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

I. Can Kazan[1,2] ¶, Prerna Sharma[1] ¶, Mohammad Imtiazur Rahman[1] ¶, Andrey Bobkov[3], Raimund Fromme[1], Giovanna Ghirlanda[1]*, S. Banu Ozkan[2] *

[1]School of Molecular Sciences, Arizona State University, Tempe, AZ

[2]Center for Biological Physics and Department of Physics, Arizona State University, Tempe, AZ

[3]Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA

*Corresponding authors

Email: Giovanna.Ghirlanda@asu.edu

Banu.Ozkan@asu.edu

¶These authors contributed equally to this work.

Abstract

We develop integrated coevolution 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 β-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 general 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.

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 3-D 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 3-D 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 3-D protein structures combine these methods with machine learning strategies i.e., 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; Larrimore 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, 2018; Modi and Ozkan, 2018; Zou et al., 2021, 2015). DCI measures the dynamic coupling between residue pairs and thus identifies positions most strongly

70 coupled to active/binding sites; these positions point to possible allosteric regulation sites
71 important for modulating function in adaptation and evolution (Butler et al., 2015; Campitelli et
72 al., 2020a, 2021; Kuriyan and Eisenberg, 2007; Lu and Liang, 2009; Modi and Ozkan, 2018; Ose
73 et al., 2020; Risso et al., 2018; Wodak et al., 2019).
74
75 In this paper, we present integrated co-evolution and dynamic coupling (ICDC) approach
76 to identify distal allosteric sites, and to assess and predict the effects of mutations on these sites on
77 function. We propose a system to classify residue positions in a binary fashion based on co-
78 evolution (co-evolved, 1 or not, 0) and dynamic coupling by DFI and DCI (dynamically coupled
79 1, or not ,0) with the functionally critical sites. This classification captures the complementarity of
80 dynamics-based and sequence-based methods. We hypothesize that positions belonging to
81 category **(1,1)**, i.e., positions both co-evolved and dynamically coupled with the functional sites,
82 will have the largest effect on function.
83
84 We validate our hypothesis first by analyzing the existing mutational fitness data for TEM-
85 1 β-lactamase, available for every position of the protein (Stiffler et al., 2015). In agreement with
86 our hypothesis, we find that mutations on category **(1,1)** positions significantly modulate the
87 function. A large fraction of mutations enhancing enzymatic activity correspond to category **(1,1)**
88 irrespective of distance from the active site. Second, we apply our ICDC approach to blindly
89 predict and experimentally validate mutations that allosterically modulate dimannose binding in a
90 natural lectin, Cyanovirin-N (CV-N). CV-N binds dimannose with nanomolar affinity and
91 remarkable specificity (Barrientos et al., 2003; Botos and Wlodawer, 2005, 2003; Mori and Boyd,
92 2001; O'Keefe et al., 2003). It is part of the CV-N family, found in a wide range of organisms

93 including cyanobacterium, ascomycetous fungi, and fern (Koharudin et al., 2008; Koharudin and
94 Gronenborn, 2013; Patsalo et al., 2011; Percudani et al., 2005; Qi et al., 2009). While the 3D folds
95 is remarkably conserved in all experimentally characterized members, the affinity and specificity
96 for different glycans and, in particular, to dimannose varies significantly (Koharudin et al., 2009,
97 2008; Matei et al., 2016; Woodrum et al., 2013). To design CV-N variants with improved binding
98 affinities for dimannose based on distal allosteric coupling, we binned each position in one of the
99 four categories based on computed DFI, DCI and co-evolution rates. We explored mutations at
100 these sites based on frequency in the sequence alignment. After obtaining the mutant models
101 through Molecular dynamics (MD) simulations, we assessed the impact of each naturally observed
102 mutation on binding affinity by docking dimannose to the mutant models via Adaptive BP-Dock
103 (Bolia et al., 2014b, 2014a; Bolia and Ozkan, 2016). We chose position I34, which belongs to
104 category **(1,1)** and is 16Å away from the binding pocket, for experimental validation. We found
105 that mutations I34K/L/Y had a diverse effect on glycan binding, either improving by two-fold or
106 abolishing completely. Through experimental and MD studies we show that the observed
107 improvement in binding affinity is due to changes in the dynamics of residues in the binding
108 pocket; mutation I34Y leads to rigidification of binding sites, thus compensating the entropic cost
109 of binding (Breiten et al., 2013; Chodera and Mobley, 2013; Cornish-Bowden, 2002; Fox et al.,
110 2018). Mutations at an additional position (A71T/S) from category **(1,1)** showed evidence of the
111 same allosteric mechanism governing the modulation of binding dynamics. Overall, this study
112 provides not only a new approach to identify distal sites whose mutations modulate binding
113 affinity, but also sheds light into mechanistic insights on how distal mutations modulate binding
114 affinity through dynamics allostery.
115

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 as well as the role of rigidity versus flexibility in allosterically modulating active/binding site dynamics. To this extend, we created four unique categories that classifies 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 table 1 & 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 beta lactamase TEM-1, 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

139 impacted activity. Mutations at other positions also affected activity; while most mutations were
140 deleterious, surprisingly, others resulted in increased activity. The impact of mutations on
141 dynamics and function of TEM-1 have been heavily explored but the distal mutational effects are
142 still poorly understood (Kolbaba-Kartchner et al., 2021; Modi et al., 2021b; Modi and Ozkan,
143 2018; Salverda et al., 2010; Schneider et al., 2021; Stiffler et al., 2015; Thomas et al., 2010;
144 Zimmerman et al., 2017; Zou et al., 2015). We applied our approach by obtaining DFI, DCI and
145 co-evolution scores for every position of TEM-1 and binning residue positions into each ICDC
146 category (Table S1). We constructed fitness distributions for each category using the
147 experimentally measured single mutant relative fitness values for all mutations per position
148 provided in the dataset (Fig 1).
149

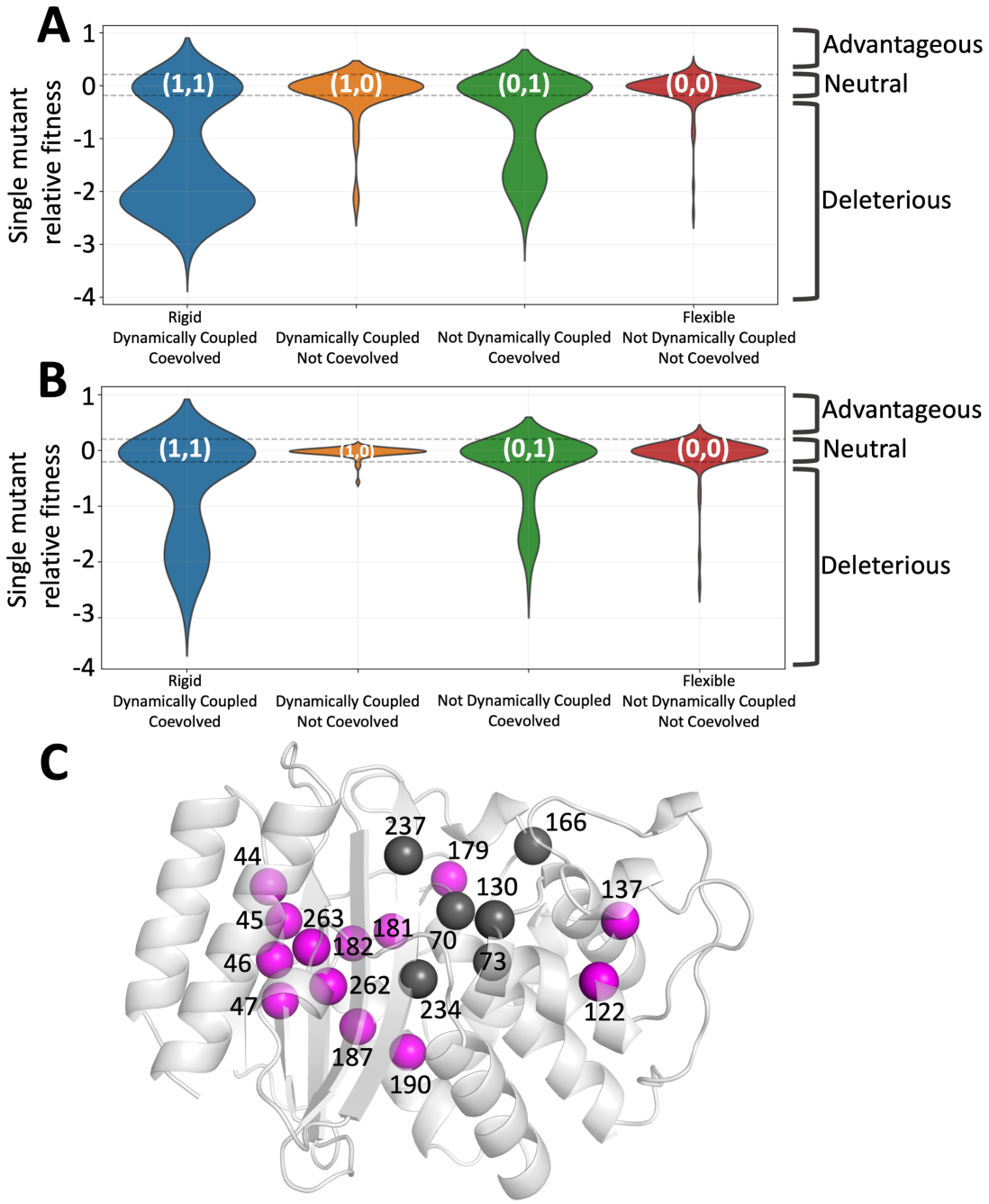

Fig 1 ICDC categories based on the dynamics and co-evolutionary analyses applied on TEM-1 β-lactamase. A) The distributions in the form of violin plots are obtained for each ICDC category using all available experimental mutational data (Stiffler et al., 2015) B) Violin plots showing the fitness values for amino acid substitutions observed in the natural sequences. C) The category **(1,1)** positions are mapped on 3-D structure. The catalytic site residues are shown in dark grey whereas category **(1,1)**

positions are shown in magenta color. The function altering category **(1,1**) positions are widely distributed over the 3-D structure.

150

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

165

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

169

170

171

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

173

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

181

182 To simplify our analyses, we used a designed CV-N variant, P51G-m4, that contains a
183 single high-affinity dimannose binding site (domain B), folds exclusively as a monomer in
184 physiological conditions, and is more stable to thermal denaturation than wild type (Fromme et
185 al., 2008, 2007). The binding pocket of domain B of CV-N has been subjected to intense scrutiny
186 to glean information on the origin of its binding specificity for dimannose (Bewley, 2001; Bolia et
187 al., 2014b; Botos and Wlodawer, 2003; Li et al., 2015; Vorontsov and Miyashita, 2009). Previous
188 mutational studies on the binding pocket residues have shown their importance in modulating
189 interaction with dimannose (Barrientos et al., 2006; Bolia et al., 2014b; Chang and Bewley, 2002;
190 Matei et al., 2008). All known substitutions of the binding residues led to decreased binding
191 affinity for dimannose on domain B (Bolia et al., 2014b; Fujimoto and Green, 2012; Kelley et al.,
192 2002; Matei et al., 2011; Ramadugu et al., 2014). Evolutionary analyses shows that the majority
193 of the binding site residues are conserved in CV-N glycan interactions, suggesting that affinity is
194 already optimized at the binding site. (Koharudin et al., 2008; Percudani et al., 2005). We

195 hypothesized that amino acid substitutions at distal positions could enhance the dimannose affinity
196 of CV-N by rigidification of the binding site and applied our ICDC approach to CV-N to identify
197 positions in each category (Table S2) (Supplementary file 1 table 2).
198
199 We generated models of CV-N variants in each ICDC category by mutating these positions
200 to amino acid types observed in the MSA of CV-N family members, choosing the subset of
201 sequences that have binding sites with identical or similar amino-acid composition to P51G-m4
202 CV-N. As discussed above this approach allows us to identify amino-acid substitutions with the
203 least impact on fold. All the substitutions identified (104 variants in total) were modeled using the
204 crystal structure of P51G-m4 CV-N (Fromme et al., 2008) and subjected to MD simulations
205 (Abraham et al., 2015; Spoel et al., 2005). The best conformation sampled for each variant obtained
206 from equilibrated production trajectories was used as a model for dimannose docking analysis. We
207 evaluated the variants using Adaptive BP-Dock (Bolia and Ozkan, 2016), a computational docking
208 tool that incorporates both ligand and receptor flexibility to accurately sample binding induced
209 conformations, and ranks them using X-scores binding energy units (XEUs) . In previous work on
210 CV-N this method yielded good correlations with experimentally measured binding affinities (K_d),
211 and established -6.0 XEU unit as a good threshold to differentiate variants that bind dimannose
212 from "non-binders" (Bolia et al., 2014b; Li et al., 2015; Woodrum et al., 2013). Here, we applied
213 Adaptive BP-Dock initially on wild-type CV-N and its variants, P51G-m4 and mutDB (a mutant
214 in which binding by domain B has been obliterated) and the results recapitulate the success of
215 previous studies (Table S3). This result shows that Adaptive BP-Dock can correctly assess the
216 dimannose binding of CV-N and its variants, thus, we applied it on new P51G-m4 CV-N variants
217 to predict the impact of mutations on dimannose binding. Figure 2 shows the distribution of

218 changes in predicted binding energy scores relative to the P51G-m4 energy scores for mutations
219 belonging to each binary category: a positive change in binding score represents an unfavorable
220 effect on binding, and, conversely, a negative change in the score indicates an enhancement in
221 binding.
222

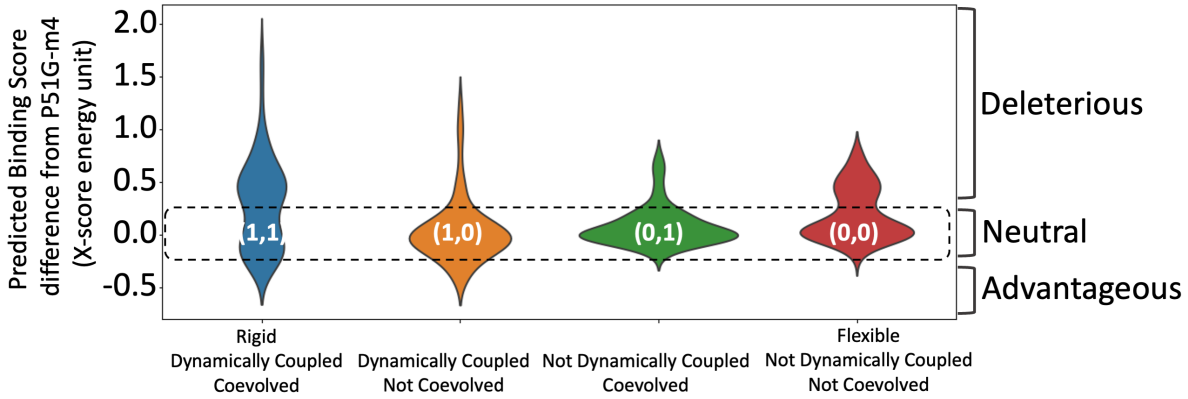

Fig 2 Predicted binding energies for each ICDC category. Mutations in category **(1,1)** positions comprise the highest number of binding energy enhancing mutations as well as deleterious mutations. Mutations in category **(0,0)** positions are mostly near neutral (Category **(1,1)** & **(0,0)** P value < 0.3).

