 2001; Atilgan et al., 2010; Atilgan and Atilgan, 2009).

We employed Adaptive BP-Dock in modeling glycan CV-N interactions starting from an unbound conformation of CV-N. The perturbed pose of the protein is calculated using PRS. The structure is then minimized, and the side chains are added at this step. The glycan is docked to the minimized structure using RosettaLigand algorithm. Rosetta samples bound conformations using a knowledge-based potential function and calculates bound pose energies. The lowest energy docked pose is selected and feed back to perturbation step, and the same procedure is followed iteratively until a convergence is reached. At the end of each iteration the lowest energy docked pose is taken and binding score is calculated using an empirical scoring function X-score. XEUs have shown to provide higher correlations with experimental results (Wang et al., 2002). The flow of the algorithm is shown in **Figure 4—figure supplement 1**. Adaptive BP-Docks iterative algorithm ensures that the sampling does not get trapped in a local minimum and reaches a global minimum. The challenge of unbound/bound modeling is solved using the iterative approach as the conformations are led toward a bound pose with the help of PRS.

## Molecular dynamics

Gromacs simulations are conducted for P51G-m4 CV-N and all the variants in unbound form, and further for P51G-m4 CV-N, 134 variants I34K, I34L, I34Y, and A71 variants A71S, A7T in bound form (Abraham et al., 2015; Van Der Spoel et al., 2005). For each simulation the all-atom system is parametrized with CHARMM36 force field and explicit water model TIP3P. The solvation box is set to be minimum 16 Å from the edge of the protein. The system is neutralized by potassium ions to sustain electroneutrality and minimized with steepest descent for 10,000 steps. A short-restrained equilibrium is conducted in the constant number of particles, pressure, and temperature ensemble (NPT) for 5 ns using the Berendsen method at 300 K temperature and 1 bar pressure. NPT production trajectories were performed with Nose-Hoover and Parrinello-Rahman temperature and pressure coupling methods for 2 μs at 300 K and 1 bar. For all cases periodic boundary conditions and particle-mesh Ewald (PME) with interaction cutoff of 12 Å is employed with Gromacs version 2018.1.

## Dynamic flexibility index

DFI is a position-specific metric that can measure the resilience of a given position to the force perturbations in a protein. It calculates the fluctuation response of a residue relative to the gross fluctuation response of the protein (Kumar et al., 2015b; Larrimore et al., 2017). DFI calculates residue response due to a perturbation by utilizing covariance matrices.

$$[\Delta R]_{3N \times 1} = [H]_{3N \times 3N}^{-1} [F]_{3N \times 1}$$

$$DFI_i = \frac{\sum_{j=1}^N |\Delta R^j|_i}{\sum_{i=1}^N \sum_{j=1}^N |\Delta R^j|_i}$$

Residue response,  $\Delta \mathbf{R}$ , is calculated using linear response theory by applying force,  $\mathbf{F}$ , in multiple directions to mimic isotropic fluctuations. Hessian matrix,  $\mathbf{H}$ , contains second derivatives of potentials. The inverse of Hessian matrix,  $\mathbf{H}^{-1}$ , contains residue covariances, and interpreted as a covariance matrix. The covariance matrices can be gathered from MD simulations, and also by using ENM of a protein. In this study, MD covariance matrices have been utilized to incorporate residue interactions accurately.

Residues with low DFI score (below 0.2) are considered as hinge points. These points are communication hubs in this 3D interaction network. Due to high coordination number, the residues exhibiting low DFI values are crucial as information gateways. While they do not exhibit high residue fluctuation to the perturbations, they quickly transfer the perturbation information to other parts, thus they are in control of collective motion of the protein. A change in low DFI positions (i.e., a mutation) will lead to a transformation in the communication grid and majority of disease-associated (i.e., function altering mutations) are often observed as hinges ([Butler et al., 2015](#); [Nevin Gerek et al., 2013](#); [Kumar et al., 2015a](#)). The substitution on these site usually alters catalytic activity or binding interaction (i.e., glycans) by modulating equilibrium dynamics ([Campitelli et al., 2020a](#)).

## Dynamic coupling index

DCI exploits the same framework of DFI ([Campitelli et al., 2020a](#); [Larrimore et al., 2017](#)). DCI utilizes the residue response fluctuation upon random force perturbation at a specific residue position to investigate residues that exhibit long-range coupling to each other. In DCI approach, a unit force is applied on functional residues (i.e., binding site residues) one by one and responses of all other residues are calculated.

$$DCI_i = \frac{\sum_{j=1}^{N_{\text{Functional}}} |\Delta R^j| / N_{\text{Functional}}}{\sum_{j=1}^N |\Delta R^j| / N}$$

With DCI scheme the residues with high response (high DCI score) indicate high long-range dynamic coupling. Residues with high DCI values with binding sites play a critical role in intercommunication of a protein with the binding residues. These coupled residues are of utmost importance in how forces propagate through amino acid chain network on a binding event. Some of the coupled residues are far from the binding site but still encompass modulation capabilities over binding pocket.

## Informing dynamics from co-evolution

Co-evolutionary data paves the way to assessing 3D structural contacts by utilizing available sequence information ([Hopf et al., 2018](#); [Marks et al., 2012](#); [Morcos et al., 2014](#)). Sequence information is more abundant compared to resolved protein structures. Exploiting the sequence information, primary contacts comparable to realistic structural contacts can be calculated and a contact matrix is formed. The accuracy of these contact maps is proved to be valuable in protein folding studies ([Kryshtafovych et al., 2019](#); [Morcos et al., 2011](#); [Wang et al., 2016](#)). EC analysis is used to collect information on how much two residues in a protein sequence are in close proximity in 3D structure. EC scores could be calculated by many different statistical approaches. In this study EC information is gathered by using RaptorX, EVcouplings, and MISTIC webservers ([Hopf et al., 2019](#); [Simonetti et al., 2013](#); [Wang et al., 2017](#)). While the limitation of these methods emerges from sequence homolog availability of a protein in MSA, RaptorX uses a deep neural network leveraging joint family approach, combining multiple ortholog protein families sharing similar function and phylogeny, to infer possible contacts. This method is proven to produce high accuracy in contact prediction compared to others ([Wang et al., 2017](#)). However, for a given MSA containing enough homolog sequences, other methods are also strong in predicting spatial contacts. EVcouplings approach uses direct information (DI) to calculate co-EC. DI metric is a modified mutual information (MI) score considering consistency between pairwise probabilities and single amino acid frequencies ([de Juan et al., 2013](#); [Morcos et al., 2011](#)). Nonetheless, MI, a global approach compared to local DI metric, is accurate in capturing true contacts, while entangling indirect contacts from direct contacts. MISTIC webserver has taken advantage of MI to calculate co-EC ([Dunn et al., 2008](#); [Gouveia-Oliveira and Pedersen, 2007](#); [Simonetti et al., 2013](#)). In their MI method they introduced a correction term to MI to surpass the low statistics gathered with an MSA containing limited number of sequences. This approach is very useful in cases where certain homologs are rare and MSA of these homologs have multiple gaps in their alignments. All of these methods are employed in this study to achieve high accuracy predictions in finding residue couplings.

