Design of novel CV-N variants: Modulation of binding dynamics through distal mutations

I. Can Kazan[1,2 ¶], Prerna Sharma[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 a computational approach to identify distal residues that allosterically modulate dynamics of binding sites by combining dynamic coupling with statistical analysis of co-evolution. Putative mutants of the predicted allosteric sites are subjected to Adaptive BP-Dock for binding analysis. Here, we apply this method to a small lectin, Cyanovirin-N (CV-N), that selectively binds to dimannose. Our method points out mutations on I34, that is 16Å away from binding pocket, can modulate binding. Experimental characterization confirms that I34K abolishes

binding, while I34Y increases affinity. Increased affinity is not due to changes in the binding region, which are conserved in the crystal structure. However, ITC analysis reveals an opposite entropic contribution, suggesting that modulation of dynamics is responsible for the change in binding affinity. Our results point to a novel approach to identify and substitute distal sites by integrating evolutionary inference with protein dynamics in glycan-binding proteins to improve binding affinity.

Introduction

Naturally occurring proteins exhibit a fine balance between protein stability and biological function. Protein sequence underpins any possible protein function, yet deciphering this encoded information is still a challenge (Salinas and Ranganathan, 2018). Evolutionarily related proteins display co-variation at specific sites, pointing to the importance of interactions among these residues in governing 3-D structure and function. Sequence-based prediction methods involving multiple sequence alignment (MSA) to map co-evolved positions through information theory based approaches usually identify the residue pairs that have direct interactions (i.e. in contact in 3-D structure), yet mask allosteric coupling between residues far from each other (long range interactions) or phylogenetic effects (Anishchenko et al., 2017; Jones et al., 2012; Kamisetty et al., 2013). As a result, it is hard to identify important sites distal from functional sites (binding or active sites that modulate function). However, directed evolution studies have shown that these long range interactions between distal residues and binding/active site play a role in function, and substitutions on these distal sites are crucial to optimize leads in computational protein design (Acar et al., 2020; Barrientos et al., 2006; Campbell et al., 2016, 2018; Campitelli et al., 2020a;

46 Keeffe et al., 2011; Otten et al., 2018; Risso et al., 2018; Romero and Arnold, 2009; Tokuriki et
47 al., 2012; Yuan et al., 2005; Zhang et al., 2006).
48
49 To overcome this challenge in protein design, here we combine co-evolution analyses
50 through MSA with protein dynamics using Dynamic Coupling Index (DCI) analysis to identify
51 distal positions potentially involved in modulating binding interactions (Campitelli et al., 2018,
52 2020a; Larrimore et al., 2017). We hypothesize that mutations at these distal sites can modulate
53 binding affinity by altering the conformational dynamics of the binding pocket. Thus, after
54 identifying an allosteric site, we first computationally screened how mutations modulate binding
55 affinity using our in-house flexible docking method Adaptive BP-dock and then validated our
56 predictions experimentally (Bolia and Ozkan, 2016).
57
58 The model protein chosen for characterizing allosteric modulation of binding mediated by
59 long range interactions with co-evolved sites is Cyanovirin-N (CV-N). CV-N is a small (11 kDa)
60 natural lectin isolated from cyanobacterium *Nostoc ellipsosporum* and has been studied
61 extensively as a model to investigate protein interactions with glycans (Barrientos et al., 2003;
62 Botos and Wlodawer, 2005, 2003; Mori and Boyd, 2001; O'Keefe et al., 2003). Most members of
63 the CV-N homology family contain two sequential repeats of about 50 residues. While residues in
64 the core region are highly conserved, residues involved in dimannose binding are not highly
65 conserved resulting in differences in specificity and selectivity of the binding pockets (Koharudin
66 et al., 2008; Koharudin and Gronenborn, 2013; Patsalo et al., 2011; Percudani et al., 2005; Qi et
67 al., 2009). Most homologs characterized so far show high conservation of the structural binding
68 units, but they lose the specificity for dimannose that is a hallmark of CV-N recognition, and bind

69 to oligomannose and other glycans at much higher micromolar concentrations (Koharudin et al.,
70 2009, 2008; Matei et al., 2016; Woodrum et al., 2013).
71
72 CV-N comprises two quasi-symmetric domains, A and B, spanning residues 1-38/90-101
73 and residues 39-89 respectively. The two domains are connected to each other by a short helical
74 linker. Despite almost having identical structures, the domains show relatively low sequence
75 homology (28% sequence identity, and 52% similarity). Functionally, they both bind dimannose,
76 yet the affinity is quite different, with domain B having tighter binding affinity (K_d= 15.3 μM),
77 and domain A showing weak affinity (K_d = 400 μM) (Balzarini, 2007; Bolmstedt et al., 2001; Li
78 et al., 2015). To simplify our analysis we used a designed CV-N variant, P51G-m4, that contains
79 a single high-affinity dimannose-binding site (domain B), folds exclusively as a monomer in
80 physiological conditions, and is more stable to thermal denaturation than wild type (Fromme et
81 al., 2008, 2007).
82
83 Application of our approach on P51G-m4 identified position I34, which is 16 Å away from
84 the binding pocket, as highly coupled and co-evolved with the binding residues of domain B. Based
85 on the co-variation with binding pocket residues, three variants of I34 (K, L, Y) were selected for
86 further analyses. The mutants were modeled by GROMACS Molecular Dynamics (MD)
87 simulations and subjected to docking using Adaptive BP dock. The docking results uncovered that
88 each variant acts differently to binding of dimannose. Experimental biophysical characterization
89 of the mutants confirmed that (i) I34Y increases the binding affinity, (ii) I34L is neutral and
90 consistent with the template, and (iii) I34K abolishes binding affinity towards dimannose. The
91 crystal structure of I34Y with and without dimannose shows high conservation of the overall fold

92 and particularly its binding motif compared to P51G-m4, despite the enhancement in binding
93 affinity. Conformational dynamics analysis by MD revealed that substitution on I34 modulates
94 equilibrium distribution of the conformations of the binding pocket towards open pocket. In
95 addition, dynamic flexibility index (DFI) analysis shows that I34Y mutant results in rigidification
96 of binding residues T57, E41 and N42, in turns decreasing the entropy cost of binding and leading
97 to the favorable binding affinity. A favorable entropy term in binding by I34Y variant is supported
98 by ITC studies. Our findings indicate that it is possible to modulate ligand binding through
99 mutation at a distal location, based on long distance dynamic coupling with binding sites, and
100 present a novel computational approach to identify those sites.
101

102 ## Results and Discussion

103

104 **Combining long-range dynamic coupling analysis with co-evolution allows to identify distal**
105 **sites that contribute to binding affinity.**

106

107 The binding pocket of domain B of CV-N has been subjected to intense scrutiny to glean
108 the origin of its binding specificity for dimannose (Bewley, 2001; Bolia et al., 2014b; Botos and
109 Wlodawer, 2003; Li et al., 2015; Vorontsov and Miyashita, 2009). Previous mutational studies on
110 the binding pocket residues have shown their importance in modulating interaction with
111 dimannose along with a role in modulating water-mediated interactions (Barrientos et al., 2006;
112 Bolia et al., 2014b; Chang and Bewley, 2002; Matei et al., 2008). So far all substitutions of the
113 binding residues led to decreased binding affinity for dimannose on domain B (Bolia et al., 2014b;
114 Fujimoto and Green, 2012; Kelley et al., 2002; Matei et al., 2011; Ramadugu et al., 2014).