223

224 The substitutions on positions in category **(1,1)** (Fig 2) yield a wide range of change in
225 binding energy scores: the tail of the distribution on the positive side reaches nearly a binding
226 score change of 2.0 XEUs and on the negative site values below -0.5 XEUs. Strikingly, the
227 positions in category **(1,1)** yield the most binding enhancing energy scores compared to all other
228 categories, mirroring TEM-1 results. Additionally, the substitutions applied in category **(1,0)** also
229 result in more favorable binding energy scores for dimannose. Mutations in both category **(1,1)**
230 and **(1,0)** present favorable binding energy scores. However, the number of mutations predicted to
231 be enhancing binding in category **(1,1)** is more than those in category **(1,0)** (26% of category **(1,1)**
232 compared to 14% of category **(1,0)**). Interestingly, the mutations in category **(1,0)** that disrupt the

233 binding energy scores is not a strong as category **(1,1)**, but similar to category **(0,1)** and (**0,0)**. The
234 observed mostly neutral behavior with category **(0,0)** agrees with the same trend obtained with
235 TEM-1 analyses.
236
237 Overall, the distribution of computational binding scores of dimannose binding to CV-N
238 in each category aligns with the distribution of experimentally characterized TEM-1 fitness results
239 of the same category. However, there are some discrepancies, for example, there are beneficial
240 mutations in category **(0,1)** in TEM-1, but we don't observe the same trend in CV-N. This is due
241 to the initial challenge faced in constructing the MSA of CV-N homologous proteins. There is
242 limited sequence information, and most of the proteins in the CV-N family exhibits binding
243 specificity to a different glycan (Fujimoto and Green, 2012; Koharudin et al., 2009). In contrast,
244 β-lactamase family proteins exhibit highest activity towards penicillin, and they have been
245 subjected to strong natural selection leading to conservation in both fold and function (Salverda et
246 al., 2010; Zou et al., 2021). Hence, the less noise in evolutionary analysis in case of β-lactamase
247 family of proteins allows us to correctly filter deleterious type of substitutions based on the MSA.
248 Regardless, however, in both cases, as hypothesized, substitutions on category **(1,1)** residues
249 impact the function most.
250
251 To further investigate the mechanism of functional modulation of category **(1,1)** mutations,
252 we chose the position with highest binding enhancing docking scores, I34, from category **(1,1)**.
253 I34 exhibits %DFI values lower than 0.2 (Fig 3A), is at least 16Å away from binding residues
254 (distal), dynamically coupled (Fig 3B) and co-evolved with the binding pocket (Table S2)
255 (Supplementary file 1 table 2). Moreover, docking scores of I34 variants suggest that the mutations

256 can modulate binding in a wide range: I34Y variant leads to an increase in binding affinity
257 (beneficial), I34K decreases the binding affinity (deleterious), and I34L yields no change (neutral)
258 (Table 1).
259

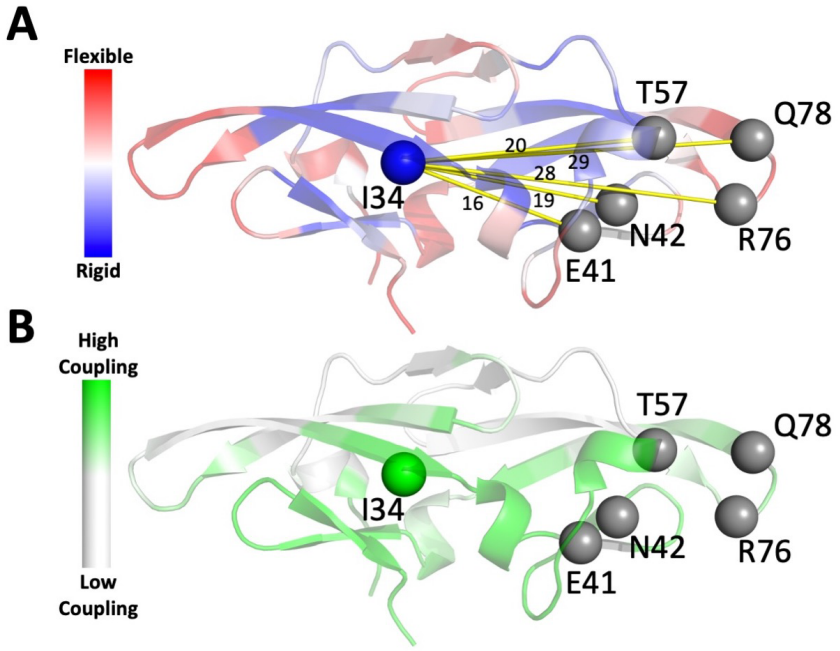

Fig 3 A) DFI profile mapped onto CV-N structure: red corresponds to high DFI, (very flexibile sites) , and blue to low DFI values (hinge sites). Position I34 (low DFI score) is highlighted B) DCI profile projected on CV-N structure with green corresponding to sites exhibiting high coupling with binding site residues.

260

261 To verify the predictions of I34 variants, we first assessed the folding and thermal stability
262 of these mutants by circular dichroism (CD) spectroscopy. Far-UV CD spectroscopy showed that
263 all mutants are well folded and adopt a fold similar to the parent protein, characterized by spectra
264 with a single negative band centered at 216 nm. We determined the stability of the mutants by CD
265 monitored thermal denaturation; the thermal denaturation curves were analyzed to obtain apparent

266 melting temperature (T_m) values. We found that the conservative mutation I34L is as stable as
267 P51G-m4, with apparent T_m of 57.8 °C and 58 °C respectively. In contrast, I34Y and I34K were
268 less thermostable than P51G-m4 as shown by apparent T_m values of 54.7 °C and 47 °C,
269 respectively. Not surprisingly, substituting a hydrophobic residue with a basic aliphatic amino acid
270 (lysine) has a large destabilizing effect, while aromatic and polar tyrosine is better tolerated. The
271 trend of thermostability is P51G-m4 ~ I34L > I34Y > I34K (Fig S1).
272

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 (μM)	ITC dimannose ΔH (kcal/mol)	ITC dimannose TΔS (kcal/mol) (T=298K)	ITC dimannose ΔG (kcal/mol)	ΔG_{H2O} (kcal/mol)	C_m (M)
P51G-m4	-6.62	117 ± 3	-12.3 ± 0.3	-7.00 ± 0.3	-5.30 ± 0.3	3.01 ± 0.047	1.46 ± 0.019
P51G-m4-I34K	-5.85	No-binding	No-binding	No-binding	No-binding	2.40 ± 0.124	0.68 ± 0.015
P51G-m4-I34L	-6.19	148 ± 2	-9.60 ± 0.1	-4.40 ± 0.1	-5.20 ± 0.1	2.95 ± 0.077	1.39 ± 0.009
P51G-m4-I34Y	-6.75	64 ± 5	-4.35 ± 0.1	1.32 ± 0.2	-5.67 ± 0.2	2.91 ± 0.157	1.13 ± 0.017

273

274 Chemical denaturation experiments were used to extract thermodynamic values, after
275 ensuring complete equilibration at each concentration of guanidinium hydrochloride by incubating
276 the samples for 72-hour (Patsalo et al., 2011). The ΔG_{H20} values and C_m values of P51G-m4, I34L,
277 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
278 respectively (Table 2). The results align with the thermal denaturation results: P51G-m4 is the
279 most stable to denaturant, followed by I34L, I34Y and I34K (Fig S2).
280

281 Next, we evaluated the impact of the mutations on the dimannose binding affinity by
282 isothermal titration calorimetry (ITC) (Fig S3); data were analyzed to extract K_d values listed in
283 Table 2. We found that I34Y binds dimannose with tightest affinity (K_d: 64 μM) of all the mutants
284 tested, a two-fold improvement over P51G-m4 (K_d: 117 μM). Binding by I34L is slightly weaker
285 with a K_d of 148 μM. No binding was observed for I34K in these conditions. Thermodynamic
286 values extracted from ITC experiments (Table 2) suggest that entropy changes play an important
287 role in the observed changes in binding affinity: surprisingly, entropy is positive for I34Y,
288 indicating an increase in disorder upon binding.
289
290 To glean more information on the mode of binding by I34Y, we determined the X-ray
291 structure of the unbound and dimannose-bound form and compared it with the template protein
292 P51G-m4. The fold is highly conserved (Fig 4) as shown by main chain RMSD of 0.16 and 0.20
293 Å with bound and unbound I34Y, respectively, and tyrosine is well tolerated at position I34. The
294 binding pocket region is also structurally conserved compared to P51G-m4. Analysis of the polar
295 contacts between dimannose and P51G-m4 and I34Y (Fig 4B) shows an identical number of
296 hydrogen bonds (11) with the ligand, indicating a conserved binding pose. We compared the
297 docked pose of I34Y acquired from Adaptive BP-Dock with the bound X-ray structure. The ligand
298 shows an RMSD value of 0.75 Å (Fig S4). These observations suggest that the increase in binding
299 affinity of I34Y towards dimannose might be mediated by equilibrium dynamics, which are not
300 captured by the crystal structure. This hypothesis is supported by the changes in entropy
301 compensation measured experimentally (ITC) in dimannose binding by P51G-m4 (negative TΔS)
302 and I34Y (positive TΔS).
303

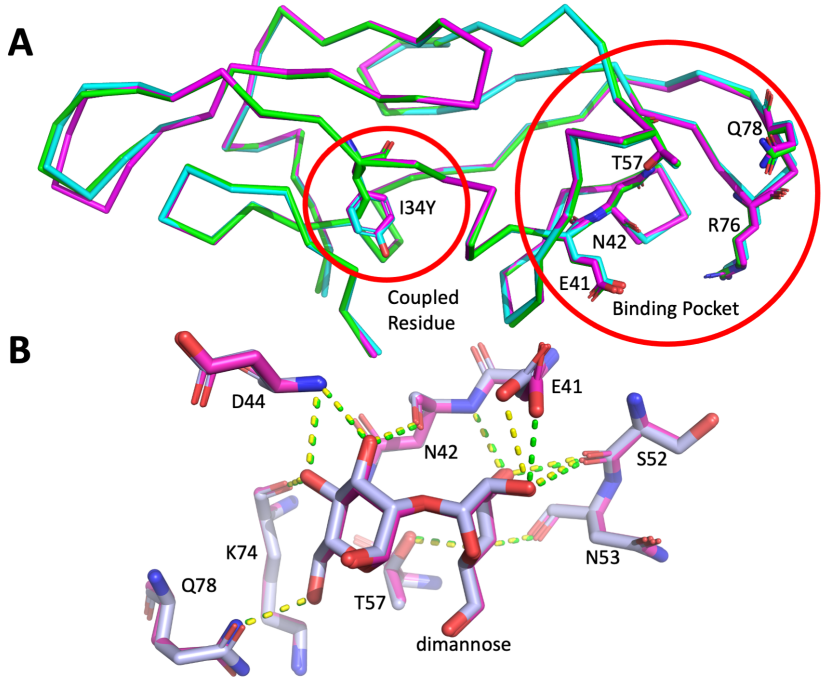

Fig 4 A) The crystal structures of I34Y (bound in magenta and unbound in cyan) and its template protein P51G-m4 (green) are superimposed. B) Overlay of bound structures of I34Y (magenta) and P51G-m4 (grey) (RMSD 0.15 Å); dashed lines depict polar interactions with dimannose.

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, 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 both in bound and unbound forms (See Methods for details of the simulations). The unbound trajectories were

315 analyzed for acquiring binding pocket hydrogen bond distances and pocket volume. Later, to learn
316 about the ligand induced conformational dynamic changes, the bound trajectories were utilized to
317 estimate computational binding free energies (Deng and Roux, 2009; Okazaki et al., 2006).
318
319 Previous computational work in our lab had linked binding affinity in the CV-N family to
320 the accessibility of the binding pocket: A hydrogen bond between the amide hydrogen of N42 and
321 carbonyl oxygen of N53 forms a closed pocket, hindering glycan accessibility, whereas the loss of
322 this hydrogen bond leads to an open pocket (Li et al., 2015). Using the formation of this hydrogen
323 bond in the trajectories of unbound WT and I34Y as metric for assessing open and closed
324 conformations, we found that I34Y variant samples the open binding pocket more often than
325 P51G-m4 (Fig S5).
326
327 Another compelling evidence differentiating I34 variants from P51G-m4 is the change in
328 their binding pocket volumes estimated by POVME pocket volume calculation tool (Wagner et
329 al., 2017). The calculated pocket volumes for I34Y, I34K and P51G-m4 were converted into
330 frequencies to obtain probability distributions (Fig 5A), revealing that I34Y variant samples a more
331 compact pocket volume compared to P51G-m4. If the pocket is too small or too large, dimannose
332 cannot maximize its interaction with the protein, and a compact conformation enables dimannose
333 to easily make the necessary hydrogen bond interactions with the protein. This optimum pocket
334 volume sampled by I34Y may also explain the different binding energetics observed by ITC, in
335 which a positive entropy change upon binding compensates for the loss in enthalpy compared to
336 P51G-m4 (Table 1) (Breiten et al., 2013; Cornish-Bowden, 2002). Pocket volume analysis reveals
337 a larger value for I34K compared to P51G-m4, suggesting that this mutant cannot accommodate

338 the necessary interactions with the dimannose resulting in loss of binding. We applied the same
339 pocket volume calculation to the X-ray structures of P51G-m4 and I34Y variant, and we found
340 volumes of 141 Å^3 and 114 Å^3 for P51G-m4 and I34Y, respectively in the unbound forms (Fig
341 5B). These volumes correlate well with the mean volumes from MD trajectories, suggesting that
342 the variants modulate the conformational dynamics of binding pocket.
343