## Mutant proteins cloning, expression, and purification

The genes for mutants (I34Y, I34K, and I34L) were generated by applying mutagenic primers to P51G-m4-gene sequence and amplifying by PCR. The constructs were subsequently cloned in pET26B vector between NdeI and XhoI sites and transformed in BL21(DE3) for expression and purification. The proteins were expressed from a 10 ml starter culture in LB broth overnight at 37°C, inoculated into 1 l LB medium. The culture was induced with 1 mM isopropyl thiogalactoside when OD reached 0.6 and grown for another 6–8 hr. Then, the cells were harvested by centrifugation, lysed in 6 M guanidine hydrochloride at pH 8.0, and sonicated for 10 min. The supernatant recovered after centrifugation was used to purify proteins with GE HisTrap HP column (GE Healthcare Bio-Sciences, Piscataway, NJ) and a Bio-Rad EconoPump (Bio-Rad, Richmond, CA) under denaturing conditions. In brief, the proteins were loaded on the column in Gu-HCl buffer, which was buffer exchanged by 8 M urea buffer. The nonspecific proteins were washed out by 4 M urea and 20 mM imidazole buffer, pH 8.0 and eluted with 2 M urea and 200 mM imidazole, pH 8.0 buffer before putting it for overnight dialysis against 10 mM Tris pH 8.0 and 100 mM NaCl buffer. The buffer was changed once during the night. The refolded protein was concentrated and re-purified to isolate the monomeric species by size exclusion chromatography using Sephadex 75 10/300 column on Agilent's Infinity 1260 system. The gel filtered protein was finally used for all the experiments.

## CD spectroscopy and T-melts

In CV-N family proteins, thermodynamic parameters like free energy of unfolding, enthalpy, and entropy cannot be extracted by thermal denaturation because the transition from folded to unfolded state is non-reversible (*Patsalo et al., 2011*), therefore melting temperatures are used. Far-UV CD spectra were recorded on a Jasco J-815 spectropolarimeter equipped with a thermostatic cell holder, PTC 424S. Spectra were measured from 250 to 200 nm, using a scanning speed of 50 nm/min and a data pitch of 1.0 nm at 25°C. Samples concentration was approximately 15  $\mu$ M in 10 mM Tris, pH 8.0, and 100 mM NaCl. For thermal denaturation experiments, the melting profile was monitored at 202 nm from 25°C to 90°C. The data points were plotted and fitted in Origin8.5 software to get apparent  $T_m$ .

## Isothermal titration calorimetry

ITC was performed at the Sanford-Burnham Medical Research Institute Protein Analysis Facility using ITC200 calorimeter from Microcal (Northampton, MA) at 23°C; 2.0  $\mu$ l aliquots of solution containing between 3 and 10 mM Man2 were injected into the cell containing between 0.057 and 0.11 mM protein. Nineteen of 2.0  $\mu$ l injections were made. The experiments were performed in 10 mM Tris, 100 mM NaCl, pH 8.0 buffer. ITC data were analyzed using Origin software provided by Microcal.

## Chemical denaturation experiments

Chemical denaturation experiments were done by monitoring the shift in the intrinsic tryptophan fluorescence on Cary Eclipse instrument (Varian). Ten  $\mu$ M of protein samples were incubated with increasing concentrations of guanidine hydrochloride in the range of 0–6 M in 50 mM Tris pH 8.0 buffer for 72 hr at 25°C. The emission spectra for the same were recorded by keeping the excitation wavelength at 295 nm and bandwidth of 1 nm. A ratio of fluorescence at 330 and 360 nm ( $I_{330}/I_{360}$ ) was plotted at respective Gu-HCl concentrations, and the data points were fit to following sigmoidal equation to obtain  $C_m$ .

$$y = A_2 + \frac{A_1 - A_2}{1 + e^{(x - x_0)/dx}}$$

where  $A_1$  and  $A_2$  are the initial and final 330/360 ratios and  $x_0$  is the concentration of Gu-HCl, where  $y = (A_1 + A_2)/2$ , or the point, where 50% of the population is unfolded. It is also denoted as  $C_m$ .

The denaturation curve was used to calculate the free energy of the protein in the absence of denaturant ( $\Delta G_{H_2O}$ ). Fraction unfolded ( $f_U$ ) was calculated using the following formula:

$$f_U = (Y_F - Y_{obs}) / (Y_F - Y_U)$$



where  $f_U$  is the fraction unfolded,  $y_F$  is the value when there is no denaturant,  $y_{obs}$  is the value at each position, and  $y_U$  is the value for unfolded protein. Since  $f_U + f_F = 1$ , the equilibrium constant,  $K$ , of the free energy change can be calculated using

$$K = f_U / f_F$$
$$K = f_U / (1 - f_U)$$
$$\Delta G = -RT \ln K$$

where  $R$  is the gas constant whose value is 1.987 cal/mol·K and  $T$  is the temperature of incubation, which was 298 K. The value of  $\Delta G$  is linear over a limited range of Gu-HCl. The linear fit over that range was extrapolated to obtain  $\Delta G_{H_2O}$ .