115 Furthermore, evolutionary analyses reassures the majority of the binding site residues are
116 conserved in CV-N glycan interactions, suggesting that it is harder to enhance affinity through
117 mutations of these conserved binding sites. (Koharudin et al., 2008; Percudani et al., 2005). Here,
118 we take a completely different approach and aim to identify positions that are not directly
119 interacting with glycans but can potentially impact glycan binding through modulating the protein
120 ensemble in equilibrium. These positions could be far away yet can modulate the binding by
121 altering the conformational dynamics of the binding pocket due to long-distance communication
122 through fluctuations arising anisotropic interaction within the 3-D network of interactions (i.e.,
123 dynamic allosteric communications) (Acar et al., 2020; Campbell et al., 2016; Gunasekaran et al.,
124 2004; Lu and Liang, 2009; Wodak et al., 2019). We call positions that exhibit dynamic coupling
125 with the binding sites and mediate long-distance communication dynamic allosteric residue
126 coupling (DARC) spots. Our protein dynamics analyses on different protein families, particularly
127 on ancestral resurrected proteins have revealed that many proteins harbor DARC spots and that
128 substitutions on the DARC spots play a critical role in modulating the function for adaptation to
129 new environments (Kumar et al., 2015b; Modi and Ozkan, 2018; Ose et al., 2020).
130
131 To seek DARC spots in the CV-N scaffold we first applied our Dynamic Flexibility Index
132 (DFI) and Dynamic coupling Index (DCI) analyses. DFI score measures the resilience of a position
133 by computing the total fluctuation response of a position when each position in the protein is
134 perturbed with a random unit force and is related to the flexibility/rigidity of a given position (see
135 methods). Our earlier work has shown that rigid positions, which we call *hinge sites*, contribute
136 the most to the equilibrium dynamics that sample conformations in native state and distal hinge

137 positions that are coupled to the active site could modulate dynamics to alter function (Gerek and
138 Ozkan, 2011; Kumar et al., 2015b; Larrimore et al., 2017; Modi and Ozkan, 2018).
139
140 As Natural selection strongly shapes the mutational landscape through protein dynamics
141 (Campitelli et al., 2020a; Modi et al., 2021), we combine evolutionary inference obtained from
142 multiple sequence alignment with position specific, dynamics-based DCI and DFI profiles of CVN
143 to modify the binding affinity of a lectin towards a specific glycan. Particularly, we searched for
144 distal rigid positions (exhibiting %DFI values 0.2 or lower and at least 15Å away from binding
145 residues) that are dynamically coupled and co-evolved with the binding pocket (showing 0.7 or
146 higher dynamic coupling with the binding site) (Supplementary file 1 table 1).
147
148 DFI profile for CV-N (Fig 1A) shows that the region around residue I34 is rigid (low score)
149 implicating this distal location in transferring allosteric communication and controlling the
150 dynamics of the binding site, over a 16Å distance (Table S1).
151

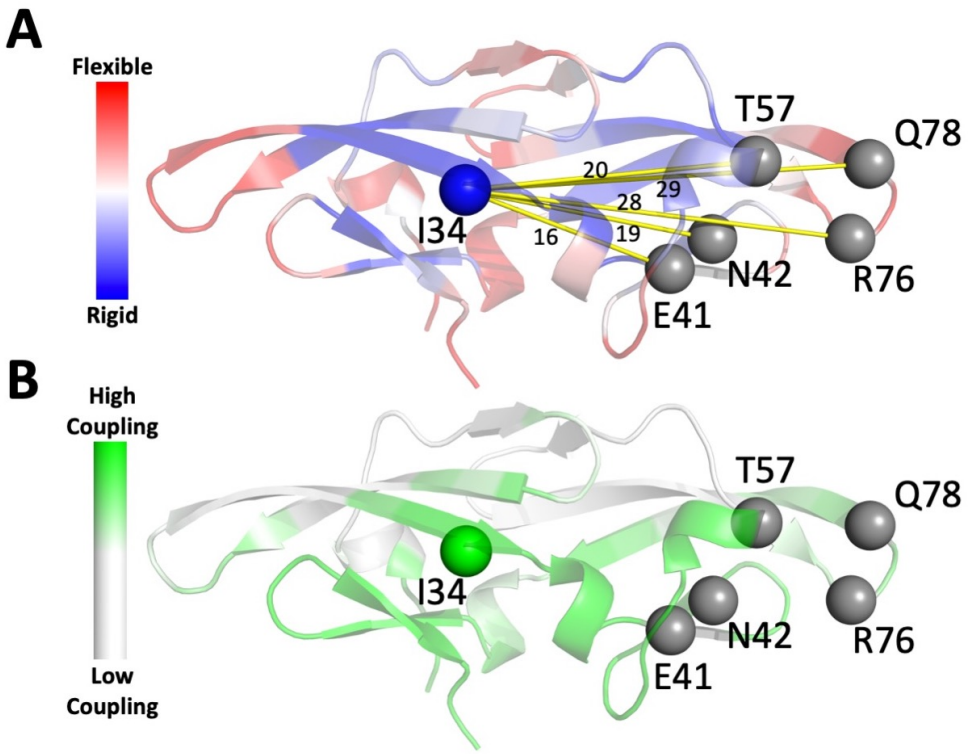

Fig 1 A) The DFI profile mapped onto CV-N structure within a color spectrum of red white blue where red presents positions with high DFI, (very flexibile sites) , and blue presents positions with low DFI values (rigid sites). Residue I34 shows low DFI score, suggesting its critical role in mediating the collective motion and its importance in function. B) DCI profile projected on CV-N structure in a color-coded spectrum sites exhibiting high DCI are green. Residue I34 is found as highly dynamically coupled to binding site residues. The color-code maps from dynamics metrics suggest I34 can potentially be an allosteric modulator for binding where substitutions can alter glycan interactions.

To further analyze potential impact of position I34 on binding, we applied the DCI metric to capture its long-distance dynamic coupling with the binding residues. DCI metric shows the fluctuation response of a position of interest when a functional group (i.e., binding sites) is

156 perturbed by unit forces relative to its fluctuation response to a perturbation to any other positions
157 in the protein. In other words, it measures how much a residue of interest is dynamically coupled
158 to another position, particularly functionally important positions (Campitelli et al., 2018;
159 Larrimore et al., 2017). We found that residue I34 is highly coupled to binding site residues of
160 domain B of CV-N (Fig 1B) (Table S1) (Supplementary file 1 table 1), suggesting that
161 substitutions at this position could modulate binding dynamics as observed in other systems
162 (Campitelli et al., 2020a; Risso et al., 2018). Additionally, we applied co-evolutionary analyses to
163 estimate evolutionary coupling between residue I34 and binding site residues (Supplementary file
164 1 table 1) (Hopf et al., 2019; Ma et al., 2015; Simonetti et al., 2013) (See methods). We found that
165 I34 is co-evolved with five binding site positions (Table S1). When both conformational dynamics
166 analyses, and co-evolution analyses are used to select residue positions that fit our criteria, I34
167 emerges as the one and only hinge position that is more than 15Å away and the only co-evolved
168 position exhibiting dynamic coupling with the binding site. We note that it is remarkable
169 considering hundred locations in CV-N. only one residue that is distal and both exhibiting high
170 dynamic coupling and co-evolved with the binding sites. emerged from our analyses. Therefore,
171 we selected I34 for further investigation.
172
173 **Substitutions at a distal site yield to change in binding affinity from enhancing to abolishing**
174 **glycan binding**
175
176 Next, we selected type of amino acid substitutions on position I34 (i.e., I34 variants) based
177 on the amino-acid frequencies observed using MSA of CV-N homolog sequences. To this aim, we
178 first generated a multiple sequence alignment using CVN family members whose binding site