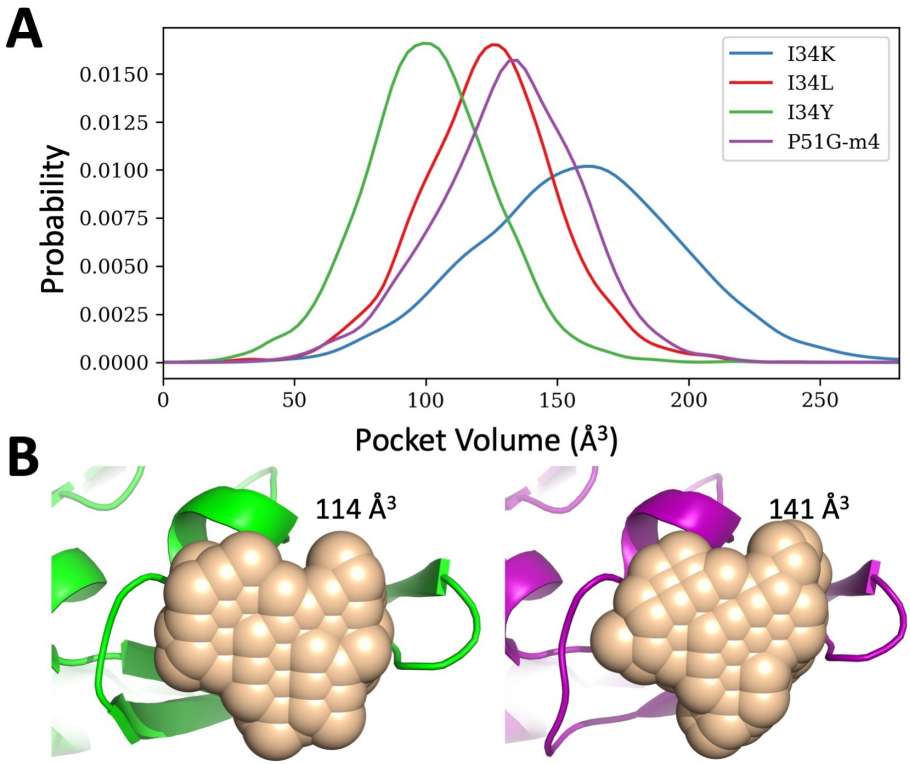

Fig 5 A) Probability distribution of the pocket volume analyses obtained from MD simulation trajectories. I34Y populates a conformation with an optimum volume more than others. P51G-m4 and I34L variant sample similar pocket volumes, but I34K variant has a larger pocket volume compared to others. B) Pocket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

344

345 Overall, the conformational dynamics analysis of the unbound conformations indicates a
346 shift of the native ensemble towards a smaller pocket volume upon I34Y mutation. This could
347 explain the decrease in the entropic cost of binding observed in ITC results. We also analyzed the
348 binding energetics by carrying out dimannose docking with 2000 different conformations sampled
349 from the binding pocket volume distributions. We found that the small volume restrict accessibility
350 to the side-chain conformations of binding residue R76 in the I34Y variant, yielding different
351 hydrogen bond patterns with the dimannose (Fig S6) and suggesting a loss in enthalpic
352 contribution.
353
354 The bound simulation trajectories were subjected to the MM-PBSA approach to estimate
355 computational binding free energies and related enthalpic and entropic contributions (He et al.,
356 2020; Rastelli et al., 2010). The results are tabulated on Table S4. The computed binding free
357 energies capture the trend of experimental binding affinities (R=0.87). The I34Y variant displays
358 a more favorable binding with dimannose compared to wild type. Interestingly, both experimental
359 and computational results show I34Y compensating the enthalpic loss with entropic gain. While
360 I34L variant enthalpic loss is greater than I34Y in computational approach, the overall binding
361 free energy mirrors the ITC results. Additionally, loss of binding of I34K variant overlaps with the
362 ITC data.
363
364
365
366

367 **Substitutions of I34 modulates the conformational ensemble leading to change in dimannose**
368 **binding affinity**
369
370 Proteins adapt to a new environment or changes the function of a protein by modulating
371 the native state ensemble through mutations of different positions while keeping the 3-D structure
372 conserved (Campitelli et al., 2020a; Kuriyan and Eisenberg, 2007; Li et al., 2015; Liu and
373 Nussinov, 2017; Modi and Ozkan, 2018; Risso et al., 2018; Tripathi et al., 2015; Woodrum et al.,
374 2013). As, we also observed a similar pattern of conservation of structure yet change in function
375 in our designed CV-N I34 variants, we further analyzed the flexibility profiles of I34 variants. The
376 DFI profiles clustered using principal component analyses (PCA) match the 2-D map of melting
377 temperature and K_d as reaction coordinates, suggesting a correlation between changes in dynamics
378 and changes in function (Fig 6). The 2-D map shows I34L, P51G-m4, and I34Y under the same
379 cluster, with I34L and P51G-m4 close, while I34K is markedly different (Fig 6A). The dendrogram
380 constructed based on the DFI profiles captures this clustering (Fig 6B) with P51G-m4 and I34L
381 variant under the same branch, suggesting their dynamics are very similar; I34Y is under the same
382 main cluster albeit in a different branch. I34K is under a separate branch, indicating different
383 dynamics. This is in agreement with our previous studies, where substitutions on DARC spots
384 modulates binding dynamics reflected in their flexibility profiles to adapt to a new environment
385 (Campitelli et al., 2021, 2020b; Kumar et al., 2015b; Modi et al., 2021a).
386

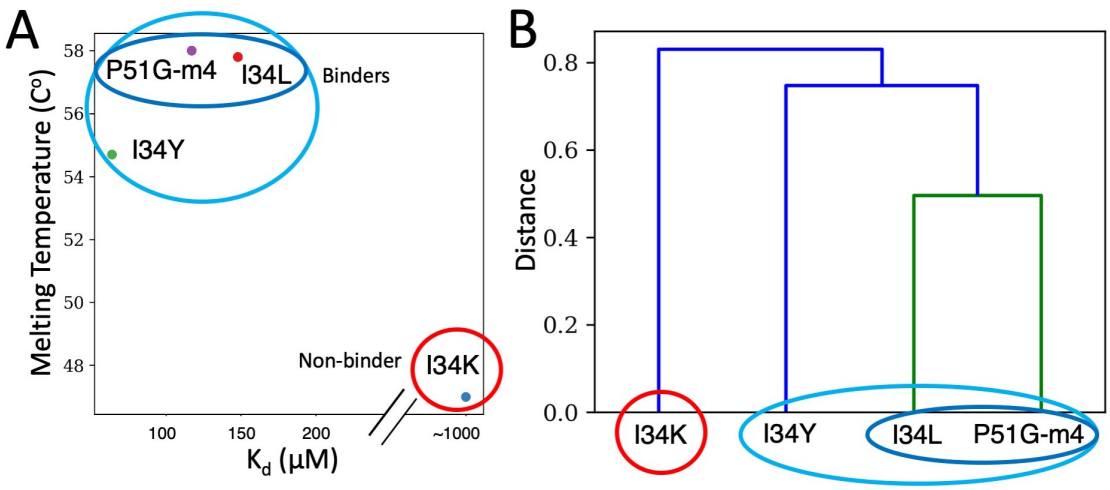

Fig 6 A) 2D map of K_d and Melting temperature of P51G-m4 and its variants B) PCA clustering on the first two principal components of the DFI profiles as a dendrogram.

387

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

396

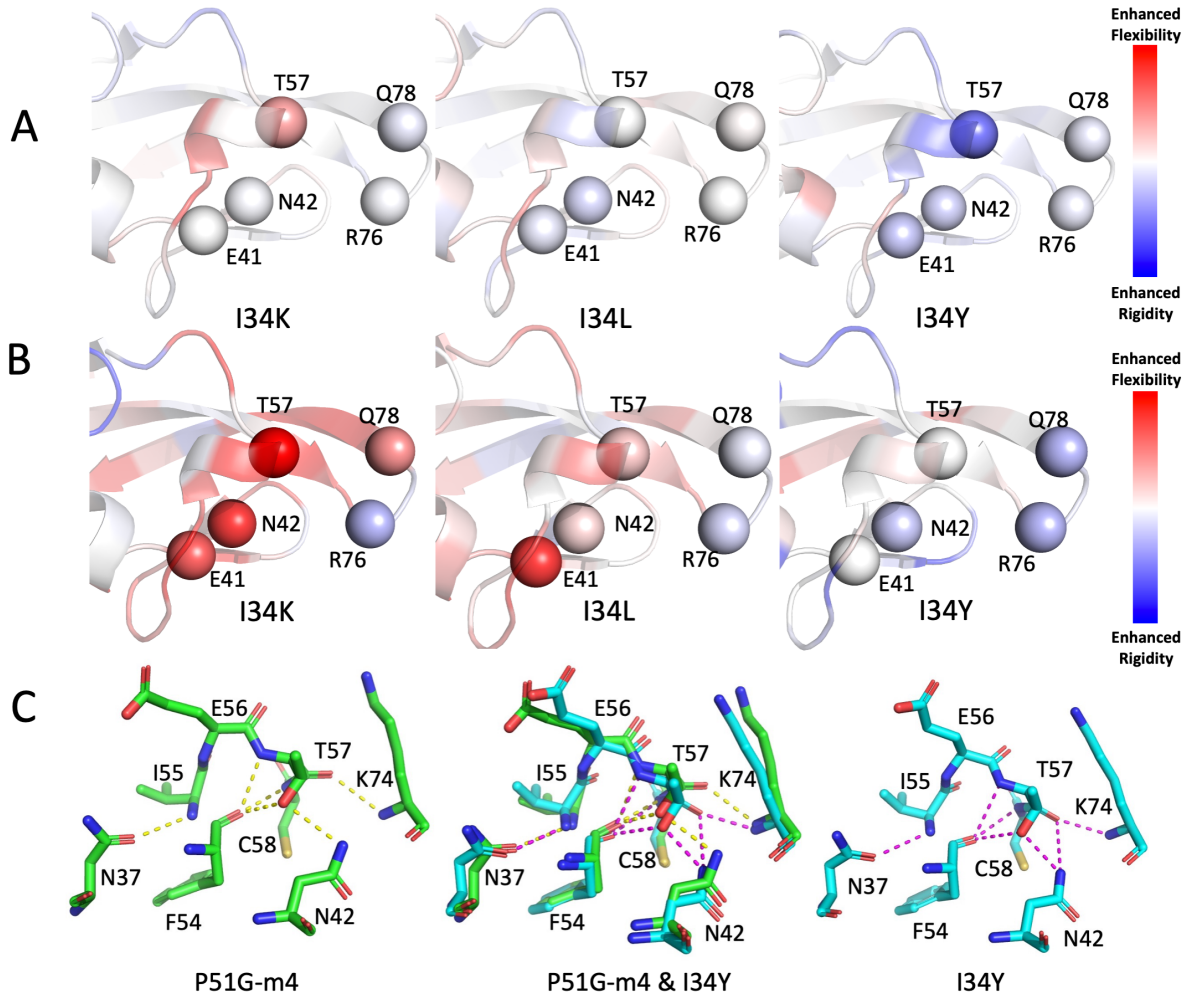

Fig 7 A) Change in flexibility of I34K, I34L, and I34Y relative to P51G-m4 in unbound form are shown. Residues E41, N42, and T57 rigidifies on I34Y compared to P51G-m4. B) Change in flexibility of I34K, I34L, and I34Y relative to P51G-m4 in bound form are projected on structure. C) Hydrogen bonding interactions of residues I55, E56, T57, and C58 are shown for P51G-m4 and I34Y variant.

397

398 Comparison of the flexibility profiles of the bound form with those of the unbound form
399 reveals that residue I34 in WT drastically gets rigidified upon binding, whereas I34Y variant does
400 not. The decreased flexibility of T57 in the unbound form of I34Y accommodate the interactions

401 with dimannose, contributing to the entropic compensation. In addition to the binding site residues
402 of Domain B, the flexibility of the rest of the residues also contribute to the total change in binding
403 free energies. Therefore, we analyzed the correlation between (i) the sum of total change in
404 flexibility of the binding site residues, (ii) the binding site residues and the residues exhibiting
405 highly coupling with the binding pocket, with the experimentally measured binding affinity
406 change. We observe a strong correlation between change in flexibility and change in affinity, as
407 expected I34Y exhibiting tighter binding also gets more rigidified upon binding compared to
408 P51G-m4. Moreover, inclusion of the highly coupled residues in addition to the domain B binding
409 sites in computing the total sum of DFI scores yields a higher correlation with the experimental
410 binding affinity change (Fig S7A). On the other hand, the correlation between the flexibility
411 change of the randomly selected residues and experimental binding affinities yields poor
412 correlation coefficient (Fig S7B). These results strongly support the role of dynamic allostery in
413 modulating binding affinity.
414
415
416 The rigidification of T57 in I34Y variant is compelling evidence that the distal mutation is
417 allosterically controlling the binding site dynamics. We further computed the network of
418 interactions that connects the residue position 34 to 57 and investigated whether distinct pathways
419 emerge after I34Y mutation. We analyzed the hydrogen bond networks, particularly computed the
420 possible network of hydrogen bonds creating pathways from 34 to 57 using the sampled snapshots
421 from the MD trajectories (Fig S8). This analysis presents a unique pathway from 34 to 57 by first
422 forming a new hydrogen bond between the side chain oxygen of the Tyrosine 34 and the nitrogen
423 of the Tyrosine 100 in I34Y variant. Furthermore, a second pathway is also found which is sampled

424 much more frequently in I34Y variant strengthening the communication between position 34 and
425 57. Thus, both pathways may contribute to the rigidification of T57. We also analyzed the
426 conformations from MD clustered with highest percentage based on alpha carbon RMSD for I34Y
427 and P51G-m4 and compared the hydrogen bond interactions of T57 and its neighboring residues.
428 The closest neighbors of T57; positions I55, E56, and C58 conserved their hydrogen bond
429 interactions with their surrounding residues between P51G-m4 and I34Y. On the other hand, T57
430 makes an additional hydrogen bond interaction in I34Y compared to P51G-m4 (Fig 7C),
431 suggesting that enhancement in hydrogen bond networking of T57 in I34Y leads to rigidification
432 of this position in equilibrium dynamics.
433
434 To gain more insight on distal dynamic modulation of binding pocket particularly the
435 decrease of binding site flexibility through distal coupling, we computationally and experimentally
436 characterized another residue, A71, belonging to category **(1,1)** and its mutations: T, S. The
437 docking scores and DFI profiles of A71 variants show high similarity to position I34 ones. The
438 variant A71T is predicted as binding enhancing by our docking scheme displaying a similar
439 binding score as I34Y (A71T predicted binding score: -6.81 XEU), whereas variant A71S is
440 predicted as analogous to I34L variant (A71S predicted binding score: -6.20 XEU). This position
441 is next to residue E72, which is within the hydrogen bond pathway (Pathway 2) (Fig S8) identified
442 previously connecting I34 and binding residue T57. Furthermore, the computed binding free
443 energies by MM-PBSA is found to be correlating with position I34 results. The A71T variant
444 shows a binding free energy near I34Y (A71T ΔG: -13.70 kcal/mol with ΔH: -29.33 kcal/mol and
445 TΔS: -15.63 kcal/mol), and A71S close to I34L (A71S ΔG: -9.98 kcal/mol with ΔH: -29.00
446 kcal/mol and TΔS: -19.02 kcal/mol). All computational analyses suggested that A71 can modulate