## Crystallization and structure determination

I34Y was purified as discussed previously and the monomeric gel filtered protein was concentrated to 8 mg/ml. We got the crystals in 2 M ammonium sulphate and 5% (v/v) 2-propanol after screening it in Index HT screen from Hampton Research. The protein crystals were reproduced using same condition in hanging drop method. For protein crystals with dimannose, the crystals were incubated in 1.2-fold molar excess of dimannose. Single needle-like crystals were picked up and cryo-preserved in 25% glycerol before freezing them for data collection at Synchrotron ALS, beamline 8.2.1. Single crystal diffraction was measured at wavelength of 0.999 Å with ADSC quantum 315r detector. The data were evaluated to resolution of 1.25 Å. The data acquired was indexed using XDS and scaled by the aimless package from CCP4i program suite. The structural coordinates and phase were determined by molecular replacement using 2RDK PDB code. The structure of I34Y of CV-N is deposited under PDB accession code 6X7H. The structure was further refined in Coot.

## Acknowledgements

SBO acknowledges support from the Gordon and Betty Moore Foundations and National Science Foundation (Award: 1715591 and 1901709). This work was supported in part by NIH award 1R21CA207832-01.

---

## Additional information

### Funding

| Funder                            | Grant reference number | Author                             |
|-----------------------------------|------------------------|------------------------------------|
| National Science Foundation       | 1715591                | S Banu Ozkan                       |
| Gordon and Betty Moore Foundation | 1901709                | S Banu Ozkan                       |
| National Institutes of Health     | 1R21CA207832-01        | Giovanna Ghirlanda<br>S Banu Ozkan |

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

### Author contributions

I Can Kazan, Conceptualization, Resources, Data curation, Software, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing – original draft, Writing – review and editing; Prerna Sharma, Conceptualization, Resources, Data curation, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing – original draft, Writing – review and editing; Mohammad Imtiazur Rahman, Resources, Formal analysis, Validation, Investigation, Visualization, Writing – review and editing; Andrey Bobkov, Raimund Fromme, Formal analysis; Giovanna Ghirlanda, S Banu Ozkan, Conceptualization, Resources, Supervision, Funding acquisition, Methodology, Writing – original draft, Project administration, Writing – review and editing

**Author ORCIDs**I Can Kazan  <http://orcid.org/0000-0003-2593-4179>Giovanna Ghirlanda  <http://orcid.org/0000-0001-5470-1484>S Banu Ozkan  <http://orcid.org/0000-0002-9351-3758>**Decision letter and Author response**Decision letter <https://doi.org/10.7554/eLife.67474.sa1>Author response <https://doi.org/10.7554/eLife.67474.sa2>

---

**Additional files****Supplementary files**

- Transparent reporting form
- Supplementary file 1. Dynamic flexibility index (DFI), dynamic coupling index (DCI), RaptorX, Evcoupling, and MISTIC metrics are used to identify residues in TEM-1  $\beta$ -lactamase for the four unique categories (**Supplementary file 5**).
- Supplementary file 2. Dynamic flexibility index (DFI), dynamic coupling index (DCI), RaptorX, Evcoupling, and MISTIC metrics are used to identify residues in cyanovirin-N (CV-N) for the integrated co-evolution and dynamic coupling (ICDC) categories (**Supplementary file 6**).
- Supplementary file 3. The predicted binding affinities of domain B and comparison with experimental isothermal titration calorimetry (ITC) data for wild type, mutDB, and P51G-m4 benchmarking.
- Supplementary file 4. Binding free energies, enthalpy and entropy values for wild-type cyanovirin-N (CV-N) and its variants calculated with MM-PBSA approach applied on dimannose bound molecular dynamics (MD) simulations.
- Supplementary file 5. Contains the complete TEM-1 dynamic flexibility index (DFI), dynamic coupling index (DCI), RaptorX, Evcoupling, and MISTIC metric data used in this study.
- Supplementary file 6. Contains the complete cyanovirin-N (CV-N) dynamic flexibility index (DFI), dynamic coupling index (DCI), RaptorX, Evcoupling, and MISTIC metric data used in this study.

**Data availability**

All data generated or analysed during this study are included in the manuscript and supporting files.

**References**

- Abraham MJ**, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. 2015. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**:19–25. DOI: <https://doi.org/10.1016/j.softx.2015.06.001>, PMID: 24793553
- Atilgan AR**, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. 2001. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophysical Journal* **80**:505–515. DOI: [https://doi.org/10.1016/S0006-3495\(01\)76033-X](https://doi.org/10.1016/S0006-3495(01)76033-X), PMID: 11159421
- Atilgan C**, Atilgan AR. 2009. Perturbation-response scanning reveals ligand entry-exit mechanisms of ferric binding protein. *PLOS Computational Biology* **5**:e1000544. DOI: <https://doi.org/10.1371/journal.pcbi.1000544>, PMID: 19851447
- Atilgan C**, Gerek ZN, Ozkan SB, Atilgan AR. 2010. Manipulation of conformational change in proteins by single-residue perturbations. *Biophysical Journal* **99**:933–943. DOI: <https://doi.org/10.1016/j.bpj.2010.05.020>, PMID: 20682272
- Balzarini J**. 2007. Targeting the glycans of glycoproteins: a novel paradigm for antiviral therapy. *Nature Reviews Microbiology* **5**:583–597. DOI: <https://doi.org/10.1038/nrmicro1707>, PMID: 17632570
- Barrientos LG**, O’Keefe BR, Bray M, Sanchez A, Gronenborn AM, Boyd MR. 2003. Cyanovirin-N binds to the viral surface glycoprotein, GP1,2 and inhibits infectivity of ebola virus. *Antiviral Research* **58**:47–56. DOI: [https://doi.org/10.1016/s0166-3542\(02\)00183-3](https://doi.org/10.1016/s0166-3542(02)00183-3), PMID: 12719006
- Barrientos LG**, Matei E, Lasala F, Delgado R, Gronenborn AM. 2006. Dissecting carbohydrate-cyanovirin-N binding by structure-guided mutagenesis: functional implications for viral entry inhibition. *Protein Engineering, Design & Selection* **19**:525–535. DOI: <https://doi.org/10.1093/protein/gzl040>, PMID: 17012344
- Bewley CA**. 2001. Solution structure of a cyanovirin-N: man alpha 1-2man alpha complex: structural basis for high-affinity carbohydrate-mediated binding to gp120. *Structure* **9**:931–940. DOI: [https://doi.org/10.1016/s0969-2126\(01\)00653-0](https://doi.org/10.1016/s0969-2126(01)00653-0), PMID: 11591348
- Bolia A**, Gerek ZN, Ozkan SB. 2014a. BP-dock: a flexible docking scheme for exploring protein-ligand interactions based on unbound structures. *Journal of Chemical Information and Modeling* **54**:913–925. DOI: <https://doi.org/10.1021/ci4004927>, PMID: 24380381