179 residues have identical or similar amino-acid composition to P51G-m4 CV-N. We then obtained
180 the frequency of amino acid types at position 34 for these proteins. This analysis indicates that the
181 most frequently observed amino acids at this position are L and Y. As the earlier studies displays
182 (Jana et al., 2014; Modi et al., 2021; Morcos, 2020; Morcos et al., 2014, 2013), negative selection
183 (i.e. elimination of amino-acid types that are detrimental to the folding) is a major force in shaping
184 the mutational landscape. We can confidently assume that these frequently observed residue types
185 on positions 34 do not disrupt folding due to their presence in the MSA. We also selected K as a
186 mutant, because it is the only positively charged amino acid observed at position
187 34. (Supplementary file 1 table 2)
188
189 These new mutants were then modeled on the coordinates of P51G-m4 CV-N and subjected
190 to Molecular Dynamics simulations (Abraham et al., 2015; Spoel et al., 2005). At the end of the
191 simulations, the conformations are clustered by their root-mean-square deviation (RMSD), and the
192 most frequently sampled conformation is selected for docking analysis using our in house
193 algorithm Adaptive BP-Dock (Bolia and Ozkan, 2016), which incorporates both ligand and
194 receptor flexibility into computational methods to accurately sample binding induced
195 conformations during docking. These features are particularly important when docking highly
196 flexible ligands such as glycans, and we previously used Adaptive BP-Dock to analyze and predict
197 binding to dimannose in a series of CV-N binding site mutants (Bolia et al., 2014b; Li et al., 2015;
198 Woodrum et al., 2013). Here, we apply Adaptive BP-Dock to experimentally characterized wild-
199 type CV-N, and its variants: P51G-m4, mutDB (a mutant in which binding by domain B has been
200 obliterated) for benchmarking, and then compare their binding score energies to those of I34
201 variants. Adaptive BP-Dock has previously used to analyze different single point mutants of P51G-

202 m4 CV-N and good correlations were obtained between X-scores binding energy scores and
203 experimentally measured binding affinities (K_d) (Bolia et al., 2014b; Li et al., 2015). These
204 previous studies also showed that -6 kcal/mol X-score binding energy is a good threshold to
205 differentiate dimannose binding CV-Ns from mutants that cannot bind.
206
207 Based on the previous analysis binding energy X-scores, Adaptive BP-Dock correctly
208 assessed binding of benchmark proteins and predicted that substitutions on I34 modulate binding
209 affinity in a wide range: I34Y variant leads to an increase in binding affinity (beneficial), I34K
210 decreases the binding affinity (deleterious), and I34L yields no change (neutral). It is interesting
211 to observe that a distal site can modulate the binding with a wide range of outcome based on amino-
212 acid substitutions. This finding has also been observed for allosterically regulated enzymes such
213 as LacI where different amino-acid substitutions on non-conserved sites lead to gradual changes
214 in function, acting like a rheostatic switch to modulate function through conformational dynamics
215 (Campitelli et al., 2020b).
216
217
218
219
220
221
222

Table 1. The predicted binding affinities of domain B and comparison with experimental ITC data for P51G-m4 and its I34 variants. Docking results shows I34Y mutant has higher binding towards dimannose improving on its P51G-m4 template binding affinity.

Protein	Predicted Binding Score (kcal/mol)	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	-6.0 ± 0.1	
mutDB	-5.97	No-binding			
P51G-m4	-6.62	117 ± 3	-12.3 ± 0.3	-7.0 ± 0.3	-5.3 ± 0.3
P51G-m4-I34K	-5.85	No-binding	No-binding	No-binding	No-binding
P51G-m4-I34L	-6.19	148 ± 2	-9.6 ± 0.1	-4.4 ± 0.1	-5.2 ± 0.1
P51G-m4-I34Y	-6.75	64 ± 5	-4.35 ± 0.1	1.32 ± 0.2	-5.67 ± 0.2

223

224 The computational predictions were validated by experimental characterization of the
225 mutant proteins, prepared as described in Methods.

226

227 We assessed the folding and thermal stability of the mutants by circular dichroism (CD)
228 spectroscopy. The Far-UV CD spectra of the mutants showed that all the mutants are well folded
229 and adopt a fold similar to the parent protein, characterized by a single negative band centered at
230 216 nm. Thermal denaturation was used to measure apparent melting temperature (T_m) values. We
231 found that P51G-m4 and I34L are equally stable, with apparent T_m of 58 and 57.8 respectively. In
232 contrast, the I34Y and I34K mutants were less thermostable than P51G-m4 as shown by apparent
233 T_m values of 54.7 and 47 °C respectively. Substituting a hydrophobic residue with a basic aliphatic

234 amino acid (lysine) has a large destabilizing effect, while aromatic and polar tyrosine is better
235 tolerated. The trend for thermostability is P51G-m4 ~ I34L > I34Y > I34K (Fig S1).
236
237 Chemical denaturation experiments were used to extract thermodynamic values, after
238 ensuring complete equilibration at each concentration of guanidinium hydrochloride by incubating
239 the samples for 72-hour (Patsalo et al., 2011). The results align with the thermal denaturation
240 results: P51G-m4 is the most stable to denaturant, followed by I34L, I34Y and I34K (Fig S2). The
241 data were analyzed as described in methods to extract ΔG_{H20} values and C_m values of for P51G-
242 m4, I34L, I34Y and I34K of 3.0, 2.94, 2.91 and 2.38 kcal/mol and of 1.45, 1.39, 1.13 and 0.68 M
243 respectively (Table 2).
244

Table 2. The ΔG_{H2O} and C_m values for CV-N mutants obtained from chemical denaturation experiments

Variants	ΔG_{H2O} (Kcal/mol)	C_m (M)
P51G-m4	3.01 ± 0.047	1.46 ± 0.019
P51G-m4-I34L	2.95 ± 0.077	1.39 ± 0.009
P51G-m4-I34Y	2.91 ± 0.157	1.13 ± 0.017
P51G-m4-I34K	2.40 ± 0.124	0.68 ± 0.015

245
246 Next, we evaluated the impact of the mutations on the function. The binding affinities of
247 the mutants for dimannose were assessed by isothermal titration calorimetry (ITC) (Fig S3),
248 obtaining K_d values listed in Table 1. We found that I34Y binds dimannose with tightest affinity
249 of all the mutants tested and showed a two-fold improvement over P51G-m4, with K_d of 64 μM
250 as compared to 117 μM. I34L showed a slight decrease in binding with a K_d of 148 μM. No binding
251 was observed for I34K. Thermodynamic values extracted from ITC experiments (Table 1) also

252 show that the change in binding is related to changes in entropy, particularly for I34Y. These data
253 support the hypothesis that I34 modulates the conformational dynamics of the pocket.
254
255 To glean more information on the effect of the mutations, we determined the X-ray
256 structure of the I34Y mutant with and without dimannose and compared it to the template protein
257 P51G-m4 by superimposing the structures. The fold is highly conserved (Fig 2) as shown by root
258 mean squared distance (RMSD) of 0.16 and 0.20 Å with bound and unbound I34Y, respectively,
259 and tyrosine is well tolerated at position I34. The binding pocket region is also structurally
260 conserved compared to P51G-m4. Analysis of the polar contacts with dimannose compared
261 between P51G-m4 and I34Y (Fig 2B) shows that both proteins make an identical number of
262 hydrogen bonds (11) with the ligand, indicating a conserved binding pose for dimannose. These
263 observations suggest that the increase in binding affinity of I34Y towards dimannose might be
264 mediated by equilibrium dynamics, which are not captured by the crystal structure. This hypothesis
265 is supported by the changes in entropy compensation measured experimentally (ITC) in dimannose
266 binding by P51G-m4 (negative TDS) and I34Y (positive TDS).
267

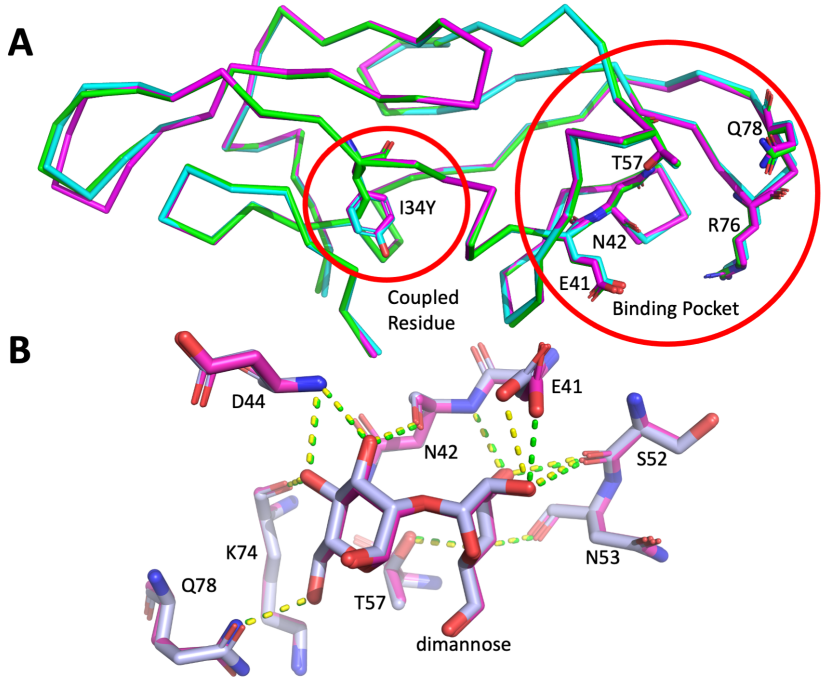

Fig 2 A) The crystal structures of I34Y (bound in magenta and unbound in cyan) and its template protein P51G-m4 (green) are superimposed. B) Overlay of bound structures of I34Y (magenta) and P51G-m4 (grey) (RMSD 0.15 Å); dashed lines depict polar interactions with dimannose.