447 binding affinity through distal dynamic coupling similar to I34, hence we experimentally
448 characterized these two variants.
449
450 The experimental binding affinity by ITC correlates with *in silico* predictions. When the
451 change in total DFI score upon binding is compared to change in free energy of binding from ITC
452 experiments (Fig S7), A71T (ΔG: -5.70 kcal/mol with ΔH: -6.00 kcal/mol and TΔS: -0.30) features
453 both a change in total DFI and ΔG closer to I34Y, and A71S (A71S ΔG: -5.10 kcal/mol with ΔH:
454 -9.10 kcal/mol and TΔS: -4.00) shows a score identical to I34L. The entropy of A71T shows a
455 similar change as I34Y experimentally (A71T TΔS: -0.30) indicating that the same compensation
456 mechanism is utilized by another category **(1,1)** residue. A71S is closer to I34L (A71S TΔS: -
457 4.00). Similar to I34Y, the melting temperature of A71T is lower than P51G-m4 (Fig S1). Results
458 of A71 variants further establishes the potential of ICDC and category **(1,1)** residues in diversely
459 tuning the binding affinity of Domain B of CV-N through playing enthalpy-entropy compensation
460 of binding process.
461
462 Our new integrated coevolution and dynamic coupling (ICDC) approach shows that it is
463 possible to identify and incorporate distal mutations into protein design bringing together
464 evolutionary inferences with long-range dynamic communications within the 3-D network of
465 interactions.
466

467 # Methods

468

469 **Adaptive BP-Dock**

470

471 Adaptive backbone perturbation docking, Adaptive BP-Dock in short, allows us to model
472 the interaction between CV-N and glycans *in silico* (Bolia and Ozkan, 2016). Adaptive BP-dock
473 combines the complex simulation of backbone flexibility of a protein into Rosetta's ligand docking
474 application (Davis and Baker, 2009). The common restriction in docking is the implementation of
475 flexibility of receptor and ligand (Davis et al., 2009; Davis and Baker, 2009; DeLuca et al., 2015;
476 Meiler and Baker, 2006). Rosetta included the flexibility of ligand in their monte-carlo sampling
477 approach but lacking full receptor flexibility. This high order challenge is overcome by utilizing
478 Perturbation Response Scanning (PRS) to compute backbone changes during docking (Atilgan and
479 Atilgan, 2009; Bolia et al., 2014b, 2014a; Ikeguchi et al., 2005). This procedure also allows the
480 modeling of transition from an unbound state to a bound state (Bolia and Ozkan, 2016). The
481 computational cost of sampling is reduced by using a coarse-grained approach employing Elastic
482 Network Model (ENM) leading to an efficient way of computing backbone perturbations,
483 mimicking the ligand interacting with receptor (Atilgan et al., 2001, 2010; Atilgan and Atilgan,
484 2009).

485

486 We employed Adaptive BP-Dock in modeling glycan CV-N interactions starting from an
487 unbound conformation of CV-N. The perturbed pose of the protein is calculated using PRS. The
488 structure is then minimized, and the side chains are added at this step. The glycan is docked to the
489 minimized structure using RosettaLigand algorithm. Rosetta samples bound conformations using

490 a knowledge based potential function and calculates bound pose energies. The lowest energy
491 docked pose is selected and feed back to perturbation step, and the same procedure is followed
492 iteratively until a convergence is reached. At the end of each iteration the lowest energy docked
493 pose is taken and binding score is calculated using an empirical scoring function X-score. X-score
494 energy units (XEUs) has shown to provide higher correlations with experimental results (Wang et
495 al., 2002). The flow of the algorithm is shown in Fig S9. Adaptive BP-Docks iterative algorithm
496 ensures the sampling does not get trapped in a local minimum and reaches a global minimum. The
497 challenge of unbound/bound modeling is solved using the iterative approach as the conformations
498 are led towards a bound pose with the help of PRS.
499
500 **Molecular Dynamics (MD)**
501
502 Gromacs simulations are conducted for P51G-m4 CV-N and all the variants in unbound
503 form, and further for P51G-m4 CV-N, I34 variants I34K, I34L, I34Y, and A71 variants A71S,
504 A7T in bound form. (Abraham et al., 2015; Spoel et al., 2005). For each simulation the all-atom
505 system is parametrized with CHARMM36 force field and explicit water model TIP3P. The
506 solvation box is set to be minimum 16Å from the edge of the protein. The system is neutralized by
507 potassium ions to sustain electroneutrality and minimized with steepest descent for 10000 steps.
508 A short-restrained equilibrium is conducted in the constant number of particles, pressure, and
509 temperature ensemble (NPT) for 5 ns using the Berendsen method at 300K temperature and 1 bar
510 pressure. NPT production trajectories were performed with Nose-Hoover and Parrinello-Rahman
511 temperature and pressure coupling methods for 2μs at 300K and 1 bar. For all cases periodic

boundary conditions and particle-mesh Ewald (PME) with interaction cutoff of 12Å is employed with Gromacs version 2018.1.

Dynamic Flexibility Index (DFI)

DFI is a position specific metric that can measures 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.

$$[\Delta \mathbf{R}]_{3Nx1} = [\mathbf{H}]^{-1}_{3Nx3N}[\mathbf{F}]_{3Nx1}$$

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

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

533 Residues with low DFI score (below 0.2) are considered as hinge points. These points are
534 communication hubs in this 3-D interaction network. Due to high coordination number, the
535 residues exhibiting low DFI values are crucial as information gateways. While they do not exhibit
536 high residue fluctuation to the perturbations, they quickly transfer the perturbation information to
537 other parts, thus they are in control of collective motion of the protein. A change in low DFI
538 positions (i.e., a mutation) will lead to a transformation in the communication grid and majority of
539 disease-associated (i.e. function altering mutations) are often observed as hinges (Butler et al.,
540 2015; Gerek et al., 2013; Kumar et al., 2015a). The substitution on these site usually alters catalytic
541 activity or binding interaction (i.e., glycans) by modulating equilibrium dynamics (Campitelli et
542 al., 2020a).
543
544 **Dynamic Coupling Index (DCI)**
545
546 Dynamic Coupling Index (DCI) exploits the same framework of DFI (Campitelli et al.,
547 2020a; Larrimore et al., 2017). DCI utilizes the residue response fluctuation upon random force
548 perturbation at a specific residue position to investigate residues that exhibit long-range coupling
549 to each other. In DCI approach, a unit force is applied on functional residues (i.e., binding site
550 residues) one by one and responses of all other residues are calculated.
551

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

552
553

554 With DCI scheme the residues with high response (high DCI score) indicates high long
555 range dynamic coupling. Residues with high DCI values with binding sites play a critical role in
556 intercommunication of a protein with the binding residues. These coupled residues are of utmost
557 importance in how forces propagate through amino acid chain network on a binding event. Some
558 of the coupled residues are far from the binding site but still encompass modulation capabilities
559 over binding pocket.
560
561 **Informing dynamics from co-evolution**
562
563 Co-evolutionary data paves the way to assessing 3-D structural contacts by utilizing
564 available sequence information (Hopf et al., 2018; Marks et al., 2012; Morcos et al., 2014).
565 Sequence information is more abundant compared to resolved protein structures. Exploiting the
566 sequence information, primary contacts comparable to realistic structural contacts can be
567 calculated and a contact matrix is formed. The accuracy of these contact maps is proved to be
568 valuable in protein folding studies (Kryshtafovych et al., 2019; Morcos et al., 2011; Wang et al.,
569 2016). Evolutionary coupling (EC) analysis is used to collect information on how much two
570 residues in a protein sequence is in close proximity in 3-D structure. EC scores could be calculated
571 by many different statistical approaches. In this study EC information is gathered by using
572 RaptorX, EVcouplings, and MISTIC webservers (Hopf et al., 2019; Simonetti et al., 2013; Wang
573 et al., 2017). While the limitation of these methods emerges from sequence homolog availability
574 of a protein in multiple sequence alignment (MSA), RaptorX uses a deep neural network
575 leveraging joint family approach, combining multiple ortholog protein families sharing similar
576 function and phylogeny, to infer possible contacts. This method is proven to produce high accuracy

577 in contact prediction compared to others (Wang et al., 2017). However, for a given MSA
578 containing enough homolog sequences other methods are also strong in predicting spatial contacts.
579 EVcouplings approach uses Direct Information (DI) to calculate co-evolutionary couplings. DI
580 metric is a modified mutual information (MI) score considering consistency between pairwise
581 probabilities and single amino acid frequencies (de Juan et al., 2013; Morcos et al., 2011).
582 Nonetheless, MI, a global approach compared to local DI metric, is accurate in capturing true
583 contacts, while entangling indirect contacts from direct contacts. MISTIC web server has taken
584 advantage of MI to calculate co-evolutionary couplings (Dunn et al., 2008; Gouveia-Oliveira and
585 Pedersen, 2007; Simonetti et al., 2013). In their MI method they introduced a correction term to
586 MI to surpass the low statistics gathered with an MSA containing limited number of sequences.
587 This approach is very useful in cases where certain homologs are rare and MSA of these homologs
588 have multiple gaps in their alignments. All of these methods are employed in this study to achieve
589 high accuracy predictions in finding residue couplings.
590
591 **Mutant proteins cloning, expression, and purification**
592
593 The genes for mutants (I34Y, I34K, and I34L) were generated by applying mutagenic
594 primers to P51G-m4-gene sequence and amplifying by PCR. The constructs were subsequently
595 cloned in pET26B vector between NdeI and XhoI sites and transformed in BL21(DE3) for
596 expression and purification. The proteins were expressed from a 10 ml starter culture in LB broth
597 overnight at 37 °C, inoculated into 1 L LB medium. The culture was induced with 1 mM isopropyl
598 thiogalactoside (IPTG) when OD reached 0.6 and grown for another 6-8 hours. Then, the cells
599 were harvested by centrifugation, lysed in 6M Guanidine hydrochloride at pH 8.0, and sonicated

600 for 10 minutes. The supernatant recovered after centrifugation was used to purify proteins with
601 GE HisTrap HP column (GE Healthcare Bio-Sciences, Piscataway, NJ) and a Bio-Rad
602 EconoPump (Bio-Rad, Richmond, CA) under denaturing conditions. In brief, the proteins were
603 loaded on the column in Gu-HCl buffer, which was buffer exchanged by 8M Urea buffer. The
604 nonspecific proteins were washed out by 4M urea and 20mM imidazole buffer, pH 8.0 and eluted
605 with 2M Urea and 200mM imidazole, pH 8.0 buffer before putting it for overnight dialysis against
606 10 mM Tris pH 8.0 and 100 mM NaCl buffer. The buffer was changed once during the night. The
607 refolded protein was concentrated and re-purified to isolate the monomeric species by size
608 exclusion chromatography using Sephadex 75 10/300 column on Agilent's Infinity 1260 system.
609 The gel filtered protein was finally used for all the experiments.
610
611 **CD spectroscopy and T-melts**
612
613 In CV-N family proteins, thermodynamic parameters like free energy of unfolding,
614 enthalpy and entropy cannot be extracted by thermal denaturation because the transition from
615 folded to unfolded state is non-reversible (Patsalo et al., 2011), therefore melting temperatures are
616 used. Far-UV CD spectra were recorded on a Jasco J-815 spectropolarimeter equipped with a
617 thermostatic cell holder, PTC 424S. Spectra were measured from 250 to 200 nm, using a scanning
618 speed of 50 nm/min and a data pitch of 1.0 nm at 25 °C. Samples concentration was approximately
619 15 μM in 10 mM Tris, pH 8.0 and 100 mM NaCl. For thermal denaturation experiments, the
620 melting profile was monitored at 202 nm from 25 °C to 90 °C. The data points were plotted and
621 fitted in Origin8.5 software to get apparent Tm.
622

623 **Isothermal titration calorimetry (ITC)**

624

625 ITC was performed at the Sanford-Burnham Medical Research Institute Protein Analysis
626 Facility using ITC200 calorimeter from Microcal (Northampton, MA) at 23 ºC. 2.0 µl aliquots of
627 solution containing between 3mM and 10 mM Man2 were injected into the cell containing between
628 0.057, and 0.11 mM protein. 19 of 2.0 µl injections were made. The experiments were performed
629 in 10mM Tris, 100mM NaCl, pH 8.0 buffer. ITC data were analyzed using Origin software
630 provided by Microcal.

631

632 **Chemical denaturation experiments**

633

634 Chemical denaturation experiments were done by monitoring the shift in the intrinsic
635 tryptophan fluorescence on Cary Eclipse instrument (Varian). 10 µM of protein samples were
636 incubated with increasing concentrations of Guanidine hydrochloride in the range of 0-6 M in 50
637 mM Tris pH8.0 buffer for 72 hours at 25 °C. The emission spectra for the same were recorded by
638 keeping the excitation wavelength at 295nm and bandwidth of 1 nm. A ratio of fluorescence at
639 330 and 360 nm ($I_{330/360}$) was plotted at respective Gu-HCl concentrations, and the data points
640 were fit to following sigmoidal equation to obtain C_m.

641 $$y = \mathrm{A2} + \frac{\mathrm{A1} - \mathrm{A2}}{1 + \mathrm{e}^{(\mathrm{x}-\mathrm{x0})/\mathrm{dx}}}$$

642 Where, A1and A2 are the initial and final 330/360 ratios and x0 is the concentration of Gu-HCl,
643 where y= (A1 +A2) /2, or the point where 50 % of the population is unfolded. It is also denoted as
644 C_m.

645 The denaturation curve was used to calculate the free energy of the protein in the absence of
646 denaturant (ΔG_{H2O}). Fraction unfolded (f_U) was calculated using the following formula:

647 $$f_U = (y_F - y_{obs})/(y_F - y_U)$$

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

651 $$K = f_U / f_F$$

652 $$K = f_U / 1 - f_F$$

653 $$\Delta G = -RT\ln K$$

654 Where R is the gas constant whose value is 1.987 cal/mol.K and T is the temperature of
655 incubation, which was 298K. The value of ΔG is linear over a limited range of Gu-HCl. The linear
656 fit over that range was extrapolated to obtain ΔG_{H2O}.