- Bolia A**, Woodrum BW, Cereda A, Ruben MA, Wang X, Ozkan SB, Ghirlanda G. 2014b. A flexible docking scheme efficiently captures the energetics of glycan-cyanovirin binding. *Biophysical Journal* **106**:1142–1151. DOI: <https://doi.org/10.1016/j.bpj.2014.01.040>
- Bolia A**, Ozkan SB. 2016. Adaptive BP-dock: an induced fit docking approach for full receptor flexibility. *Journal of Chemical Information and Modeling* **56**:734–746. DOI: <https://doi.org/10.1021/acs.jcim.5b00587>, PMID: 26971620
- Bolmstedt AJ**, O’Keefe BR, Shenoy SR, McMahon JB, Boyd MR. 2001. Cyanovirin-N defines a new class of antiviral agent targeting N-linked, high-mannose glycans in an oligosaccharide-specific manner. *Molecular Pharmacology* **59**:949–954. DOI: <https://doi.org/10.1124/mol.59.5.949>, PMID: 11306674
- Botos I**, Wlodawer A. 2003. Cyanovirin-N: a sugar-binding antiviral protein with a new twist. *Cellular and Molecular Life Sciences* **60**:277–287. DOI: <https://doi.org/10.1007/s000180300023>, PMID: 12678493
- Botos I**, Wlodawer A. 2005. Proteins that bind high-mannose sugars of the HIV envelope. *Progress in Biophysics and Molecular Biology* **88**:233–282. DOI: <https://doi.org/10.1016/j.pbiomolbio.2004.05.001>, PMID: 15572157
- Bowman GR**, Geissler PL. 2012. Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. *PNAS* **109**:11681–11686. DOI: <https://doi.org/10.1073/pnas.1209309109>, PMID: 22753506
- Breiten B**, Lockett MR, Sherman W, Fujita S, Al-Sayah M, Lange H, Bowers CM, Heroux A, Krilov G, Whitesides GM. 2013. Water networks contribute to enthalpy/entropy compensation in protein-ligand binding. *Journal of the American Chemical Society* **135**:15579–15584. DOI: <https://doi.org/10.1021/ja4075776>, PMID: 24044696
- Butler BM**, Gerek ZN, Kumar S, Ozkan SB. 2015. Conformational dynamics of nonsynonymous variants at protein interfaces reveals disease association. *Proteins* **83**:428–435. DOI: <https://doi.org/10.1002/prot.24748>, PMID: 25546381
- Campbell E**, Kaltenbach M, Correy GJ, Carr PD, Porebski BT, Livingstone EK, Afriat-Jurnou L, Buckle AM, Weik M, Hollfelder F, Tokuriki N, Jackson CJ. 2016. The role of protein dynamics in the evolution of new enzyme function. *Nature Chemical Biology* **12**:944–950. DOI: <https://doi.org/10.1038/nchembio.2175>, PMID: 27618189
- Campitelli P**, Guo J, Zhou HX, Ozkan SB. 2018. Hinge-shift mechanism modulates allosteric regulations in human Pin1. *The Journal of Physical Chemistry. B* **122**:5623–5629. DOI: <https://doi.org/10.1021/acs.jpcc.7b11971>, PMID: 29361231
- Campitelli P**, Modi T, Kumar S, Ozkan SB. 2020a. The role of conformational dynamics and allostery in modulating protein evolution. *Annual Review of Biophysics* **49**:267–288. DOI: <https://doi.org/10.1146/annurev-biophys-052118-115517>, PMID: 32075411
- Campitelli P**, Ozkan SB, Swint-Kruse L. 2020b. Asymmetry in dynamic allosteric residue coupling (DARC) interactions captures evolutionary landscape. *Biophysical Journal* **118**:52a. DOI: <https://doi.org/10.1016/j.bpj.2019.11.464>
- Campitelli P**, Swint-Kruse L, Ozkan SB. 2021. Substitutions at nonconserved rheostat positions modulate function by rewiring long-range, dynamic interactions. *Molecular Biology and Evolution* **38**:201–214. DOI: <https://doi.org/10.1093/molbev/msaa202>, PMID: 32780837
- Chang LC**, Bewley CA. 2002. Potent inhibition of HIV-1 fusion by cyanovirin-N requires only a single high affinity carbohydrate binding site: characterization of low affinity carbohydrate binding site knockout mutants. *Journal of Molecular Biology* **318**:1–8. DOI: [https://doi.org/10.1016/S0022-2836\(02\)00045-1](https://doi.org/10.1016/S0022-2836(02)00045-1), PMID: 12054763
- Chodera JD**, Mobley DL. 2013. Entropy-enthalpy compensation: role and ramifications in biomolecular ligand recognition and design. *Annual Review of Biophysics* **42**:121–142. DOI: <https://doi.org/10.1146/annurev-biophys-083012-130318>, PMID: 23654303
- Cornish-Bowden A**. 2002. Enthalpy-entropy compensation: a phantom phenomenon. *Journal of Biosciences* **27**:121–126. DOI: <https://doi.org/10.1007/BF02703768>, PMID: 11937682
- Davis IW**, Baker D. 2009. RosettaLigand docking with full ligand and receptor flexibility. *Journal of Molecular Biology* **385**:381–392. DOI: <https://doi.org/10.1016/j.jmb.2008.11.010>, PMID: 19041878
- Davis IW**, Raha K, Head MS, Baker D. 2009. Blind docking of pharmaceutically relevant compounds using rosettaligand. *Protein Science* **18**:1998–2002. DOI: <https://doi.org/10.1002/pro.192>, PMID: 19554568
- de Juan D**, Pazos F, Valencia A. 2013. Emerging methods in protein co-evolution. *Nature Reviews. Genetics* **14**:249–261. DOI: <https://doi.org/10.1038/nrg3414>, PMID: 23458856
- DeLuca S**, Khar K, Meiler J. 2015. Fully flexible docking of medium sized ligand libraries with rosettaligand. *PLOS ONE* **10**:e0132508. DOI: <https://doi.org/10.1371/journal.pone.0132508>, PMID: 26207742
- Deng Y**, Roux B. 2009. Computations of standard binding free energies with molecular dynamics simulations. *The Journal of Physical Chemistry. B* **113**:2234–2246. DOI: <https://doi.org/10.1021/jp807701h>, PMID: 19146384
- Dunn SD**, Wahl LM, Gloor GB. 2008. Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. *Bioinformatics* **24**:333–340. DOI: <https://doi.org/10.1093/bioinformatics/btm604>, PMID: 18057019
- Fox JM**, Zhao M, Fink MJ, Kang K, Whitesides GM. 2018. The molecular origin of enthalpy/entropy compensation in biomolecular recognition. *Annual Review of Biophysics* **47**:223–250. DOI: <https://doi.org/10.1146/annurev-biophys-070816-033743>, PMID: 29505727
- Fromme R**, Katilene Z, Giomarelli B, Bogani F, Mc Mahon J, Mori T, Fromme P, Ghirlanda G. 2007. A monovalent mutant of cyanovirin-N provides insight into the role of multiple interactions with gp120 for antiviral activity. *Biochemistry* **46**:9199–9207. DOI: <https://doi.org/10.1021/bi700666m>, PMID: 17636873