To gather atomic level detail on how the substitutions on I34 dynamically modulate the binding affinity, we performed two microsecond molecular dynamics simulations for each variant and P51G-m4 to obtain equilibrium dynamics (See Methods for details of the simulations). Our earlier work showed that residues N53 and N42 play a key role in regulating the binding pocket accessibility to the glycans: A hydrogen bond between the amide hydrogen of N42 and carbonyl oxygen of N53 forms a closed pocket, hindering glycan accessibility, whereas the loss of this hydrogen bond leads to an open pocket (Li et al., 2015). Comparison of open and closed conformations based on this hydrogen bond shows that I34Y variant samples the open binding pocket more often than P51G-m4 (Fig S4).

279 Another compelling evidence differentiating I34 variants from P51G-m4 is the change in
280 their binding pocket volumes. The MD simulations were analyzed by POVME pocket volume
281 calculation tool (Wagner et al., 2017). The calculated pocket volumes for I34Y, I34K and P51G-
282 m4 were converted into frequencies to obtain probability distribution of the pocket volumes (Fig
283 3A), revealing that I34Y variant samples a more compact pocket volume compared to P51G-m4.
284 If the pocket is too small or too large dimannose cannot maximize its interaction with the protein,
285 and a compact conformation enables dimannose to easily make the necessary hydrogen bond
286 interactions with the protein. This optimum pocket volume sampled by I34Y may also explain the
287 different binding energetics observed by ITC, in which a positive entropy change upon binding
288 compensates for the loss in enthalpy compared to P51G-m4. Pocket volume analysis reveals a
289 larger value for I34K compared to P51G-m4, suggesting that this mutant cannot accommodate the
290 necessary interactions with the dimannose resulting in loss of binding. We applied the same pocket
291 volume calculation to the X-ray structures of P51G-m4 and I34Y variant, and we found volumes
292 of 141Å^3 and 114Å^3 for P51G-m4 and I34Y, respectively in the unbound forms (Fig 3B). These
293 volumes from structures correlates well with the mean volumes from MD simulation trajectories,
294 hinting that the variants modulate the conformational dynamics of binding pocket. Overall, in
295 agreement with ITC analysis, the results show a shift in native state ensemble upon allosteric
296 mutations, modifying the glycan-CV-N interactions. While our conformational dynamics analysis
297 suggest that the shift of the native ensemble towards a smaller pocket volume upon I34Y mutation
298 can explain the decrease in the entropic cost of binding, observed as positive entropy change in
299 ITC, the drastic decrease in enthalpic component in ITC fit is further investigated by analyzing
300 binding energetics of the docked poses obtained through performing dimannose docking on 2000
301 different conformations sampled accordingly from the binding pocket volume distributions. This

302 analysis shows that the small volume restrict accessibility to the side-chain conformations of
303 binding residue R76 in I34Y variant yielding different hydrogen bond patterns with the dimannose
304 (Fig S5), suggesting a loss in enthalpic contribution.
305

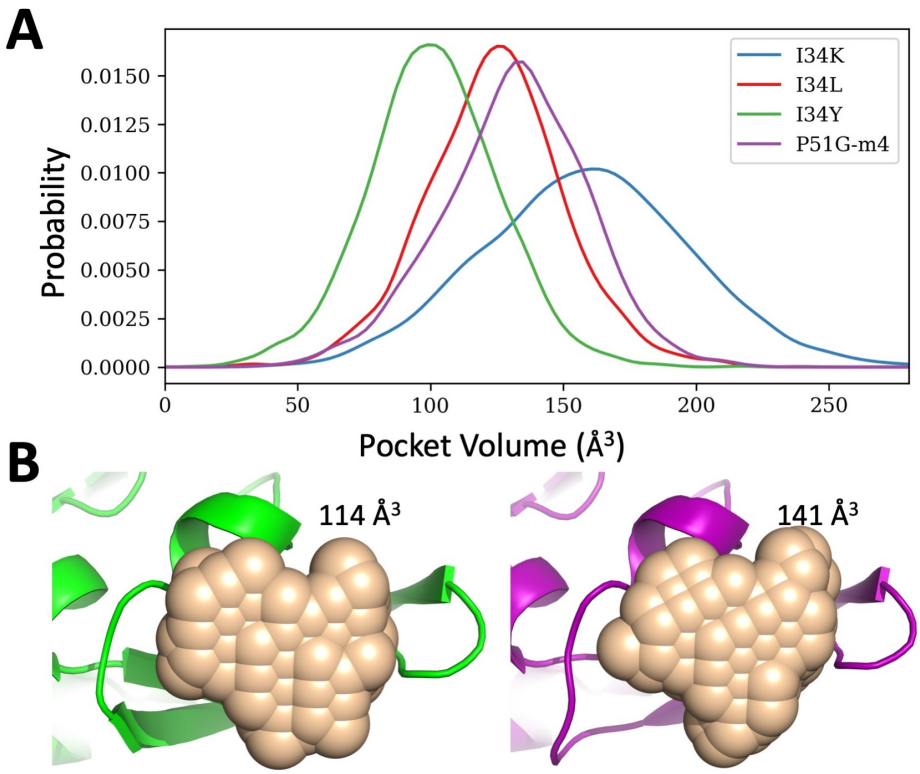

Fig 3 A) Probability distribution of the pocket volume analyses obtained from MD simulation trajectories. Comparison between different variants presents that I34Y stays in an optimum volume more than others favoring a suitable pocket for binding. P51G-m4 and I34L variant sample similar pocket volumes. 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 colored structure on right) and I34Y variant (green colored structure on left).

306

307 **Substitutions of I34 modulates conformational ensemble leading to change in glycan binding**
308 **affinity**
309
310 We and others have shown that nature adapts to a new environment or changes the function
311 of a protein by modulating the native state ensemble through mutations of different positions while
312 keeping the 3-D structure conserved (Campitelli et al., 2020a; Kuriyan and Eisenberg, 2007; Li et
313 al., 2015; Liu and Nussinov, 2017; Modi and Ozkan, 2018; Risso et al., 2018; Tripathi et al., 2015;
314 Woodrum et al., 2013). Furthermore, the shift in native state ensemble is reflected in equilibrium
315 dynamics and the flexibility profiles obtained by DFI capture this shift. DFI profiles that show the
316 flexibility/rigidity of each position withing the 3-D structure correlate well with the changes in
317 function throughout evolution. We observed a similar pattern of conservation of structure yet
318 change in function in our designed CV-N I34 variants, thus we further analyzed the DFI profiles
319 of I34 variants. We clustered the DFI profiles of I34 mutants and P51G-m4 using principal
320 component analyses (PCA) and compared with a 2-D map of these mutants in which melting
321 temperature and K_d are used as reaction coordinates (Fig 4). The 2-D map reveals I34L, P51G-m4,
322 and I34Y clustered together, particularly I34L, P51G-m4, is the closest in this 2-D map while I34K
323 is on the other corner of the map based on stability and binding affinity (Fig 4A). Interestingly, the
324 dendrogram constructed based on the DFI similarities of these proteins using PCA analysis
325 captures biophysical properties (Fig 4B). The clustered DFI profiles put P51G-m4 and I34L variant
326 under the same branch, suggesting their dynamics are very similar; thus, function associated with
327 these dynamics should also be similar. I34Y is also under the same main cluster with the other two
328 proteins that also exhibit dimannose affinity, yet it is in a different branch. This presents that I34Y
329 exhibits slightly different DFI profile suggesting that slight modification in its DFI profile may be