657

658 **Crystallization and structure determination**

659

660 I34Y was purified as discussed previously and the monomeric gel filtered protein was
661 concentrated to 8mg/ml. We got the crystals in 2M ammonium sulphate and 5 % (v/v) 2-propanol
662 after screening it in Index HT screen from Hampton Research. The protein crystals were
663 reproduced using same condition in hanging drop method. For protein crystals with dimannose,
664 the crystals were incubated in 1.2 fold molar excess of dimannose. Single needle-like crystals were
665 picked up and cryo-preserved in 25 % glycerol before freezing them for data collection at
666 Synchrotron ALS, beamline 8.2.1. Single crystal diffraction was measured at wavelength of 0.999
667 A with ADSC quantum 315 r detector. The data were evaluated to resolution of 1.25 A. 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

S.B.O. 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.

Competing interests

No competing interests declared.

Supplementary Data

Table S1. DFI, DCI, RaptorX, Evcoupling, and MISTIC metrics are used to identify residues in TEM-1 β-lactamase for the four unique categories (Supplementary file 1 table 1).

	Dynamically Coupled to Binding Sites 1										Not Dynamically Coupled to Binding Sites 0									
	Residue	DFI Score	DCI Score	Distance		RAPTORX Score	MISTIC Score	EVCOUPLING Score	Current Amino acid	Available Amino acid	Residue	DFI Score	DCI Score	Distance		RAPTORX Score	MISTIC Score	EVCOUPLING Score	Current Amino acid	Available Amino acid
				Minimum	Average									Minimum	Average					
Coevolved with Binding Sites 1	44	0.20	0.94	14	19	0.92	0.87	0.32	V	AFILMV	51	0.56	0.35	17	24	0.77	0.77	0.92	L	ACEFGIKLMNPQRSTV
	45	0.10	0.89	15	19	0.93	0.99	0.67	G	AGR	121	0.55	0.63	13	17	0.66	0.91	0.91	E	ADEGHKMNQRSTV
	46	0.05	0.86	15	19	0.90	0.88	0.69	Y	AFGILMTVY	142	0.35	0.52	13	19	0.83	0.61	0.77	I	ACFHILMNQRSTVY
	47	0.13	0.85	16	20	0.89	0.84	0.97	I	ACDEFGHILMNQRSTVWY	148	0.30	0.68	11	16	0.94	0.78	0.95	L	ADFILMQTVWY
	122	0.15	0.92	10	14	0.71	0.92	0.86	L	ACFHILMSV	155	0.58	0.51	17	23	0.68	0.91	0.91	M	ACEFGHIKLMNPQRSTVW
	137	0.13	0.93	10	15	0.93	0.95	0.95	L	AEFGIKLMQRSVY	183	0.04	0.50	14	17	0.84	0.97	0.75	P	ALPS
	179	0.25	0.85	11	13	0.98	0.97	0.93	D	ADGNR	199	0.61	0.45	17	23	0.74	0.99	0.67	L	ILV
	181	0.19	0.90	12	15	0.92	0.99	0.71	T	ACILSTV	220	0.49	0.67	10	16	0.95	0.96	0.99	L	ACEFGHKLMQRSTVWY
	182	0.10	0.70	15	18	0.66	0.98	0.65	M	ACKMNRSTV	222	0.61	0.56	11	18	0.93	0.99	0.74	R	ACGKLNPQRSV
	187	0.08	0.73	13	17	0.84	0.91	0.99	A	ACEGILMNQRSTV	223	0.71	0.51	14	20	0.88	0.88	0.64	S	ADEGHKLPQRSTV
	190	0.19	0.78	12	16	0.95	0.94	0.86	L	AFHILMNSTVWY	224	0.69	0.59	16	22	0.83	0.68	0.75	A	ADEGHIKLNPRSTVY
	262	0.01	0.94	10	15	0.96	0.64	0.99	V	ACGILNSTV	225	0.74	0.45	15	22	0.84	0.76	0.99	L	AFIKLMQSTVW
	263	0.01	0.90	11	15	0.94	0.80	0.99	I	ACFGILMSTV	227	0.98	0.67	20	27	0.66	0.62	0.89	A	ADEFGHIKLMNPQRSTV
Not Coevolved with Binding Sites 0	83	0.74	0.89	16	22	0.16	0.41	0.44	R	ADEGHKLMNQRSTVY	35	0.80	0.81	21	27	0.52	0.09	0.53	D	ADEGHIKLMNPQRSTVWY
	84	0.77	0.84	17	23	0.33	0.41	0.48	V	ACDEFGHIKLMNQRSTVWY	52	0.75	0.28	21	27	0.44	0.49	0.50	N	ADEFGHIKMNPQRSTV
	93	0.92	0.92	19	24	0.16	0.35	0.51	R	ADEHKLNQRSTVY	197	0.71	0.42	21	26	0.33	0.39	0.49	E	ACDEGHKLNPQRSTY
	94	0.90	0.93	19	23	0.16	0.29	0.40	R	ACDEFGHIKLMNPQRSTVW	201	0.79	0.43	20	25	0.33	0.09	0.32	L	ACDEFGHIKLMNPQRSTVY
											289	0.85	0.72	23	30	0.52	0.10	0.50	H	ACDEFGHIKLMNPQRSTVWY

683

Table S2. DFI, DCI, RaptorX, Evcoupling, and MISTIC metrics are used to identify residues in CV-N for the ICDC categories (Supplementary file 1 table 2).

	Dynamically Coupled to Binding Sites 1										Not Dynamically Coupled to Binding Sites 0									
	Residue	DFI Score	DCI Score	Distance		RAPTORX Score	MISTIC Score	EVCOUPLING Score	Current Amino acid	Available Amino acid	Residue	DFI Score	DCI Score	Distance		RAPTORX Score	MISTIC Score	EVCOUPLING Score	Current Amino acid	Available Amino acid
				Minimum	Average									Minimum	Average					
Coevolved with Binding Sites 1	34	0.17	0.71	16	22	0.65	0.93	0.71	I	FIKLMVY	18	0.54	0.51	16	21	0.65	0.98	0.84	L	FILMV
	61	0.15	0.83	11	15	0.81	0.76	0.86	T	ACFIKLMQSTVWY	66	0.95	0.27	21	26	0.27	0.91	0.94	S	ADEFGKLMNPQRST
	71	0.01	0.93	11	13	0.94	0.92	0.98	A	ACDEGQSTV										
Not Coevolved with Binding Sites 0	32	0.02	0.91	23	29	0.44	0.92	0.48	S	ACDHLNSTV	65	0.88	0.49	21	24	0.26	0.7	0.69	G	ADEFGHIKNPQRSTWY
	99	0.35	0.85	18	26	0.53	0.3	0.37	K	ADEFHIKLMNQRSTV	67	0.82	0.29	18	23	0.34	0.79	0.86	S	ADEGHIKLMNPQRSTVY
											88	0.74	0.55	16	23	0.59	0.72	0.73	D	ADEGHNRST

684

Table S3. The predicted binding affinities of domain B and comparison with experimental ITC data for wild type, mutDB, and P51G-m4 benchmarking.

Protein	Predicted Binding Score (X-score energy unit)	ITC dimannose K_d (μM)	ITC dimannose ΔH (kcal/mol)	ITC dimannose TΔS (kcal/mol) (T=298K)	ITC dimannose ΔG (kcal/mol)
Wild Type	-7.08	16 ± 1	-12.5 ± 0.3	-6.00 ± 0.1	-6.50 ± 0.3
mutDB	-5.97	No-binding	No-binding	No-binding	No-binding
P51G-m4	-6.62	117 ± 3	-12.3 ± 0.3	-7.00 ± 0.3	-5.30 ± 0.3

685

Table S4. Binding free energies, enthalpy and entropy values for wild type CV-N and its variants calculated with MM-PBSA approach applied on dimannose bound MD simulations.

Protein	ΔH (kcal/mol)	TΔS (kcal/mol)	ΔG (kcal/mol)
P51G-m4	-31.13	-18.34	-12.80
I34K	-0.03	-51.07	51.04
I34L	-27.54	-17.97	-9.57
I34Y	-29.76	-15.72	-14.05

* The ΔG scores displayed in this table correlates with experimental binding scores with an R value of 0.87.

686

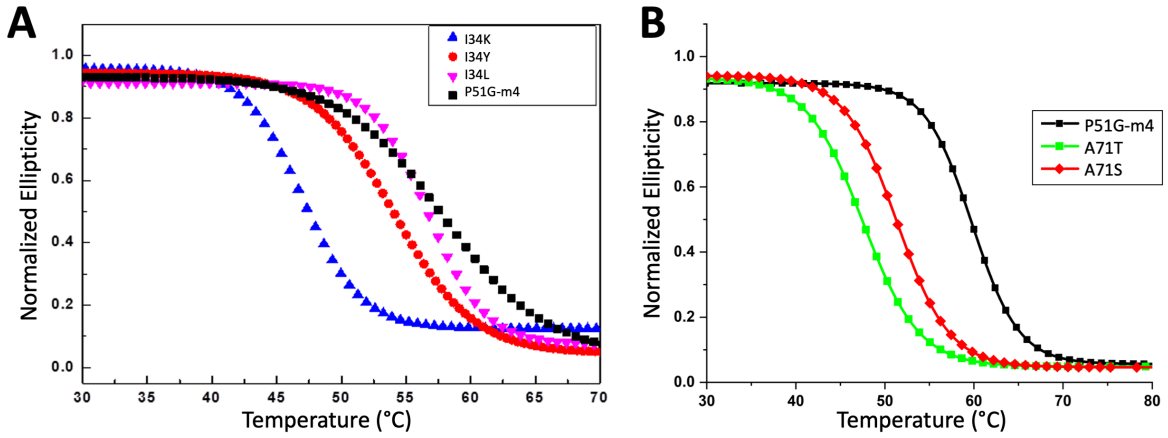

Fig S1 Fits for thermal melts of the CV-N mutants A) I34 variants, and B) A71 variants.

687

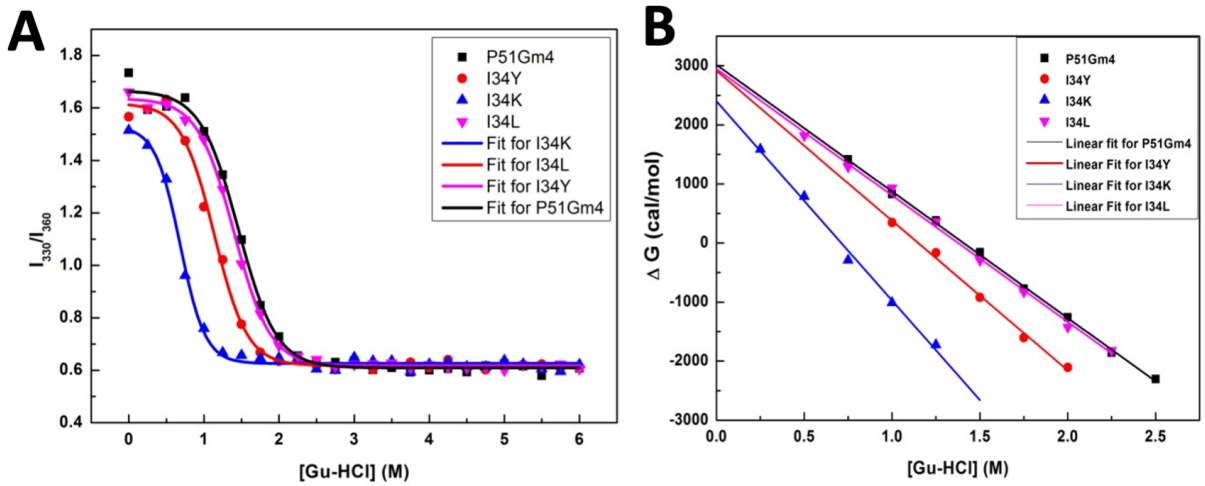

Fig S2 A) Chemical denaturation curve showing I_{330}/ I_{360} ratio as a function of Gu-HCl concentration, b) ΔG_{H2O} versus Gu-HCl concentration plot for CV-N mutants.

688

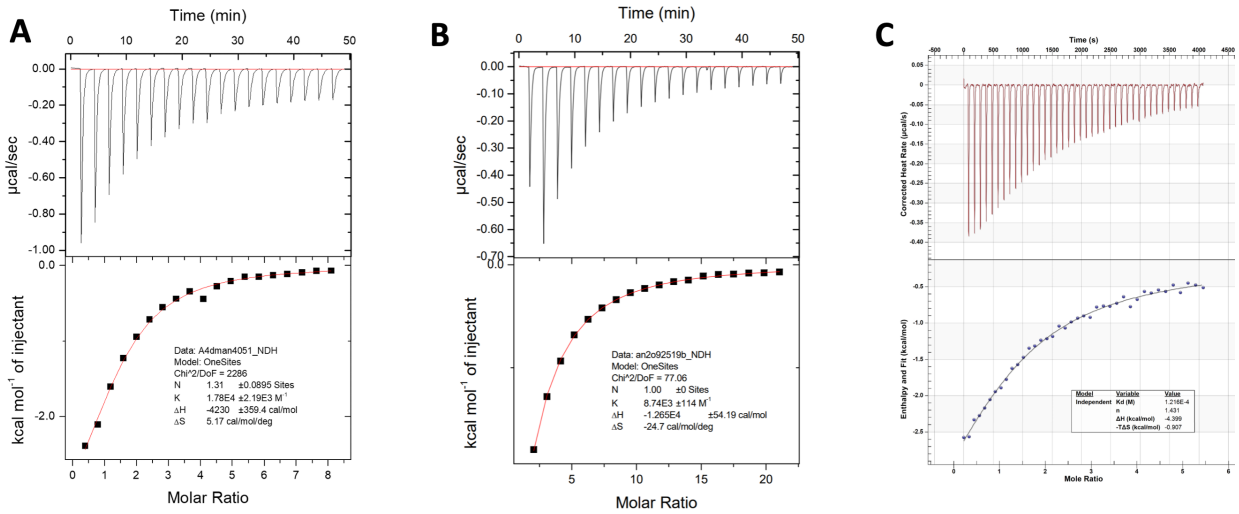

Fig S3 Binding isotherms of CV-N mutants upon titration with dimannose: A) I34Y and B) P51G-m4 C) A71T.