- Fromme R**, Katilene Z, Fromme P, Ghirlanda G. 2008. Conformational gating of dimannose binding to the antiviral protein cyanovirin revealed from the crystal structure at 1.35 Å resolution. *Protein Science* **17**:939–944. DOI: <https://doi.org/10.1110/ps.083472808>, PMID: 18436959
- Fujimoto YK**, Green DF. 2012. Carbohydrate recognition by the antiviral lectin cyanovirin-N. *Journal of the American Chemical Society* **134**:19639–19651. DOI: <https://doi.org/10.1021/ja305755b>, PMID: 23057413
- Gerek ZN**, Ozkan SB. 2011. Change in allosteric network affects binding affinities of PDZ domains: analysis through perturbation response scanning. *PLoS Computational Biology* **7**:e1002154. DOI: <https://doi.org/10.1371/journal.pcbi.1002154>, PMID: 21998559
- Gouveia-Oliveira R**, Pedersen AG. 2007. Finding coevolving amino acid residues using row and column weighting of mutual information and multi-dimensional amino acid representation. *Algorithms for Molecular Biology* **2**:12. DOI: <https://doi.org/10.1186/1748-7188-2-12>, PMID: 17915013
- He X**, Liu S, Lee TS, Ji B, Man VH, York DM, Wang J. 2020. Fast, accurate, and reliable protocols for routine calculations of protein-ligand binding affinities in drug design projects using amber GPU-TI with ff14sb/GAFF. *ACS Omega* **5**:4611–4619. DOI: <https://doi.org/10.1021/acsomega.9b04233>, PMID: 32175507
- Hopf TA**, Schärfe CPI, Rodrigues J, Green AG, Kohlbacher O, Sander C, Bonvin A, Marks DS. 2018. Sequence co-evolution gives 3D contacts and structures of protein complexes. *eLife* **3**:e03430. DOI: <https://doi.org/10.7554/eLife.03430>, PMID: 25255213
- Hopf TA**, Green AG, Schubert B, Mersmann S, Schärfe CPI, Ingraham JB, Toth-Petroczy A, Brock K, Riesselman AJ, Palmedo P, Kang C, Sheridan R, Draizen EJ, Dallago C, Sander C, Marks DS. 2019. The eCouplings python framework for coevolutionary sequence analysis. *Bioinformatics* **35**:1582–1584. DOI: <https://doi.org/10.1093/bioinformatics/bty862>, PMID: 30304492
- Ikeguchi M**, Ueno J, Sato M, Kidera A. 2005. Protein structural change upon ligand binding: linear response theory. *Physical Review Letters* **94**:078102. DOI: <https://doi.org/10.1103/PhysRevLett.94.078102>, PMID: 15783858
- Jana B**, Morcos F, Onuchic JN. 2014. From structure to function: the convergence of structure based models and co-evolutionary information. *Physical Chemistry Chemical Physics* **16**:6496–6507. DOI: <https://doi.org/10.1039/c3cp55275f>, PMID: 24603809
- Jiménez-Osés G**, Osuna S, Gao X, Sawaya MR, Gilson L, Collier SJ, Huisman GW, Yeates TO, Tang Y, Houk KN. 2014. The role of distant mutations and allosteric regulation on lovd active site dynamics. *Nature Chemical Biology* **10**:431–436. DOI: <https://doi.org/10.1038/nchembio.1503>, PMID: 24727900
- Jumper J**, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, et al. 2021. Highly accurate protein structure prediction with alphafold. *Nature* **596**:583–589. DOI: <https://doi.org/10.1038/s41586-021-03819-2>, PMID: 34265844
- Kamisetty H**, Ovchinnikov S, Baker D. 2013. Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era. *PNAS* **110**:15674–15679. DOI: <https://doi.org/10.1073/pnas.1314045110>, PMID: 24009338
- Kelley BS**, Chang LC, Bewley CA. 2002. Engineering an obligate domain-swapped dimer of cyanovirin-N with enhanced anti-HIV activity. *Journal of the American Chemical Society* **124**:3210–3211. DOI: <https://doi.org/10.1021/ja025537m>, PMID: 11916396
- Kim DE**, Dimaio F, Yu-Ruei Wang R, Song Y, Baker D. 2014. One contact for every twelve residues allows robust and accurate topology-level protein structure modeling. *Proteins* **82 Suppl 2**:208–218. DOI: <https://doi.org/10.1002/prot.24374>, PMID: 23900763
- Kim H**, Zou T, Modi C, Dörner K, Grunkemeyer TJ, Chen L, Fromme R, Matz MV, Ozkan SB, Wachter RM. 2015. A hinge migration mechanism unlocks the evolution of green-to-red photoconversion in GFP-like proteins. *Structure* **23**:34–43. DOI: <https://doi.org/10.1016/j.str.2014.11.011>, PMID: 25565105
- Koharudin LMI**, Viscomi AR, Jee JG, Ottonello S, Gronenborn AM. 2008. The evolutionarily conserved family of cyanovirin-N homologs: structures and carbohydrate specificity. *Structure* **16**:570–584. DOI: <https://doi.org/10.1016/j.str.2008.01.015>, PMID: 18400178
- Koharudin LMI**, Furey W, Gronenborn AM. 2009. A designed chimeric cyanovirin-N homolog lectin: structure and molecular basis of sucrose binding. *Proteins* **77**:904–915. DOI: <https://doi.org/10.1002/prot.22514>, PMID: 19639634
- Koharudin LMI**, Gronenborn AM. 2013. Sweet entanglements -- protein: glycan interactions in two HIV-inactivating lectin families. *Biopolymers* **99**:196–202. DOI: <https://doi.org/10.1002/bip.22106>, PMID: 23023834
- Kolbaba-Kartchner B**, Kazan IC, Mills JH, Ozkan SB. 2021. The role of rigid residues in modulating TEM-1 β-lactamase function and thermostability. *International Journal of Molecular Sciences* **22**:2895. DOI: <https://doi.org/10.3390/ijms22062895>, PMID: 33809335
- Kryshtafovych A**, Schwede T, Topf M, Fidelis K, Moutl J. 2019. Critical assessment of methods of protein structure prediction (CASP) -round XIII. *Proteins* **87**:1011–1020. DOI: <https://doi.org/10.1002/prot.25823>, PMID: 31589781
- Kumar A**, Butler BM, Kumar S, Ozkan SB. 2015a. Integration of structural dynamics and molecular evolution via protein interaction networks: a new era in genomic medicine. *Current Opinion in Structural Biology* **35**:135–142. DOI: <https://doi.org/10.1016/j.sbi.2015.11.002>, PMID: 26684487
- Kumar A**, Glemba TJ, Ozkan SB. 2015b. The role of conformational dynamics and allostery in the disease development of human ferritin. *Biophysical Journal* **109**:1273–1281. DOI: <https://doi.org/10.1016/j.bpj.2015.06.060>, PMID: 26255589