330 associated with enhanced binding affinity. On the other hand, I34K is under a separate branch,
331 showing that its dynamics is completely different than others. Therefore, the DFI clustering mirrors
332 the clusters observed in the functional 2-D map, implying a correlation between changes in
333 dynamics and changes in fitness. This is in agreement with our previous studies, where
334 substitutions on DARC spots modulates flexibility profiles to adapt to a new environment
335 (Campitelli et al., 2020b; Kumar et al., 2015b).
336

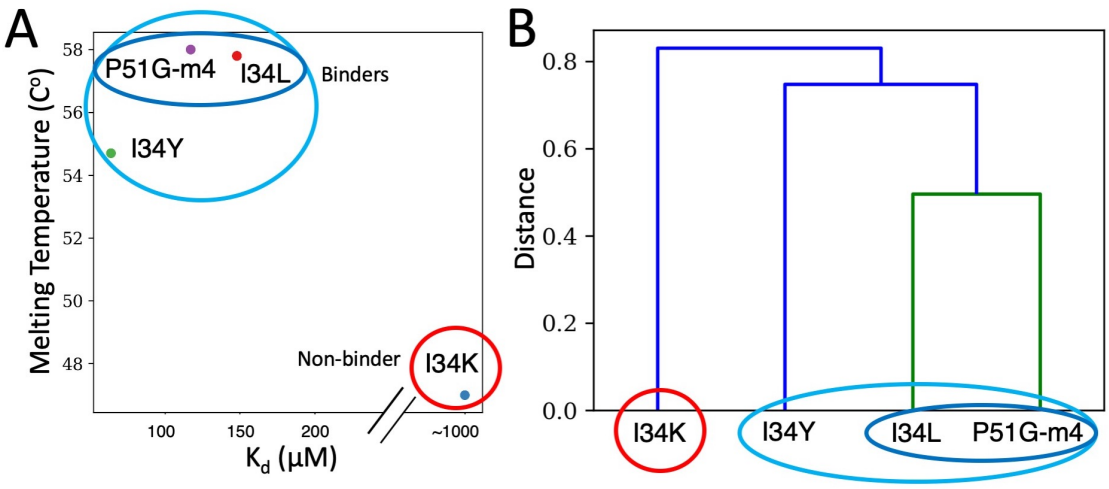

Fig 4 A) K_d and Melting temperature of P51G-m4 and its variants plotted together B) PCA clustering on the first two principal components of the DFI profiles of P51G-m4 and its variants as a dendrogram.

337

338 As clustering of mutants using their DFI profiles aligns with their biophysical
339 characterization, we further investigated the change in DFI profiles of the binding sites with respect
340 to the P51G-m4 for each mutant (Fig 5A). This analysis reveals that change in DFI values of the
341 binding site position T57 exhibits distinct variation among I34 mutants. Position T57 in the case
342 of deleterious mutant (abolishing binding) I34K, exhibits higher flexibility, suggesting a higher

343 entropic penalty for binding interactions. The flexibility profile of I34L variant in position T57 is
344 similar to that of P51G-m4. Accordingly, they show similar binding affinities. In contrast, T57
345 becomes much more rigid for I34Y mutant, suggesting that a decrease in the entropic cost can
346 contribute to the binding affinity enhancement of this mutant. In short, the change in flexibility of
347 the position T57 correlates well with the change in binding affinities of these different mutants.
348 This is indeed not a surprising observation, since our previous study of single point mutations of
349 the binding pocket residues have shown that T57 provides one of the key side chain interactions
350 with dimannose.
351
352 The rigidification of T57, a position far from the mutation location I34, is a compelling
353 evidence that the distal mutation is allosterically controlling the binding site dynamics. We further
354 computed the network of interactions that connects the residue position 34 to 57 and investigated
355 whether distinct pathways emerge after mutation of I34Y. We analyzed the hydrogen bond
356 networks, particularly computed the possible network of hydrogen bonds creating pathways from
357 34 to 57 using the sampled snapshots in our MD trajectories (Fig S6). This analysis presents a
358 unique pathway from 34 to 57 by first forming a new hydrogen bond between the side chain
359 oxygen of the Tyrosine of residue 34 and the nitrogen of the Tyrosine of residue 100 in I34Y
360 variant. Furthermore, a second pathway is also found which is sampled much more frequently in
361 I34Y variant strengthening the communication between position 34 and 57. Thus, both pathways
362 may contribute to the rigidification of T57. We also analyzed the structures from MD clustered
363 with highest percentage based on alpha carbon RMSD for I34Y and P51G-m4 and compared the
364 hydrogen bond interactions of T57 and its neighboring residues. The closest neighbors of T57;
365 positions I55, E56, and C58 conserved their hydrogen bond interactions with their surrounding

366 residues between P51G-m4 and I34Y. On the other hand, T57 is observed to make an extra
367 hydrogen bond interaction in I34Y compared to P51G-m4 (Fig 5B), suggesting that enhancement
368 in hydrogen bond networking of T57 in I34Y leads to rigidification of this position in equilibrium
369 dynamics.
370

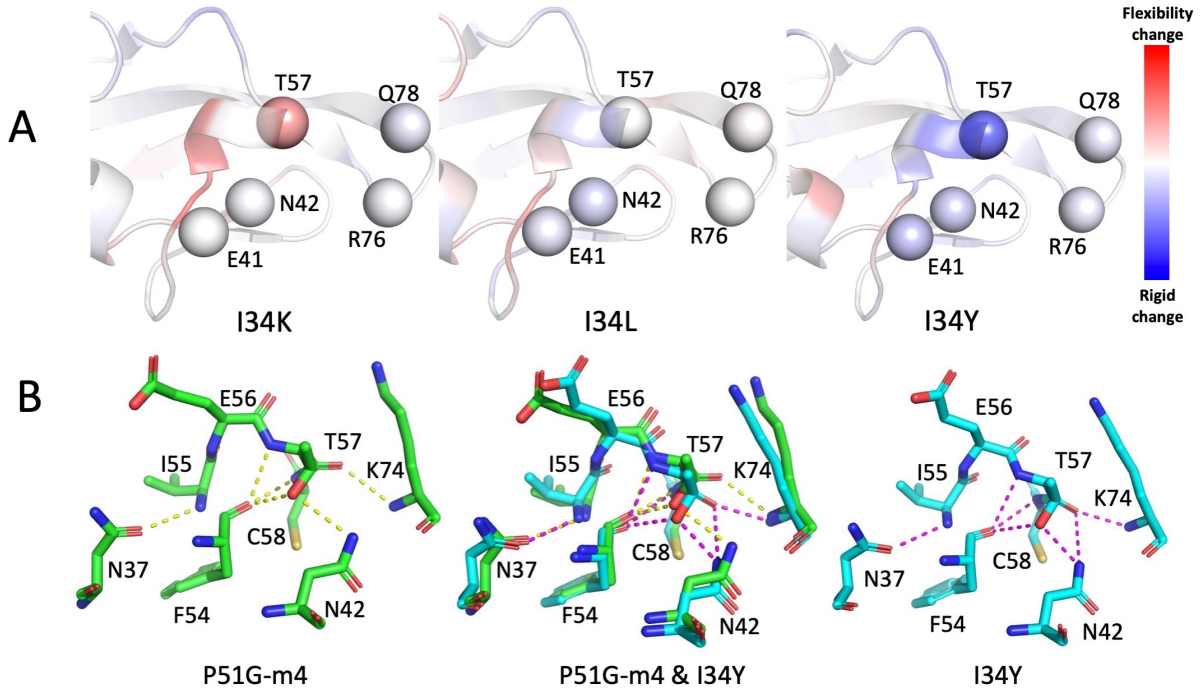

Fig 5 A) Change in DFI of I34K, I34L, and I34Y relative to P51G-m4 are projected on structure. Residues E41, N42, and T57 rigidifies on I34Y compared to P51G-m4. B) Hydrogen bonding interactions of residues I55, E56, T57, and C58 are shown for P51G-m4 and I34Y variant. The rigidification on location T57 can be understood by its neighboring interactions. Compared to P51G-m4, the I34Y variant makes an extra hydrogen bond making it a more rigid residue. This rigidity is reflected on the binding as T57 makes direct interaction with dimannose.