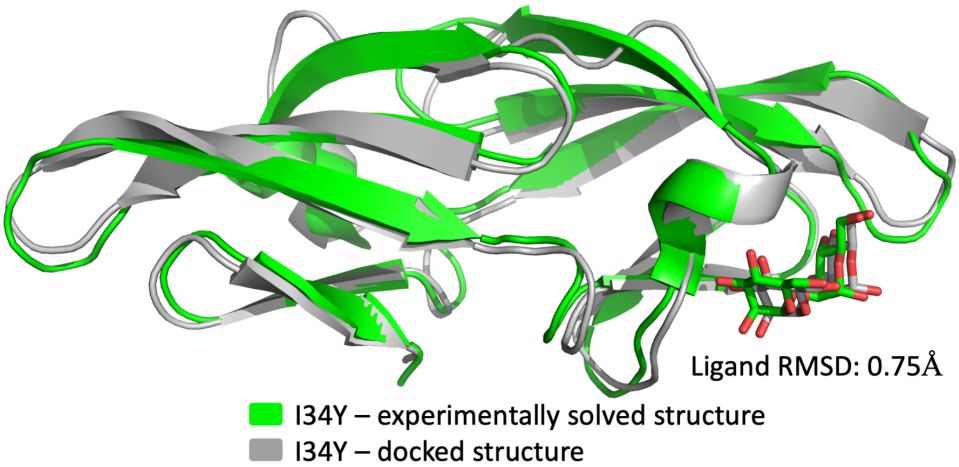

Fig S4 Comparison of experimentally solved I34Y structure with docked pose from Adaptive BP dock algorithm. The RMSD of the ligand is calculated as 0.75 Å.

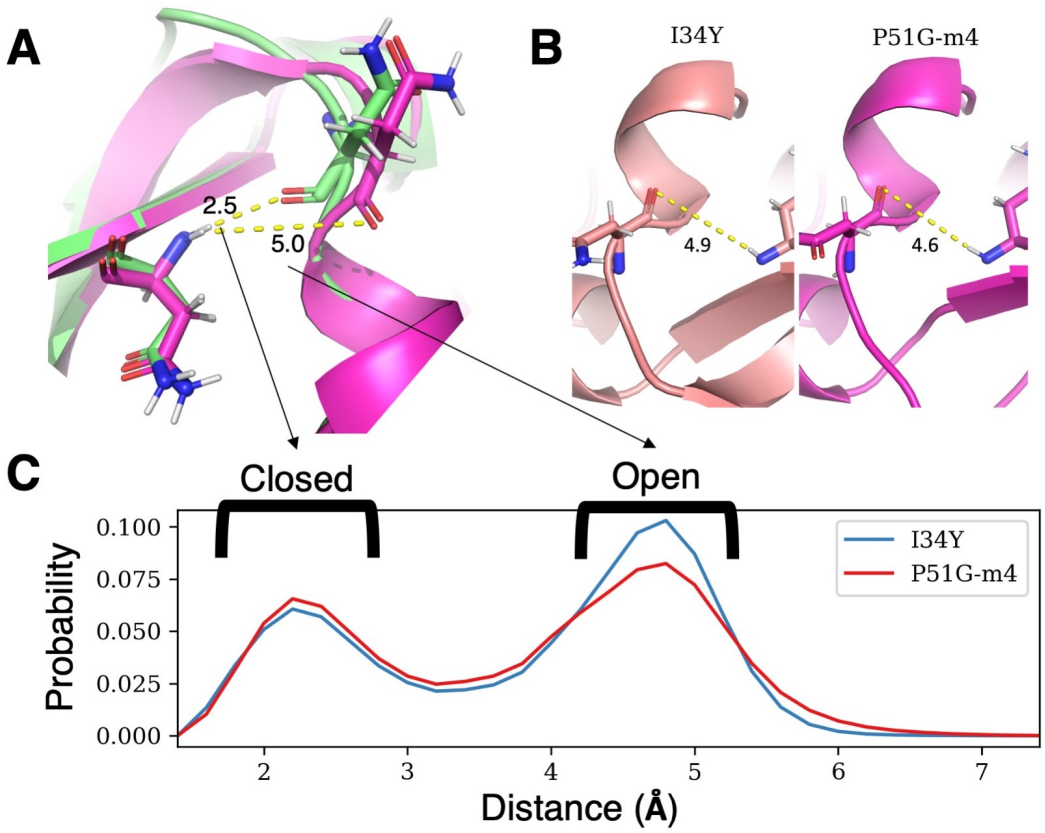

Fig S5 A) Structural difference of open vs closed conformation based on the hydrogen bond distance between residue N42 and N53 B) Hydrogen bond distance between residue N42 and N53 from crystal structures of P51G-m4, and P51G-m4-I34Y C) Frequencies of hydrogen bond distance between residue N42 and N53 from GROMACS production runs showing I34Y variant sampling more open conformation compared to P51G-m4.

691

692

693

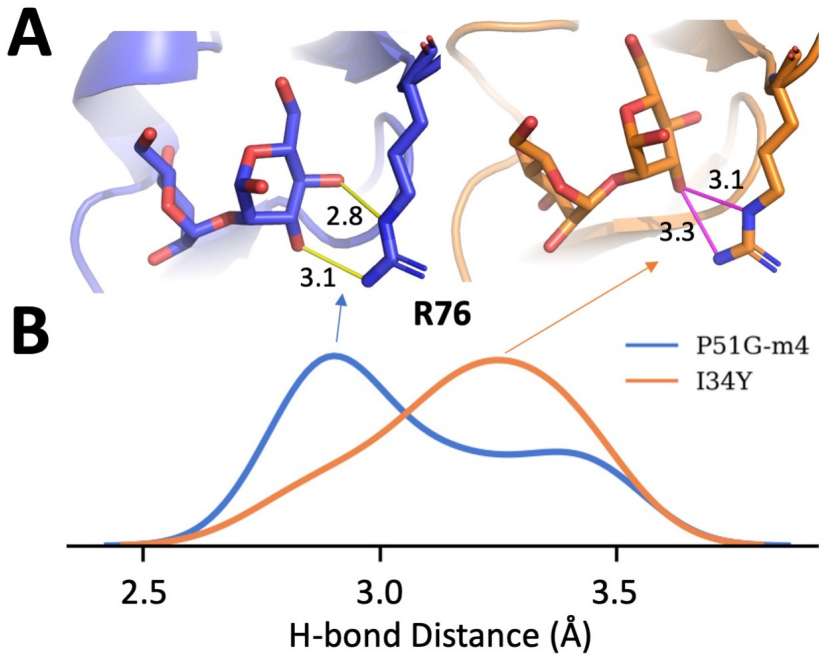

Fig S6 We sampled 2000 different conformations from MD simulations for P51G-m4 CV-N and I34Y mutant and performed dimannose docking to obtained docked poses and then analyzed hydrogen bond patterns A) Hydrogen bonds (representing the peak of the distribution on panel B) and their distances are shown between dimannose and residue R76 for P51G-m4 (blue) and I34Y (orange) B) H-bond distance distribution between dimannose and residue R76.

694

695

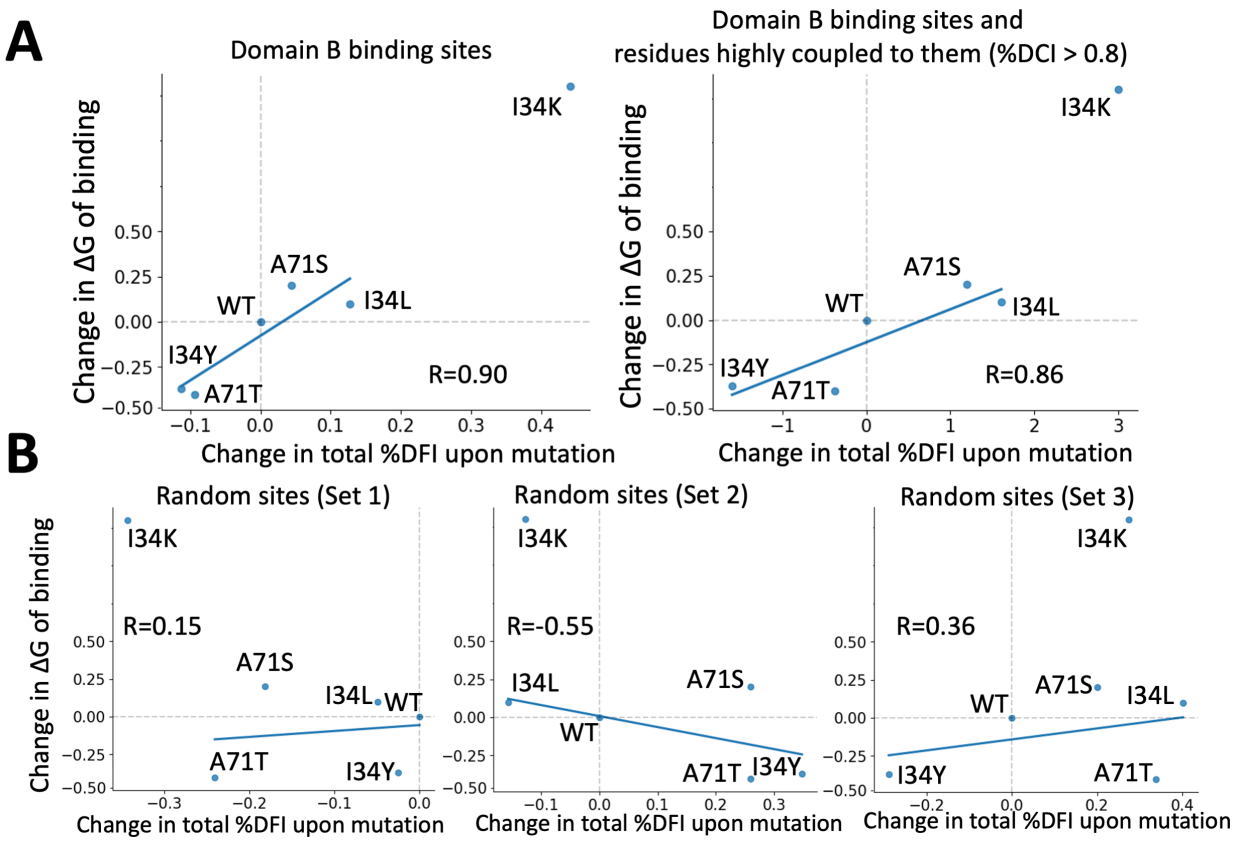

Fig S7 A) Change in ΔG of binding (ΔG_{mut} - ΔG_{wt}) is compared with change in total DFI scores ($\sum DFI_{mut} - DFI_{wt}$) for selected residues. The correlation with experimental binding scores is compared with the total sum of DFI values considering only domain B binding site residues first, and also summing over the domain B binding sites as well as the residues highly coupled (coupling greater than 0.8) to them. The observed high correlations indicates that these residues play an important role in the binding modulation upon mutations. The similar trend has also been seen in LacI (Campitelli et al., 2021) B) In addition we randomly selected residues in domain B to calculate total DFI change over these positions upon mutations. Three different randomly selected residue sets all show poor correlation with change in experimental binding free energy.

696
697

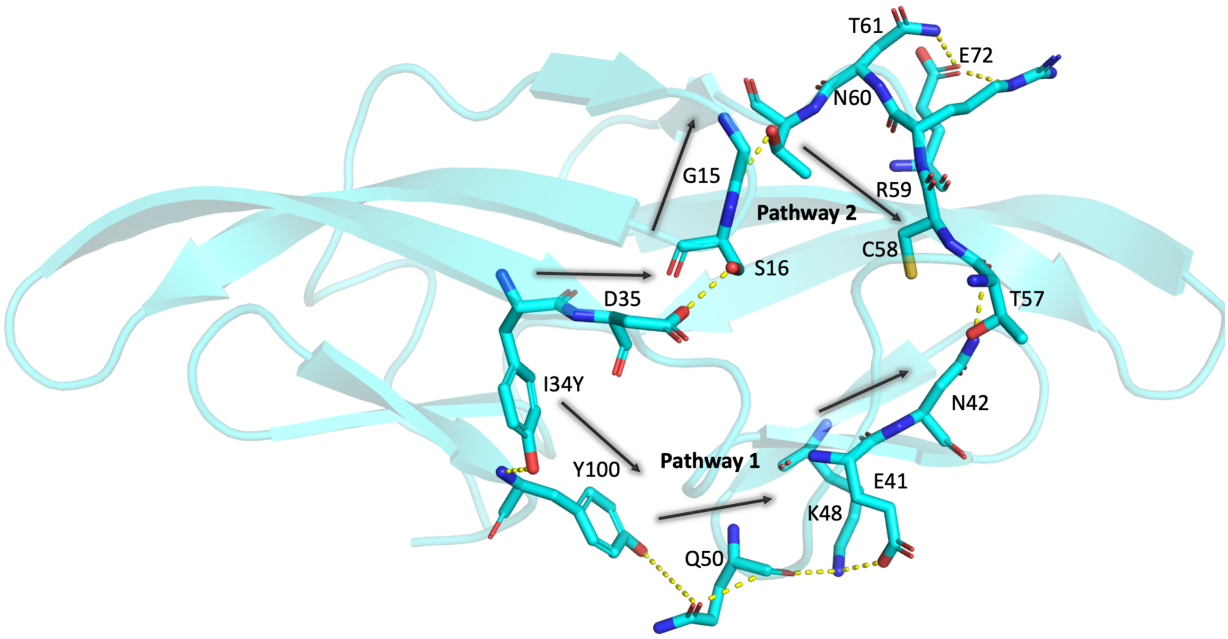

Fig S8 Network of hydrogen bond interactions connecting residue location 34 to T57 is investigated in I34Y variant and P51G-m4 CV-N. Two Hydrogen bond pathways are found connecting residue 34 to 57. Pathway 1 is unique to I34Y. Pathway 2 is also observed in P51G-m4 CV-N but sampled much more frequently in I34Y variant.

698

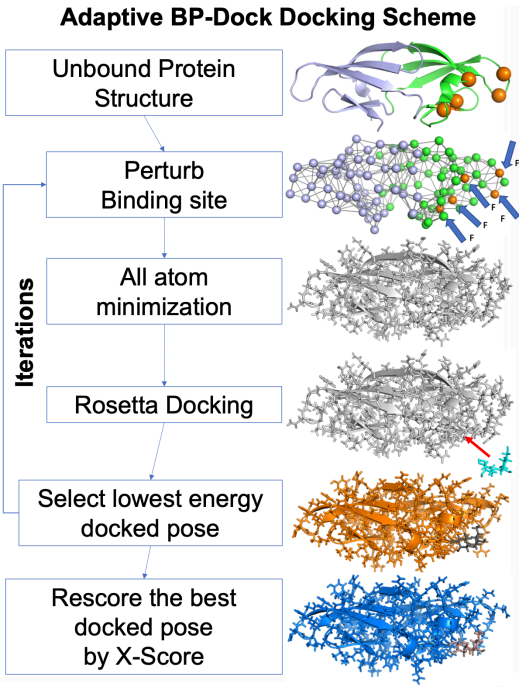

Fig S9 The flow of Adaptive BP-Dock. Starting from an unbound structure the backbone atoms are perturbed using PRS and perturbed pose is fed into minimization. The minimized conformation is docked with the glycan using RosettaLigand docking approach. The lowest scored docked pose is selected and used for next iteration step until convergence is established. For every lowest energy docked pose from an iteration a new binding score is reevaluated using X-score empirical scoring function.