- Kuriyan J, Eisenberg D. 2007. The origin of protein interactions and allostery in colocalization. *Nature* **450**:983–990. DOI: <https://doi.org/10.1038/nature06524>, PMID: 18075577
- Larrimore KE, Kazan IC, Kannan L, Kendle RP, Jamal T, Barcus M, Bolia A, Brimijoin S, Zhan C-G, Ozkan SB, Mor TS. 2017. Plant-expressed cocaine hydrolase variants of butyrylcholinesterase exhibit altered allosteric effects of cholinesterase activity and increased inhibitor sensitivity. *Scientific Reports* **7**:10419. DOI: <https://doi.org/10.1038/s41598-017-10571-z>, PMID: 28874829
- Li Z, Bolia A, Maxwell JD, Bobkov AA, Ghirlanda G, Ozkan SB, Margulis CJ. 2015. A rigid hinge region is necessary for high-affinity binding of dimannose to cyanovirin and associated constructs. *Biochemistry* **54**:6951–6960. DOI: <https://doi.org/10.1021/acs.biochem.5b00635>, PMID: 26507789
- Liu J, Nussinov R. 2017. Energetic redistribution in allostery to execute protein function. *PNAS* **114**:7480–7482. DOI: <https://doi.org/10.1073/pnas.1709071114>, PMID: 28696318
- Lu HM, Liang J. 2009. Perturbation-Based markovian transmission model for probing allosteric dynamics of large macromolecular assembling: a study of GroEL-GroES. *PLoS Computational Biology* **5**:e1000526. DOI: <https://doi.org/10.1371/journal.pcbi.1000526>, PMID: 19798437
- Marks DS, Hopf TA, Sander C. 2012. Protein structure prediction from sequence variation. *Nature Biotechnology* **30**:1072–1080. DOI: <https://doi.org/10.1038/nbt.2419>, PMID: 23138306
- Matei E, Furey W, Gronenborn AM. 2008. Solution and crystal structures of a sugar binding site mutant of cyanovirin-N: no evidence of domain swapping. *Structure* **16**:1183–1194. DOI: <https://doi.org/10.1016/j.str.2008.05.011>, PMID: 18682220
- Matei E, Louis JM, Jee J, Gronenborn AM. 2011. Nmr solution structure of a cyanovirin homolog from wheat head blight fungus. *Proteins* **79**:1538–1549. DOI: <https://doi.org/10.1002/prot.22981>, PMID: 21365681
- Matei E, Basu R, Furey W, Shi J, Calnan C, Aiken C, Gronenborn AM. 2016. Structure and glycan binding of a new cyanovirin-N homolog. *The Journal of Biological Chemistry* **291**:18967–18976. DOI: <https://doi.org/10.1074/jbc.M116.740415>, PMID: 27402833
- Meiler J, Baker D. 2006. ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility. *Proteins* **65**:538–548. DOI: <https://doi.org/10.1002/prot.21086>, PMID: 16972285
- Meinhardt S, Manley MW, Parente DJ, Swint-Kruse L. 2013. Rheostats and toggle switches for modulating protein function. *PLOS ONE* **8**:e83502. DOI: <https://doi.org/10.1371/journal.pone.0083502>, PMID: 24386217
- Miller M, Bromberg Y, Swint-Kruse L. 2017. Computational predictors fail to identify amino acid substitution effects at rheostat positions. *Scientific Reports* **7**:41329. DOI: <https://doi.org/10.1038/srep41329>, PMID: 28134345
- Modi T, Huihui J, Ghosh K, Ozkan SB. 2018. Ancient thioredoxins evolved to modern-day stability-function requirement by altering native state ensemble. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **373**:20170184. DOI: <https://doi.org/10.1098/rstb.2017.0184>, PMID: 29735738
- Modi T, Ozkan SB. 2018. Mutations utilize dynamic allostery to confer resistance in TEM-1  $\beta$ -lactamase. *International Journal of Molecular Sciences* **19**:3808. DOI: <https://doi.org/10.3390/ijms19123808>, PMID: 30501088
- Modi T, Campitelli P, Kazan IC, Ozkan SB. 2021a. Protein folding stability and binding interactions through the lens of evolution: a dynamical perspective. *Current Opinion in Structural Biology* **66**:207–215. DOI: <https://doi.org/10.1016/j.sbi.2020.11.007>, PMID: 33388636
- Modi T, Risso VA, Martinez-Rodriguez S, Gavira JA, Mebrat MD, Van Horn WD, Sanchez-Ruiz JM, Banu Ozkan S. 2021b. Hinge-shift mechanism as a protein design principle for the evolution of  $\beta$ -lactamases from substrate promiscuity to specificity. *Nature Communications* **12**:1852. DOI: <https://doi.org/10.1038/s41467-021-22089-0>, PMID: 33767175
- Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, Sander C, Zecchina R, Onuchic JN, Hwa T, Weigt M. 2011. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *PNAS* **108**:E1293–E1301. DOI: <https://doi.org/10.1073/pnas.1111471108>, PMID: 22106262
- Morcos F, Jana B, Hwa T, Onuchic JN. 2013. Coevolutionary signals across protein lineages help capture multiple protein conformations. *PNAS* **110**:20533–20538. DOI: <https://doi.org/10.1073/pnas.1315625110>, PMID: 24297889
- Morcos F, Schafer NP, Cheng RR, Onuchic JN, Wolynes PG. 2014. Coevolutionary information, protein folding landscapes, and the thermodynamics of natural selection. *PNAS* **111**:12408–12413. DOI: <https://doi.org/10.1073/pnas.1413575111>, PMID: 25114242
- Morcos F. 2020. Protein conformations à la carte, a step further in de novo protein design. *PNAS* **117**:8674–8676. DOI: <https://doi.org/10.1073/pnas.2004188117>, PMID: 32291334
- Mori T, Boyd MR. 2001. Cyanovirin-N, a potent human immunodeficiency virus-inactivating protein, blocks both CD4-dependent and CD4-independent binding of soluble gp120 (sgp120) to target cells, inhibits scd4-induced binding of sgp120 to cell-associated CXCR4, and dissociates bound sgp120 from target cells. *Antimicrobial Agents and Chemotherapy* **45**:664–672. DOI: <https://doi.org/10.1128/AAC.45.3.664-672.2001>, PMID: 11181340
- Nevin Gerek Z, Kumar S, Banu Ozkan S. 2013. Structural dynamics flexibility informs function and evolution at a proteome scale. *Evolutionary Applications* **6**:423–433. DOI: <https://doi.org/10.1111/eva.12052>, PMID: 23745135
- Okazaki K, Koga N, Takada S, Onuchic JN, Wolynes PG. 2006. Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: structure-based molecular dynamics simulations. *PNAS* **103**:11844–11849. DOI: <https://doi.org/10.1073/pnas.0604375103>, PMID: 16877541