371

372 Taken together, these analyses are in agreement with experimental observation of
373 modulation of entropy of binding to alter affinity where mutation on I34 distally regulates the
374 binding pocket dynamics particularly the position T57, leading to a change in binding energies.
375

376 **Conclusion**

377

378 Multiple evidence point to the importance of mutations on sites distal from functional sites
379 in modulating function during evolution, (Acar et al., 2020; Campbell et al., 2016, 2016;
380 Campitelli et al., 2020a; Gunasekaran et al., 2004; Lu and Liang, 2009; Otten et al., 2018; Wodak
381 et al., 2019) yet predicting which sites are involved through computational analysis is challenging.
382 Indeed, many of these sites are identified only after rounds of directed evolution in protein design
383 (Barrientos et al., 2006; Keeffe et al., 2011; Romero and Arnold, 2009; Yuan et al., 2005). Here,
384 we showed that it is possible to identify these distal sites and fine-tune the binding interactions by
385 integrating long range dynamic coupling with coevolution principles. Coupling DCI approach
386 which selectively identify distal positions that are much more sensitive to the fluctuations of
387 functional sites with evolutionary inference of mutational landscape allowed us to identify mutants
388 that may modify function. We first computationally screened impact of these mutants via Adaptive
389 -BP dock and distinguished beneficial variants from those deleterious. Specifically, we observed
390 that substitutions on these distal sites altered equilibrium dynamics, while conserving the 3-D fold
391 and the binding motif, as also highlighted as change in change in entropic contribution in ITC
392 binding analysis.
393

394 Our new approach that couples evolutionary inferences with long-range dynamic
395 communications within the 3-D network of interactions paves the way to incorporating distal
396 mutants into protein design.
397

398 ## Methods

399

400 **Adaptive BP-Dock**

401

402 Adaptive backbone perturbation docking, Adaptive BP-Dock in short, allows us to model
403 the interaction between CV-N and glycans *in silico* (Bolia and Ozkan, 2016). Adaptive BP-dock
404 combines the complex simulation of backbone flexibility of a protein into Rosetta's ligand docking
405 application (Davis and Baker, 2009). The common restriction in docking is the implementation of
406 flexibility of receptor and ligand (Davis et al., 2009; Davis and Baker, 2009; DeLuca et al., 2015;
407 Meiler and Baker, 2006). Rosetta included the flexibility of ligand in their monte-carlo sampling
408 approach but lacking full receptor flexibility. This high order challenge is overcome by utilizing
409 Perturbation Response Scanning (PRS) to compute backbone changes during docking (Atilgan and
410 Atilgan, 2009; Bolia et al., 2014b, 2014a; Ikeguchi et al., 2005). This procedure also allows the
411 modeling of transition from an unbound state to a bound state (Bolia and Ozkan, 2016). The
412 computational cost of sampling is reduced by using a coarse-grained approach employing Elastic
413 Network Model (ENM) leading to an efficient way of computing backbone perturbations,
414 mimicking the ligand interacting with receptor (Atilgan et al., 2001, 2010; Atilgan and Atilgan,
415 2009).
416

We employed Adaptive BP-Dock in modeling glycan CV-N interactions starting from an unbound conformation of CV-N. The perturbed pose of the protein is calculated using PRS. The structure is then minimized, and the side chains are added at this step. The glycan is docked to the minimized structure using RosettaLigand algorithm. Rosetta samples bound conformations using a knowledge based potential function and calculates bound pose energies. The lowest energy docked pose is selected and feed back to perturbation step, and the same procedure is followed iteratively until a convergence is reached. At the end of each iteration the lowest energy docked pose is taken and binding score is calculated using an empirical scoring function X-score. X-score has shown to provide higher correlations with experimental results (Wang et al., 2002). The flow of the algorithm is shown in Fig 6.

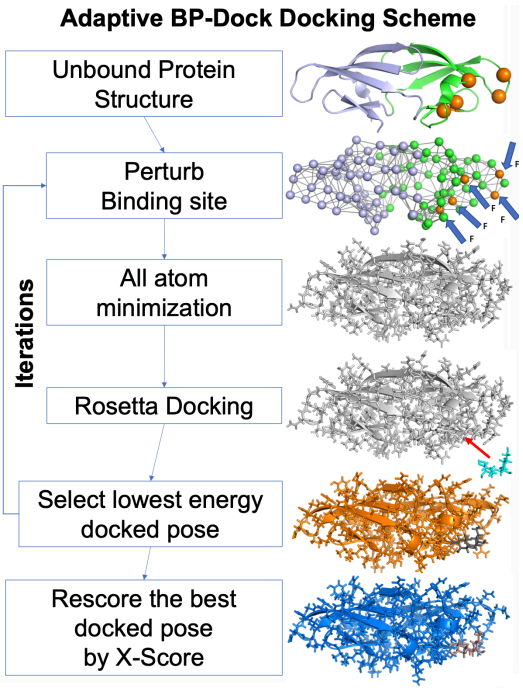

Fig 6 The flow of Adaptive BP-Dock. Starting from an unbound structure the backbone atoms are perturbed using PRS and perturbed pose is fed into minimization. The minimized

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 (Wang et al., 2002).

Adaptive BP-Docks iterative algorithm ensures the sampling does not stuck in a local minimum and reaches a global minimum. The challenge of unbound/bound modeling is solved using the iterative approach as the conformations are led towards a bound pose with the help of PRS.

Molecular Dynamics (MD)

Gromacs simulation is conducted for the mutants (I34K, I34L, I34Y) and P51G-m4 CV-N (Abraham et al., 2015; Spoel et al., 2005). For each simulation the all-atom system is parametrized with CHARMM36 force field and explicit water model TIP3P. The solvation box is set to be minimum 16Å from protein edge. The system is then neutralized by potassium ions to sustain electroneutrality and minimized with steepest descent for 10000 steps. A short-restrained equilibrium is conducted in the constant number of particles, pressure, and temperature ensemble (NPT) for 5 ns using the Berendsen method at 300K temperature and 1 bar pressure. NPT production trajectories were performed with Nose-Hoover and Parrinello-Rahman 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.

447

448 **Dynamic Flexibility Index (DFI)**

449

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

454

455 $$[\Delta \mathbf{R}]_{3Nx1} = [\mathbf{H}]^{-1}_{3Nx3N}[\mathbf{F}]_{3Nx1}$$

456

457 $$DFI_i = \frac{\sum_{j=N}^{N} |\Delta R^j|_i}{\sum_{i=1}^{N} \sum_{j=1}^{N} |\Delta R^j|_i}$$

458

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

465

466 Residues with low DFI score (below 0.2) are considered as hinge points. These points are
467 communication hubs in this 3-D interaction network. Due to high coordination number, the

468 residues exhibiting low DFI values are crucial as information gateways. While they do not exhibit
469 high residue fluctuation to the perturbations, they quickly transfer the perturbation information to
470 other parts, thus they are in control of collective motion of the protein. A change in low DFI
471 positions (i.e., a mutation) will lead to a transformation in the communication grid and majority of
472 disease-associated (i.e. function altering mutations) are often observed in hinge, low DFI sites
473 (Butler et al., 2015; Gerek et al., 2013; Kumar et al., 2015a). The substitution on these site usually
474 alters catalytic activity or binding interaction (i.e., glycans) by modulating equilibrium dynamics
475 (Campitelli et al., 2020a).
476
477 **Dynamic Coupling Index (DCI)**
478
479 Dynamic Coupling Index (DCI) exploits the same framework of DFI (Campitelli et al.,
480 2020a; Larrimore et al., 2017). DCI utilizes the residue response fluctuation upon random force
481 perturbation at a specific residue positions to investigate residues that exhibit long-range coupling
482 to each other. In DCI approach, a unit force is applied on functional residues (i.e., binding site
483 residues) one by one and responses of all other residues are calculated.
484