Additional files

Supplementary file 1

This file contains Supplementary Table 1 & 2 (referenced in the main text).

706 Supplementary Table 1 contains the complete TEM-1 DFI, DCI, RaptorX, Evcoupling, and

707 MISTIC metric data used in this study.

708 Supplementary Table 2 contains the complete CV-N DFI, DCI, RaptorX, Evcoupling, and

709 MISTIC metric data used in this study.

710

711 **References**

712

713 Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. 2015. GROMACS:
714 High performance molecular simulations through multi-level parallelism from laptops to
715 supercomputers. *SoftwareX* **1–2**:19–25. doi:10.1016/j.softx.2015.06.001
716 Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. 2001. Anisotropy of
717 Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophysical Journal*
718 **80**:505–515. doi:10.1016/S0006-3495(01)76033-X
719 Atilgan C, Atilgan AR. 2009. Perturbation-Response Scanning Reveals Ligand Entry-Exit
720 Mechanisms of Ferric Binding Protein. *PLoS Comput Biol* **5**.
721 doi:10.1371/journal.pcbi.1000544
722 Atilgan C, Gerek ZN, Ozkan SB, Atilgan AR. 2010. Manipulation of Conformational Change in
723 Proteins by Single-Residue Perturbations. *Biophysical Journal* **99**:933–943.
724 doi:10.1016/j.bpj.2010.05.020
725 Balzarini J. 2007. Targeting the glycans of glycoproteins: a novel paradigm for antiviral therapy.
726 *Nature Reviews Microbiology* **5**:583–597. doi:10.1038/nrmicro1707
727 Barrientos LG, Matei E, Lasala F, Delgado R, Gronenborn AM. 2006. Dissecting carbohydrate–
728 Cyanovirin-N binding by structure-guided mutagenesis: functional implications for viral
729 entry inhibition. *Protein Eng Des Sel* **19**:525–535. doi:10.1093/protein/gzl040
730 Barrientos LG, O'Keefe BR, Bray M, Sanchez A, Gronenborn AM, Boyd MR. 2003.
731 Cyanovirin-N binds to the viral surface glycoprotein, GP1,2 and inhibits infectivity of
732 Ebola virus. *Antiviral Research* **58**:47–56. doi:10.1016/S0166-3542(02)00183-3
733 Bewley CA. 2001. Solution Structure of a Cyanovirin-N:Manα1-2Manα Complex: Structural
734 Basis for High-Affinity Carbohydrate-Mediated Binding to gp120. *Structure* **9**:931–940.
735 doi:10.1016/S0969-2126(01)00653-0
736 Bolia A, Gerek ZN, Ozkan SB. 2014a. BP-Dock: A Flexible Docking Scheme for Exploring
737 Protein–Ligand Interactions Based on Unbound Structures. *J Chem Inf Model* **54**:913–
738 925. doi:10.1021/ci4004927
739 Bolia A, Ozkan SB. 2016. Adaptive BP-Dock: An Induced Fit Docking Approach for Full
740 Receptor Flexibility. *J Chem Inf Model* **56**:734–746. doi:10.1021/acs.jcim.5b00587
741 Bolia A, Woodrum BW, Cereda A, Ruben MA, Wang X, Ozkan SB, Ghirlanda G. 2014b. A
742 Flexible Docking Scheme Efficiently Captures the Energetics of Glycan-Cyanovirin
743 Binding. *Biophysical Journal* **106**:1142–1151. doi:10.1016/j.bpj.2014.01.040

744 Bolmstedt AJ, O'Keefe BR, Shenoy SR, McMahon JB, Boyd MR. 2001. Cyanovirin-N Defines
745 a New Class of Antiviral Agent Targeting N-Linked, High-Mannose Glycans in an
746 Oligosaccharide-Specific Manner. *Mol Pharmacol* **59**:949–954.
747 doi:10.1124/mol.59.5.949
748 Botos I, Wlodawer A. 2005. Proteins that bind high-mannose sugars of the HIV envelope.
749 *Progress in Biophysics and Molecular Biology*, Structure-guided design of AIDs
750 Antivirals **88**:233–282. doi:10.1016/j.pbiomolbio.2004.05.001
751 Botos I, Wlodawer A. 2003. Cyanovirin-N: a sugar-binding antiviral protein with a new twist.
752 *CMLS, Cell Mol Life Sci* **60**:277–287. doi:10.1007/s000180300023
753 Bowman GR, Geissler PL. 2012. Equilibrium fluctuations of a single folded protein reveal a
754 multitude of potential cryptic allosteric sites. *Proceedings of the National Academy of*
755 *Sciences* **109**:11681–11686.
756 Breiten B, Lockett MR, Sherman W, Fujita S, Al-Sayah M, Lange H, Bowers CM, Heroux A,
757 Krilov G, Whitesides GM. 2013. Water Networks Contribute to Enthalpy/Entropy
758 Compensation in Protein–Ligand Binding. *J Am Chem Soc* **135**:15579–15584.
759 doi:10.1021/ja4075776
760 Butler BM, Gerek ZN, Kumar S, Ozkan SB. 2015. Conformational dynamics of nonsynonymous
761 variants at protein interfaces reveals disease association. *Proteins: Structure, Function,*
762 *and Bioinformatics* **83**:428–435. doi:10.1002/prot.24748
763 Campbell E, Kaltenbach M, Correy GJ, Carr PD, Porebski BT, Livingstone EK, Afriat-Jurnou L,
764 Buckle AM, Weik M, Hollfelder F, Tokuriki N, Jackson CJ. 2016. The role of protein
765 dynamics in the evolution of new enzyme function. *Nature Chemical Biology* **12**:944–
766 950. doi:10.1038/nchembio.2175
767 Campitelli P, Guo J, Zhou H-X, Ozkan SB. 2018. Hinge-Shift Mechanism Modulates Allosteric
768 Regulations in Human Pin1. *J Phys Chem B* **122**:5623–5629.
769 doi:10.1021/acs.jpcb.7b11971
770 Campitelli P, Modi T, Kumar S, Ozkan SB. 2020a. The Role of Conformational Dynamics and
771 Allostery in Modulating Protein Evolution. *Annual Review of Biophysics* **49**:267–288.
772 doi:10.1146/annurev-biophys-052118-115517
773 Campitelli P, Ozkan SB, Swint-Kruse L. 2020b. Asymmetry in Dynamic Allosteric Residue
774 Coupling (DARC) Interactions Captures Evolutionary Landscape. *Biophysical Journal*
775 **118**:52a. doi:10.1016/j.bpj.2019.11.464
776 Campitelli P, Swint-Kruse L, Ozkan SB. 2021. Substitutions at Nonconserved Rheostat Positions
777 Modulate Function by Rewiring Long-Range, Dynamic Interactions. *Mol Biol Evol*
778 **38**:201–214. doi:10.1093/molbev/msaa202
779 Chang LC, Bewley CA. 2002. Potent Inhibition of HIV-1 Fusion by Cyanovirin-N Requires
780 Only a Single High Affinity Carbohydrate Binding Site: Characterization of Low Affinity
781 Carbohydrate Binding Site Knockout Mutants. *Journal of Molecular Biology* **318**:1–8.
782 doi:10.1016/S0022-2836(02)00045-1
783 Chodera JD, Mobley DL. 2013. Entropy-enthalpy compensation: Role and ramifications in
784 biomolecular ligand recognition and design. *Annu Rev Biophys* **42**:121–142.
785 doi:10.1146/annurev-biophys-083012-130318
786 Cornish-Bowden A. 2002. Enthalpy—entropy compensation: a phantom phenomenon. *Journal*
787 *of Biosciences* **27**:121–126.
788 Davis IW, Baker D. 2009. RosettaLigand Docking with Full Ligand and Receptor Flexibility.
789 *Journal of Molecular Biology* **385**:381–392. doi:10.1016/j.jmb.2008.11.010

790 Davis IW, Raha K, Head MS, Baker D. 2009. Blind docking of pharmaceutically relevant
791 compounds using RosettaLigand. *Protein Science* **18**:1998–2002. doi:10.1002/pro.192
792 de Juan D, Pazos F, Valencia A. 2013. Emerging methods in protein co-evolution. *Nature*
793 *Reviews Genetics* **14**:249–261. doi:10.1038/nrg3414
794 DeLuca S, Khar K, Meiler J. 2015. Fully Flexible Docking of Medium Sized Ligand Libraries
795 with RosettaLigand. *PLoS One* **10**. doi:10.1371/journal.pone.0132508
796 Deng Y, Roux B. 2009. Computations of Standard Binding Free Energies with Molecular
797 Dynamics Simulations. *J Phys Chem B* **113**:2234–2246. doi:10.1021/jp807701h
798 Dunn SD, Wahl LM, Gloor GB. 2008. Mutual information without the influence of phylogeny or
799 entropy dramatically improves residue contact prediction. *Bioinformatics* **24**:333–340.
800 doi:10.1093/bioinformatics/btm604
801 Fox JM, Zhao M, Fink MJ, Kang K, Whitesides GM. 2018. The Molecular Origin of
802 Enthalpy/Entropy Compensation in Biomolecular Recognition. *Annual Review of*
803 *Biophysics* **47**:223–250. doi:10.1146/annurev-biophys-070816-033743
804 Fromme R, Katiliene Z, Fromme P, Ghirlanda G. 2008. Conformational gating of dimannose
805 binding to the antiviral protein cyanovirin revealed from the crystal structure at 1.35 Å
806 resolution. *Protein Science* **17**:939–944. doi:10.1110/ps.083472808
807 Fromme R, Katiliene Z, Giomarelli B, Bogani F, Mc Mahon J, Mori T, Fromme P, Ghirlanda G.
808 2007. A Monovalent Mutant of Cyanovirin-N Provides Insight into the Role of Multiple
809 Interactions with gp120 for Antiviral Activity,. *Biochemistry* **46**:9199–9207.
810 doi:10.1021/bi700666m
811 Fujimoto YK, Green DF. 2012. Carbohydrate Recognition by the Antiviral Lectin Cyanovirin-N.
812 *J Am Chem Soc* **134**:19639–19651. doi:10.1021/ja305755b
813 Gerek ZN, Kumar S, Ozkan SB. 2013. Structural dynamics flexibility informs function and
814 evolution at a proteome scale. *Evolutionary Applications* **6**:423–433.
815 doi:10.1111/eva.12052
816 Gerek ZN, Ozkan SB. 2011. Change in Allosteric Network Affects Binding Affinities of PDZ
817 Domains: Analysis through Perturbation Response Scanning. *PLOS Computational*
818 *Biology* **7**:e1002154. doi:10.1371/journal.pcbi.1002154
819 Gouveia-Oliveira R, Pedersen AG. 2007. Finding coevolving amino acid residues using row and
820 column weighting of mutual information and multi-dimensional amino acid
821 representation. *Algorithms for Molecular Biology* **2**:12. doi:10.1186/1748-7188-2-12
822 He X, Liu S, Lee T-S, Ji B, Man VH, York DM, Wang J. 2020. Fast, Accurate, and Reliable
823 Protocols for Routine Calculations of Protein–Ligand Binding Affinities in Drug Design
824 Projects Using AMBER GPU-TI with ff14SB/GAFF. *ACS Omega* **5**:4611–4619.
825 doi:10.1021/acsomega.9b04233
826 Hopf TA, Green AG, Schubert B, Mersmann S, Schärfe CPI, Ingraham JB, Toth-Petroczy A,
827 Brock K, Riesselman AJ, Palmedo P, Kang C, Sheridan R, Draizen EJ, Dallago C,
828 Sander C, Marks DS. 2019. The EVcouplings Python framework for coevolutionary
829 sequence analysis. *Bioinformatics* **35**:1582–1584. doi:10.1093/bioinformatics/bty862
830 Hopf TA, Schärfe CPI, Rodrigues JPGLM, Green AG, Kohlbacher O, Sander C, Bonvin AMJJ,
831 Marks DS. 2018. Sequence co-evolution gives 3D contacts and structures of protein
832 complexes. *eLife* **3**. doi:10.7554/eLife.03430
833 Ikeguchi M, Ueno J, Sato M, Kidera A. 2005. Protein Structural Change Upon Ligand Binding:
834 Linear Response Theory. *Phys Rev Lett* **94**:078102. doi:10.1103/PhysRevLett.94.078102

Jana B, Morcos F, Onuchic JN. 2014. From structure to function: the convergence of structure based models and co-evolutionary information. *Phys Chem Chem Phys* **16**:6496–6507. doi:10.1039/C3CP55275F

Jiménez-Osés G, Osuna S, Gao X, Sawaya MR, Gilson L, Collier SJ, Huisman GW, Yeates TO, Tang Y, Houk K. 2014. The role of distant mutations and allosteric regulation on LovD active site dynamics. *Nature chemical biology* **10**:431–436.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**:583–589. doi:10.1038/s41586-021-03819-2

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:10.1073/pnas.1314045110

Kelley BS, Chang LC, Bewley CA. 2002. Engineering an Obligate Domain-Swapped Dimer of Cyanovirin-N with Enhanced Anti-HIV Activity. *J Am Chem Soc* **124**:3210–3211. doi:10.1021/ja025537m

Kim DE, DiMaio F, Wang RY-R, Song Y, Baker D. 2014. One contact for every twelve residues allows robust and accurate topology-level protein structure modeling. *Proteins: Structure, Function, and Bioinformatics* **82**:208–218. doi:10.1002/prot.24374

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.