- O'Keefe BR**, Smee DF, Turpin JA, Saucedo CJ, Gustafson KR, Mori T, Blakeslee D, Buckheit R, Boyd MR. 2003. Potent anti-influenza activity of cyanovirin-N and interactions with viral hemagglutinin. *Antimicrobial Agents and Chemotherapy* **47**:2518–2525. DOI: <https://doi.org/10.1128/AAC.47.8.2518-2525.2003>, PMID: 12878514
- Ose N**, Butler BM, Kumar A, Ozkan SB, Kumar S. 2020. Dynamic allosteric residue coupling reveals disease mechanism for gaucher disease and NSNVs across the proteome. *Biophysical Journal* **118**:53a. DOI: <https://doi.org/10.1016/j.bpj.2019.11.472>
- Patsalo V**, Raleigh DP, Green DF. 2011. Rational and computational design of stabilized variants of cyanovirin-N that retain affinity and specificity for glycan ligands. *Biochemistry* **50**:10698–10712. DOI: <https://doi.org/10.1021/bi201411c>, PMID: 22032696
- Percudani R**, Montanini B, Ottonello S. 2005. The anti-HIV cyanovirin-N domain is evolutionarily conserved and occurs as a protein module in eukaryotes. *Proteins* **60**:670–678. DOI: <https://doi.org/10.1002/prot.20543>, PMID: 16003744
- Qi X**, Yang Y, Su Y, Wang T. 2009. Molecular cloning and sequence analysis of cyanovirin-N homology gene in *Ceratopteris thalictroides*. *American Fern Journal* **99**:78–92. DOI: <https://doi.org/10.1640/0002-8444-99.2.78>
- Ramadugu SK**, Li Z, Kashyap HK, Margulis CJ. 2014. The role of Glu41 in the binding of dimannose to P51G-m4-CVN. *Biochemistry* **53**:1477–1484. DOI: <https://doi.org/10.1021/bi4014159>, PMID: 24524298
- Rastelli G**, Del Rio A, Degliesposti G, Sgobba M. 2010. Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. *Journal of Computational Chemistry* **31**:797–810. DOI: <https://doi.org/10.1002/jcc.21372>, PMID: 19569205
- Risso VA**, Sanchez-Ruiz JM, Ozkan SB. 2018. Biotechnological and protein-engineering implications of ancestral protein resurrection. *Current Opinion in Structural Biology* **51**:106–115. DOI: <https://doi.org/10.1016/j.sbi.2018.02.007>, PMID: 29660672
- Rivoire O**, Reynolds KA, Ranganathan R. 2016. Evolution-based functional decomposition of proteins. *PLOS Computational Biology* **12**:e1004817. DOI: <https://doi.org/10.1371/journal.pcbi.1004817>, PMID: 27254668
- Romero PA**, Arnold FH. 2009. Exploring protein fitness landscapes by directed evolution. *Nature Reviews Molecular Cell Biology* **10**:866–876. DOI: <https://doi.org/10.1038/nrm2805>, PMID: 19935669
- Salinas VH**, Ranganathan R. 2018. Coevolution-based inference of amino acid interactions underlying protein function. *eLife* **7**:e34300. DOI: <https://doi.org/10.7554/eLife.34300>, PMID: 30024376
- Salverda MLM**, De Visser JAGM, Barlow M. 2010. Natural evolution of TEM-1  $\beta$ -lactamase: experimental reconstruction and clinical relevance. *FEMS Microbiology Reviews* **34**:1015–1036. DOI: <https://doi.org/10.1111/j.1574-6976.2010.00222.x>, PMID: 20412308
- Schneider SH**, Kozuch J, Boxer SG. 2021. The interplay of electrostatics and chemical positioning in the evolution of antibiotic resistance in TEM  $\beta$ -lactamases. *ACS Central Science* **7**:1996–2008. DOI: <https://doi.org/10.1021/acscentsci.1c00880>, PMID: 34963893
- Simonetti FL**, Teppa E, Chernomoretz A, Nielsen M, Marino Buslje C. 2013. MISTIC: mutual information server to infer coevolution. *Nucleic Acids Res* **41**:W8–W14. DOI: <https://doi.org/10.1093/nar/gkt427>
- Stiffler MA**, Hekstra DR, Ranganathan R. 2015. Evolvability as a function of purifying selection in TEM-1  $\beta$ -lactamase. *Cell* **160**:882–892. DOI: <https://doi.org/10.1016/j.cell.2015.01.035>, PMID: 25723163
- Swint-Kruse L**, Matthews KS, Smith PE, Pettitt BM. 1998. Comparison of simulated and experimentally determined dynamics for a variant of the lacI DNA-binding domain, nlac-P. *Biophysical Journal* **74**:413–421. DOI: [https://doi.org/10.1016/s0006-3495\(98\)77798-7](https://doi.org/10.1016/s0006-3495(98)77798-7)
- Thomas VL**, McReynolds AC, Shoichet BK. 2010. Structural bases for stability-function tradeoffs in antibiotic resistance. *Journal of Molecular Biology* **396**:47–59. DOI: <https://doi.org/10.1016/j.jmb.2009.11.005>, PMID: 19913034
- Tokuriki N**, Jackson CJ, Afriat-Jurnou L, Wyganowski KT, Tang R, Tawfik DS. 2012. Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. *Nature Communications* **3**:1257. DOI: <https://doi.org/10.1038/ncomms2246>, PMID: 23212386
- Torgeson KR**, Clarkson MW, Granata D, Lindorff-Larsen K, Page R, Peti W. 2022. Conserved conformational dynamics determine enzyme activity. *Science Advances* **8**:eabo5546. DOI: <https://doi.org/10.1126/sciadv.abo5546>, PMID: 35921420
- Tripathi S**, Waxham MN, Cheung MS, Liu Y. 2015. Lessons in protein design from combined evolution and conformational dynamics. *Scientific Reports* **5**:14259. DOI: <https://doi.org/10.1038/srep14259>, PMID: 26388515
- Van Der Spoel D**, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. 2005. GROMACS: fast, flexible, and free. *Journal of Computational Chemistry* **26**:1701–1718. DOI: <https://doi.org/10.1002/jcc.20291>, PMID: 16211538
- Vorontsov II**, Miyashita O. 2009. Solution and crystal molecular dynamics simulation study of m4-cyanovirin-N mutants complexed with di-mannose. *Biophysical Journal* **97**:2532–2540. DOI: <https://doi.org/10.1016/j.bpj.2009.08.011>, PMID: 19883596
- Wagner JR**, Sørensen J, Hensley N, Wong C, Zhu C, Perison T, Amaro RE. 2017. POVME 3.0: software for mapping binding pocket flexibility. *Journal of Chemical Theory and Computation* **13**:4584–4592. DOI: <https://doi.org/10.1021/acs.jctc.7b00500>, PMID: 28800393
- Wang R**, Lai L, Wang S. 2002. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *Journal of Computer-Aided Molecular Design* **16**:11–26. DOI: <https://doi.org/10.1023/a:1016357811882>, PMID: 12197663