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

485
486
487 With DCI scheme the residues with high response (high DCI score) indicates high long
488 range dynamic coupling. Residues with high DCI values with binding sites play a critical role in
489 intercommunication of a protein with the binding residues. These coupled residues are of utmost

490 importance in how forces propagate through amino acid chain network on a binding event. Some
491 of the coupled residues are far from the binding site but still encompass modulation capabilities
492 over binding pocket.
493
494 **Informing dynamics from coevolution**
495
496 Coevolutionary data paves the way to assessing 3-D structural contacts by utilizing
497 available sequence information (Hopf et al., 2018; Marks et al., 2012; Morcos et al., 2014).
498 Sequence information is more abundant compared to resolved protein structures. Exploiting the
499 sequence information, primary contacts comparable to realistic structural contacts can be
500 calculated and a contact matrix is formed. The accuracy of these contact maps is proved to be
501 valuable in protein folding studies (Kryshtafovych et al., 2019; Morcos et al., 2011; Wang et al.,
502 2016). Evolutionary coupling (EC) analysis is used to collect information on how much two
503 residues in a protein sequence is in close proximity in 3-D structure. EC scores could be calculated
504 by many different statistical approaches. In this study EC information is gathered by using
505 RaptorX, EVcouplings, and MISTIC webservers (Hopf et al., 2019; Simonetti et al., 2013; Wang
506 et al., 2017). While the limitation of these methods emerges from sequence homolog availability
507 of a protein in multiple sequence alignment (MSA), RaptorX uses a deep neural network
508 leveraging joint family approach, combining multiple ortholog protein families sharing similar
509 function and phylogeny, to infer possible contacts. This method is proven to produce high accuracy
510 in contact prediction compared to others (Wang et al., 2017). However, for a given MSA
511 containing enough homolog sequences other methods are also strong in predicting spatial contacts.
512 EVcouplings approach uses Direct Information (DI) to calculate coevolutionary couplings. DI

513 metric is a modified mutual information (MI) score considering consistency between pairwise
514 probabilities and single amino acid frequencies (de Juan et al., 2013; Morcos et al., 2011).
515 Nonetheless, MI, a global approach compared to local DI metric, is accurate in capturing true
516 contacts, while entangling indirect contacts from direct contacts. MISTIC web server has taken
517 advantage of MI to calculate coevolutionary couplings (Dunn et al., 2008; Gouveia-Oliveira and
518 Pedersen, 2007; Simonetti et al., 2013). In their MI method they introduced a correction term to
519 MI to surpass the low statistics gathered with an MSA containing limited number of sequences.
520 This approach is very useful in cases where certain homologs are rare and MSA of these homologs
521 have multiple gaps in their alignments. All of these methods are employed in this study to achieve
522 high accuracy predictions in finding residue couplings.
523
524 **Mutant proteins cloning, expression and purification**
525
526 The genes for mutants (I34Y, I34K, and I34L) were generated by applying mutagenic
527 primers to P51G-m4-gene sequence and amplifying by PCR. The constructs were subsequently
528 cloned in pET26B vector between NdeI and XhoI sites and transformed in BL21(DE3) for
529 expression and purification. The proteins were expressed from a 10 ml starter culture in LB broth
530 overnight at 37 °C, inoculated into 1 L LB medium. The culture was induced with 1 mM isopropyl
531 thiogalactoside (IPTG) when OD reached 0.6 and grown for another 6-8 hours. The cells were
532 harvested by centrifugation, lysed in 6M Guanidine hydrochloride at pH 8.0, and sonicated for 10
533 minutes. The supernatant recovered after centrifugation was used to purify proteins with GE
534 HisTrap HP column (GE Healthcare Bio-Sciences, Piscataway, NJ) and a Bio-Rad EconoPump
535 (Bio-Rad, Richmond, CA) under denaturing conditions. In brief, the proteins were loaded on the

536 column in Gu-HCl buffer, which was buffer exchanged by 8M Urea buffer. The nonspecific
537 proteins were washed out by 4M urea and 20mM imidazole buffer, pH 8.0 and eluted with 2M
538 Urea and 200mM imidazole, pH 8.0 buffer before putting it for overnight dialysis against 10 mM
539 Tris pH 8.0 and 100 mM NaCl buffer. The buffer was changed once during the night. The refolded
540 protein was concentrated and re-purified to isolate the monomeric species by size exclusion
541 chromatography using Sephadex 75 10/300 column on Agilent's Infinity 1260 system. The gel
542 filtered protein was finally used for all the experiments.
543
544 **CD spectroscopy and T-melts**
545
546 In CV-N family proteins, thermodynamic parameters like free energy of unfolding,
547 enthalpy and entropy cannot be extracted by thermal denaturation because the transition from
548 folded to unfolded state is non-reversible (Patsalo et al., 2011), therefore melting temperatures are
549 used. Far-UV CD spectra were recorded on a Jasco J-815 spectropolarimeter equipped with a
550 thermostatic cell holder, PTC 424S. Spectra were measured from 250 to 200 nm, using a scanning
551 speed of 50 nm/min and a data pitch of 1.0 nm at 25 °C. Samples concentration was approximately
552 15 μM in 10 mM Tris, pH 8.0 and 100 mM NaCl. For thermal denaturation experiments, the
553 melting profile was monitored at 202 nm from 25 °C to 90 °C. The data points were plotted and
554 fitted in Origin8.5 software to get apparent Tm.
555
556
557
558

559 **Isothermal titration calorimetry (ITC)**

560

561 ITC was performed at the Sanford-Burnham Medical Research Institute Protein Analysis
562 Facility using ITC200 calorimeter from Microcal (Northampton, MA) at 23 ºC. 2.0 µl aliquots of
563 solution containing between 3mM and 10 mM Man2 were injected into the cell containing between
564 0.057, and 0.11 mM protein. 19 of 2.0 µl injections were made. The experiments were performed
565 in 10mM Tris, 100mM NaCl, pH 8.0 buffer. ITC data were analyzed using Origin software
566 provided by Microcal.

567

568 **Chemical denaturation experiments**

569

570 Chemical denaturation experiments were done by monitoring the shift in the intrinsic
571 tryptophan fluorescence on Cary Eclipse instrument (Varian). 10 µM of protein samples were
572 incubated with increasing concentrations of Guanidine hydrochloride in the range of 0-6 M in 50
573 mM Tris pH8.0 buffer for 72 hours at 25 °C. The emission spectra for the same were recorded by
574 keeping the excitation wavelength at 295nm and bandwidth of 1 nm. A ratio of fluorescence at
575 330 and 360 nm ($I_{330/360}$) was plotted at respective Gu-HCl concentrations and the data points were
576 fit to following sigmoidal equation to obtain C_m.

577 $$y = \mathrm{A2} + \frac{\mathrm{A1} - \mathrm{A2}}{1 + \mathrm{e}^{(\mathrm{x}-\mathrm{x0})/\mathrm{dx}}}$$

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

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

583 $$f_U = (y_F - y_{obs})/(y_F - y_U)$$

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

587 $$K = f_U / f_F$$

588 $$K = f_U / 1 - f_F$$

589 $$\Delta G = -RT\ln K$$

590 Where R is the gas constant whose value is 1.987 cal/mol.K and T is the temperature of incubation,
591 which was 25 °C. The value of ΔG is linear over a limited range of Gu-HCl. The linear fit over
592 that range was extrapolated to obtain ΔG_{H2O}.