Koharudin LMI, Furey W, Gronenborn AM. 2009. A designed chimeric cyanovirin-N homolog lectin: Structure and molecular basis of sucrose binding. *Proteins: Structure, Function, and Bioinformatics* **77**:904–915. doi:10.1002/prot.22514

Koharudin LMI, Gronenborn AM. 2013. Sweet entanglements—protein: Glycan interactions in two HIV-inactivating lectin families. *Biopolymers* **99**:196–202. doi:10.1002/bip.22106

Koharudin LMI, Viscomi AR, Jee J-G, Ottonello S, Gronenborn AM. 2008. The Evolutionarily Conserved Family of Cyanovirin-N Homologs: Structures and Carbohydrate Specificity. *Structure* **16**:570–584. doi:10.1016/j.str.2008.01.015

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:10.3390/ijms22062895

Kryshtafovych A, Schwede T, Topf M, Fidelis K, Moult J. 2019. Critical assessment of methods of protein structure prediction (CASP)—Round XIII. *Proteins: Structure, Function, and Bioinformatics* **87**:1011–1020. doi:10.1002/prot.25823

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*, Catalysis and regulation • Protein-protein interactions **35**:135–142. doi:10.1016/j.sbi.2015.11.002

Kumar A, Glembo TJ, Ozkan SB. 2015b. The Role of Conformational Dynamics and Allostery in the Disease Development of Human Ferritin. *Biophysical Journal* **109**:1273–1281. doi:10.1016/j.bpj.2015.06.060

Kuriyan J, Eisenberg D. 2007. The origin of protein interactions and allostery in colocalization. *Nature* **450**:983–990. doi:10.1038/nature06524

Larrimore KE, Kazan IC, Kannan L, Kendle RP, Jamal T, Barcus M, Bolia A, Brimijoin S, Zhan C-G, Ozkan SB. 2017. Plant-expressed cocaine hydrolase variants of butyrylcholinesterase exhibit altered allosteric effects of cholinesterase activity and increased inhibitor sensitivity. *Scientific reports* **7**:10419.

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:10.1021/acs.biochem.5b00635

Liu J, Nussinov R. 2017. Energetic redistribution in allostery to execute protein function. *Proc Natl Acad Sci U S A* **114**:7480–7482. doi:10.1073/pnas.1709071114

Lu H-M, Liang J. 2009. Perturbation-based Markovian transmission model for probing allosteric dynamics of large macromolecular assembling: a study of GroEL-GroES. *PLoS Comput Biol* **5**:e1000526. doi:10.1371/journal.pcbi.1000526

Marks DS, Hopf TA, Sander C. 2012. Protein structure prediction from sequence variation. *Nature Biotechnology* **30**:1072–1080. doi:10.1038/nbt.2419

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. *J Biol Chem* **291**:18967–18976. doi:10.1074/jbc.M116.740415

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:10.1016/j.str.2008.05.011

Matei E, Louis JM, Jee J, Gronenborn AM. 2011. NMR solution structure of a cyanovirin homolog from wheat head blight fungus. *Proteins: Structure, Function, and Bioinformatics* **79**:1538–1549. doi:10.1002/prot.22981

Meiler J, Baker D. 2006. ROSETTALIGAND: Protein–small molecule docking with full side-chain flexibility. *Proteins: Structure, Function, and Bioinformatics* **65**:538–548. doi:10.1002/prot.21086

Meinhardt S, Jr MWM, Parente DJ, Swint-Kruse L. 2013. Rheostats and Toggle Switches for Modulating Protein Function. *PLOS ONE* **8**:e83502. doi:10.1371/journal.pone.0083502

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:10.1038/srep41329

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:10.1016/j.sbi.2020.11.007

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 B: Biological Sciences* **373**:20170184. doi:10.1098/rstb.2017.0184

Modi T, Ozkan SB. 2018. Mutations Utilize Dynamic Allostery to Confer Resistance in TEM-1 β-lactamase. *Int J Mol Sci* **19**. doi:10.3390/ijms19123808

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 β-lactamases from substrate promiscuity to specificity. *Nat Commun* **12**:1852. doi:10.1038/s41467-021-22089-0

927 Morcos F. 2020. Protein conformations à la carte, a step further in de novo protein design. *PNAS*
928 **117**:8674–8676. doi:10.1073/pnas.2004188117
929 Morcos F, Jana B, Hwa T, Onuchic JN. 2013. Coevolutionary signals across protein lineages
930 help capture multiple protein conformations. *PNAS* **110**:20533–20538.
931 doi:10.1073/pnas.1315625110
932 Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, Sander C, Zecchina R, Onuchic JN, Hwa
933 T, Weigt M. 2011. Direct-coupling analysis of residue coevolution captures native
934 contacts across many protein families. *PNAS* **108**:E1293–E1301.
935 doi:10.1073/pnas.1111471108
936 Morcos F, Schafer NP, Cheng RR, Onuchic JN, Wolynes PG. 2014. Coevolutionary information,
937 protein folding landscapes, and the thermodynamics of natural selection. *PNAS*
938 **111**:12408–12413. doi:10.1073/pnas.1413575111
939 Mori T, Boyd MR. 2001. Cyanovirin-N, a Potent Human Immunodeficiency Virus-Inactivating
940 Protein, Blocks both CD4-Dependent and CD4-Independent Binding of Soluble gp120
941 (sgp120) to Target Cells, Inhibits sCD4-Induced Binding of sgp120 to Cell-Associated
942 CXCR4, and Dissociates Bound sgp120 from Target Cells. *Antimicrobial Agents and*
943 *Chemotherapy* **45**:664–672. doi:10.1128/AAC.45.3.664-672.2001
944 Okazaki K, Koga N, Takada S, Onuchic JN, Wolynes PG. 2006. Multiple-basin energy
945 landscapes for large-amplitude conformational motions of proteins: Structure-based
946 molecular dynamics simulations. *Proceedings of the National Academy of Sciences*
947 **103**:11844–11849.
948 O'Keefe BR, Smee DF, Turpin JA, Saucedo CJ, Gustafson KR, Mori T, Blakeslee D, Buckheit
949 R, Boyd MR. 2003. Potent Anti-Influenza Activity of Cyanovirin-N and Interactions with
950 Viral Hemagglutinin. *Antimicrobial Agents and Chemotherapy* **47**:2518–2525.
951 doi:10.1128/AAC.47.8.2518-2525.2003
952 Ose N, Butler BM, Kumar A, Ozkan SB, Kumar S. 2020. Dynamic Allosteric Residue Coupling
953 Reveals Disease Mechanism for Gaucher Disease and NSNVS Across the Proteome.
954 *Biophysical Journal* **118**:53a. doi:10.1016/j.bpj.2019.11.472
955 Patsalo V, Raleigh DP, Green DF. 2011. Rational and Computational Design of Stabilized
956 Variants of Cyanovirin-N That Retain Affinity and Specificity for Glycan Ligands.
957 *Biochemistry* **50**:10698–10712. doi:10.1021/bi201411c
958 Percudani R, Montanini B, Ottonello S. 2005. The anti-HIV cyanovirin-N domain is
959 evolutionarily conserved and occurs as a protein module in eukaryotes. *Proteins:*
960 *Structure, Function, and Bioinformatics* **60**:670–678. doi:10.1002/prot.20543
961 Qi X, Yang Y, Su Y, Wang T. 2009. Molecular Cloning and Sequence Analysis of Cyanovirin-N
962 Homology Gene in Ceratopteris thalictroides. *amfj* **99**:78–92. doi:10.1640/0002-8444-
963 99.2.78
964 Ramadugu SK, Li Z, Kashyap HK, Margulis CJ. 2014. The Role of Glu41 in the Binding of
965 Dimannose to P51G-m4-CVN. *Biochemistry* **53**:1477–1484. doi:10.1021/bi4014159
966 Rastelli G, Rio AD, Degliesposti G, Sgobba M. 2010. Fast and accurate predictions of binding
967 free energies using MM-PBSA and MM-GBSA. *Journal of Computational Chemistry*
968 **31**:797–810. doi:10.1002/jcc.21372
969 Risso VA, Sanchez-Ruiz JM, Ozkan SB. 2018. Biotechnological and protein-engineering
970 implications of ancestral protein resurrection. *Current Opinion in Structural Biology*,
971 Engineering and design: New applications • Membranes **51**:106–115.
972 doi:10.1016/j.sbi.2018.02.007

973 Rivoire O, Reynolds KA, Ranganathan R. 2016. Evolution-Based Functional Decomposition of
974 Proteins. *PLOS Computational Biology* **12**:e1004817. doi:10.1371/journal.pcbi.1004817
975 Romero PA, Arnold FH. 2009. Exploring protein fitness landscapes by directed evolution.
976 *Nature Reviews Molecular Cell Biology* **10**:866–876. doi:10.1038/nrm2805
977 Salinas VH, Ranganathan R. 2018. Coevolution-based inference of amino acid interactions
978 underlying protein function. *eLife* **7**:e34300. doi:10.7554/eLife.34300
979 Salverda ML, De Visser JAG, Barlow M. 2010. Natural evolution of TEM-1 β-lactamase:
980 experimental reconstruction and clinical relevance. *FEMS microbiology reviews*
981 **34**:1015–1036.
982 Schneider SH, Kozuch J, Boxer SG. 2021. The Interplay of Electrostatics and Chemical
983 Positioning in the Evolution of Antibiotic Resistance in TEM β-Lactamases. *ACS Central*
984 *Science* **7**:1996–2008.
985 Simonetti FL, Teppa E, Chernomoretz A, Nielsen M, Marino Buslje C. 2013. MISTIC: mutual
986 information server to infer coevolution. *Nucleic Acids Res* **41**:W8–W14.
987 doi:10.1093/nar/gkt427
988 Spoel DVD, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. 2005. GROMACS:
989 Fast, flexible, and free. *Journal of Computational Chemistry* **26**:1701–1718.
990 doi:10.1002/jcc.20291
991 Stiffler MA, Hekstra DR, Ranganathan R. 2015. Evolvability as a Function of Purifying
992 Selection in TEM-1 β-Lactamase. *Cell* **160**:882–892. doi:10.1016/j.cell.2015.01.035
993 Swint-Kruse L, Matthews KS, Smith PE, Pettitt BM. 1998. Comparison of simulated and
994 experimentally determined dynamics for a variant of the LacI DNA-binding domain,
995 NLac-P. *Biophysical journal* **74**:413–421.
996 Thomas VL, McReynolds AC, Shoichet BK. 2010. Structural bases for stability–function
997 tradeoffs in antibiotic resistance. *Journal of molecular biology* **396**:47–59.
998 Tokuriki N, Jackson CJ, Afriat-Jurnou L, Wyganowski KT, Tang R, Tawfik DS. 2012.
999 Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme.
1000 *Nature Communications* **3**:1257. doi:10.1038/ncomms2246
1001 Torgeson KR, Clarkson MW, Granata D, Lindorff-Larsen K, Page R, Peti W. 2022. Conserved
1002 conformational dynamics determine enzyme activity. *Science Advances* **8**:eabo5546.
1003 Tripathi S, Waxham MN, Cheung MS, Liu Y. 2015. Lessons in Protein Design from Combined
1004 Evolution and Conformational Dynamics. *Scientific Reports* **5**:14259.
1005 doi:10.1038/srep14259
1006 Vorontsov II, Miyashita O. 2009. Solution and Crystal Molecular Dynamics Simulation Study of
1007 m4-Cyanovirin-N Mutants Complexed with Di-Mannose. *Biophysical Journal* **97**:2532–
1008 2540. doi:10.1016/j.bpj.2009.08.011
1009 Wagner JR, Sørensen J, Hensley N, Wong C, Zhu C, Perison T, Amaro RE. 2017. POVME 3.0:
1010 Software for Mapping Binding Pocket Flexibility. *J Chem Theory Comput* **13**:4584–
1011 4592. doi:10.1021/acs.jctc.7b00500
1012 Wang R, Lai L, Wang S. 2002. Further development and validation of empirical scoring
1013 functions for structure-based binding affinity prediction. *J Comput Aided Mol Des* **16**:11–
1014 26. doi:10.1023/A:1016357811882
1015 Wang S, Li W, Zhang R, Liu S, Xu J. 2016. CoinFold: a web server for protein contact
1016 prediction and contact-assisted protein folding. *Nucleic Acids Res* **44**:W361–W366.
1017 doi:10.1093/nar/gkw307

1018 Wang S, Sun S, Li Z, Zhang R, Xu J. 2017. Accurate De Novo Prediction of Protein Contact
1019 Map by Ultra-Deep Learning Model. *PLOS Computational Biology* **13**:e1005324.
1020 doi:10.1371/journal.pcbi.1005324
1021 Wei G, Xi W, Nussinov R, Ma B. 2016. Protein Ensembles: How Does Nature Harness
1022 Thermodynamic Fluctuations for Life? The Diverse Functional Roles of Conformational
1023 Ensembles in the Cell. *Chem Rev* **116**:6516–6551. doi:10.1021/acs.chemrev.5b00562
1024 Wodak SJ, Paci E, Dokholyan NV, Berezovsky IN, Horovitz A, Li J, Hilser VJ, Bahar I,
1025 Karanicolas J, Stock G, Hamm P, Stote RH, Eberhardt J, Chebaro Y, Dejaegere A,
1026 Cecchini M, Changeux J-P, Bolhuis PG, Vreede J, Faccioli P, Orioli S, Ravasio R, Yan
1027 L, Brito C, Wyart M, Gkeka P, Rivalta I, Palermo G, McCammon JA, Panecka-Hofman
1028 J, Wade RC, Di Pizio A, Niv MY, Nussinov R, Tsai C-J, Jang H, Padhorny D, Kozakov
1029 D, McLeish T. 2019. Allostery in Its Many Disguises: From Theory to Applications.
1030 *Structure* **27**:566–578. doi:10.1016/j.str.2019.01.003
1031 Woodrum BW, Maxwell JD, Bolia A, Ozkan SB, Ghirlanda G. 2013. The antiviral lectin
1032 cyanovirin-N: probing multivalency and glycan recognition through experimental and
1033 computational approaches. *Biochem Soc Trans* **41**:1170–1176.
1034 doi:10.1042/BST20130154
1035 Xu J. 2019. Distance-based protein folding powered by deep learning. *Proceedings of the*
1036 *National Academy of Sciences* **116**:16856–16865. doi:10.1073/pnas.1821309116
1037 Yang G, Hong N, Baier F, Jackson CJ, Tokuriki N. 2016. Conformational tinkering drives
1038 evolution of a promiscuous activity through indirect mutational effects. *Biochemistry*
1039 **55**:4583–4593.
1040 Zimmerman MI, Hart KM, Sibbald CA, Frederick TE, Jimah JR, Knoverek CR, Tolia NH,
1041 Bowman GR. 2017. Prediction of new stabilizing mutations based on mechanistic
1042 insights from Markov state models. *ACS central science* **3**:1311–1321.
1043 Zou T, Risso VA, Gavira JA, Sanchez-Ruiz JM, Ozkan SB. 2015. Evolution of Conformational
1044 Dynamics Determines the Conversion of a Promiscuous Generalist into a Specialist
1045 Enzyme. *Molecular Biology and Evolution* **32**:132–143. doi:10.1093/molbev/msu281
1046 Zou T, Woodrum BW, Halloran N, Campitelli P, Bobkov AA, Ghirlanda G, Ozkan SB. 2021.
1047 Local Interactions That Contribute Minimal Frustration Determine Foldability. *J Phys*
1048 *Chem B* **125**:2617–2626. doi:10.1021/acs.jpcb.1c00364
1049

1050

1051

1052