- Wang S**, Li W, Zhang R, Liu S, Xu J. 2016. CoinFold: a web server for protein contact prediction and contact-assisted protein folding. *Nucleic Acids Research* **44**:W361–W366. DOI: <https://doi.org/10.1093/nar/gkw307>, PMID: 27112569
- Wang S**, Sun S, Li Z, Zhang R, Xu J. 2017. Accurate de novo prediction of protein contact map by ultra-deep learning model. *PLoS Computational Biology* **13**:e1005324. DOI: <https://doi.org/10.1371/journal.pcbi.1005324>, PMID: 28056090
- Wei G**, Xi W, Nussinov R, Ma B. 2016. Protein ensembles: how does nature harness thermodynamic fluctuations for life? the diverse functional roles of conformational ensembles in the cell. *Chemical Reviews* **116**:6516–6551. DOI: <https://doi.org/10.1021/acs.chemrev.5b00562>, PMID: 26807783
- Wodak SJ**, Paci E, Dokholyan NV, Berezovsky IN, Horovitz A, Li J, Hilser VJ, Bahar I, Karanicolas J, Stock G, Hamm P, Stote RH, Eberhardt J, Chebaro Y, Dejaegere A, Cecchini M, Changeux JP, Bolhuis PG, Vreede J, Faccioli P, et al. 2019. Allostery in its many disguises: from theory to applications. *Structure* **27**:566–578. DOI: <https://doi.org/10.1016/j.str.2019.01.003>, PMID: 30744993
- Woodrum BW**, Maxwell JD, Bolia A, Ozkan SB, Ghirlanda G. 2013. The antiviral lectin cyanovirin-N: probing multivalency and glycan recognition through experimental and computational approaches. *Biochemical Society Transactions* **41**:1170–1176. DOI: <https://doi.org/10.1042/BST20130154>, PMID: 24059504
- Xu J**. 2019. Distance-based protein folding powered by deep learning. *PNAS* **116**:16856–16865. DOI: <https://doi.org/10.1073/pnas.1821309116>, PMID: 31399549
- Yang G**, Hong N, Baier F, Jackson CJ, Tokuriki N. 2016. Conformational tinkering drives evolution of a promiscuous activity through indirect mutational effects. *Biochemistry* **55**:4583–4593. DOI: <https://doi.org/10.1021/acs.biochem.6b00561>, PMID: 27444875
- Zimmerman MI**, Hart KM, Sibbald CA, Frederick TE, Jimah JR, Knoverek CR, Tolia NH, Bowman GR. 2017. Prediction of new stabilizing mutations based on mechanistic insights from Markov state models. *ACS Central Science* **3**:1311–1321. DOI: <https://doi.org/10.1021/acscentsci.7b00465>, PMID: 29296672
- Zou T**, Riso VA, Gavira JA, Sanchez-Ruiz JM, Ozkan SB. 2015. Evolution of conformational dynamics determines the conversion of a promiscuous generalist into a specialist enzyme. *Molecular Biology and Evolution* **32**:132–143. DOI: <https://doi.org/10.1093/molbev/msu281>, PMID: 25312912
- Zou T**, Woodrum BW, Halloran N, Campitelli P, Bobkov AA, Ghirlanda G, Ozkan SB. 2021. Local interactions that contribute minimal frustration determine foldability. *The Journal of Physical Chemistry. B* **125**:2617–2626. DOI: <https://doi.org/10.1021/acs.jpcc.1c00364>, PMID: 33687216