593

594 **Crystallization and structure determination**

595

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

604 acquired was indexed using XDS and scaled by the aimless package from CCP4i program suite.
605 The structural coordinates and phase were determined by molecular replacement using 2RDK PDB
606 code. The structure of I34Y of CV-N is deposited under PDB accession code 6X7H. The structure
607 was further refined in Coot.
608

609 **Acknowledgements**

610 S.B.O. acknowledges support from the Gordon and Betty Moore Foundations and National
611 Science Foundation (Award: 1715591 and 1901709). G.G. and S.B.O. is grateful for the grant
612 1R21CA207832-01.
613

614 **Competing interests**

615 No competing interests declared.
616

617 **Supplementary Data**

618

Table S1. DFI, DCI, RaptorX, Evcoupling, and MISTIC metrics are used to identify critical residues both rigid, distal from binding site, allosterically coupled and coevolved with binding sites (Supplementary file 1 table 1). The analyses pointed out four residues matching the rigidity (DFI score less than 0.2), allosteric coupling (DCI score higher than 0.7), coevolution (RAPTORX, MISTIC, and EVCOUPLING all higher than 0.6) and distance (Minimum distance to binding site higher than 10Å) criteria. Only residue I34 out of four residues satisfy a far

distance measure (Both minimum and average distance to binding site higher than 15Å). We selected residue I34 for further analyses.

Residue	DFI Score	DCI Score	Distance from Binding Sites (Å)		RAPTORX Score	MISTIC Score	EVCOUPLING Score
			Minimum	Average			
34	0.17	0.71	16	22	0.65	0.93	0.71
61	0.15	0.83	11	15	0.81	0.76	0.86
71	0.01	0.93	11	13	0.94	0.92	0.98
84	0.12	0.89	11	14	0.87	0.86	0.83

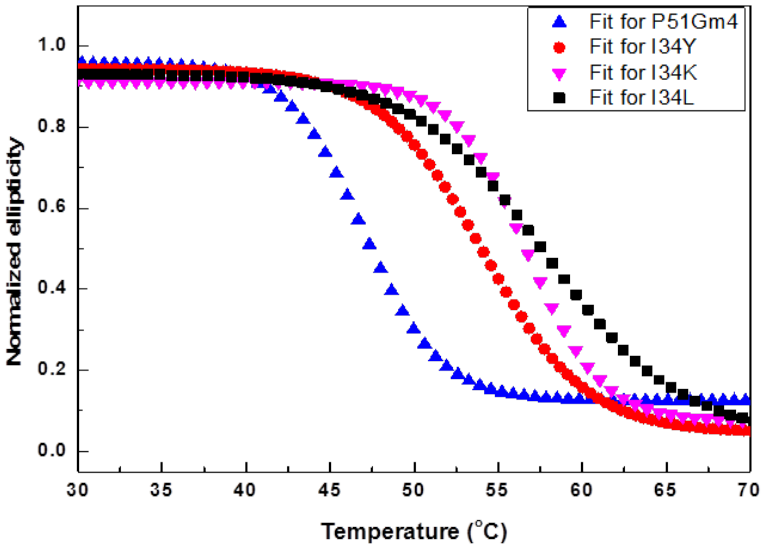

Fig S1 Fits for thermal melts of the CV-N mutants

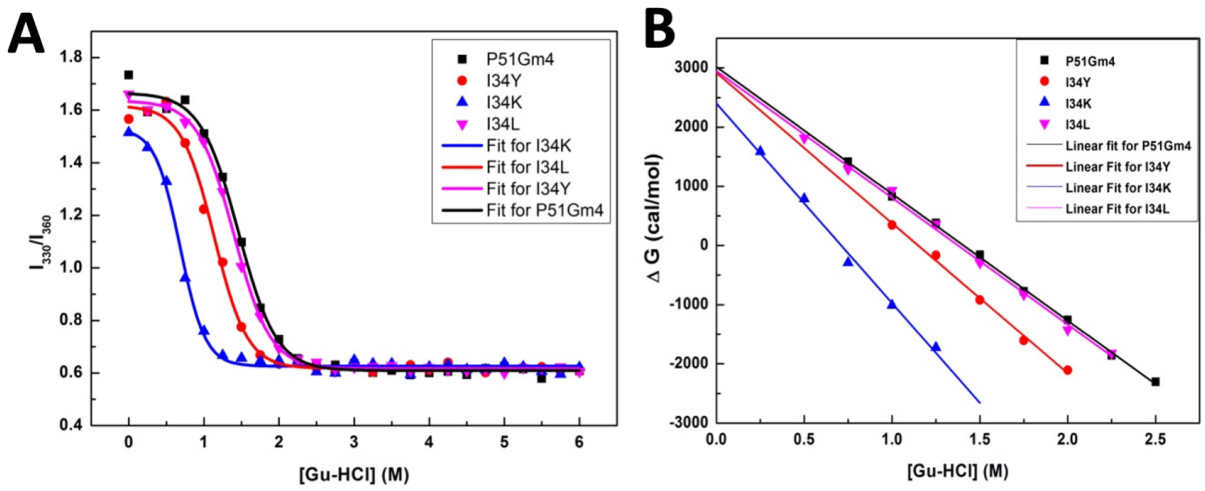

Fig S2 A) Chemical denaturation curve showing I_{330}/ I_{360} ratio as a function of Gu-HCl concentration, b) ΔG_{H2O} versus Gu-HCl concentration plot for CV-N mutants.

621

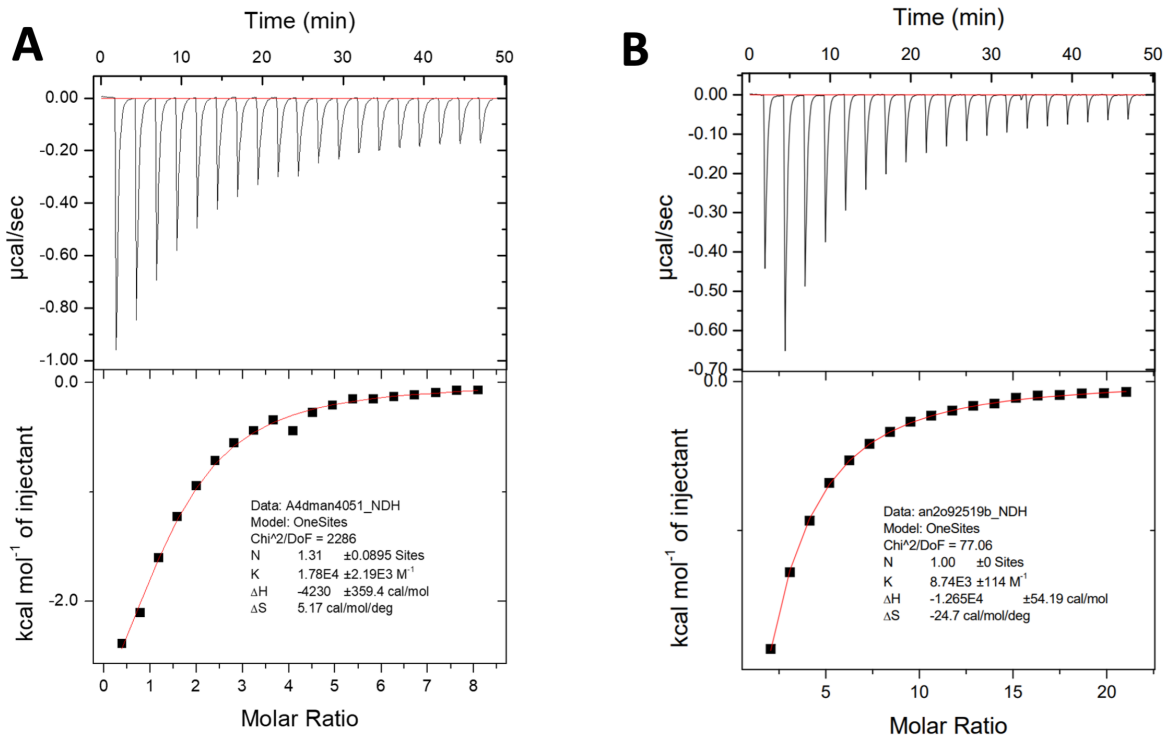

Fig S3 Binding isotherms of CV-N mutants upon titration with dimannose: A) I34Y and B) P51G-m4

622

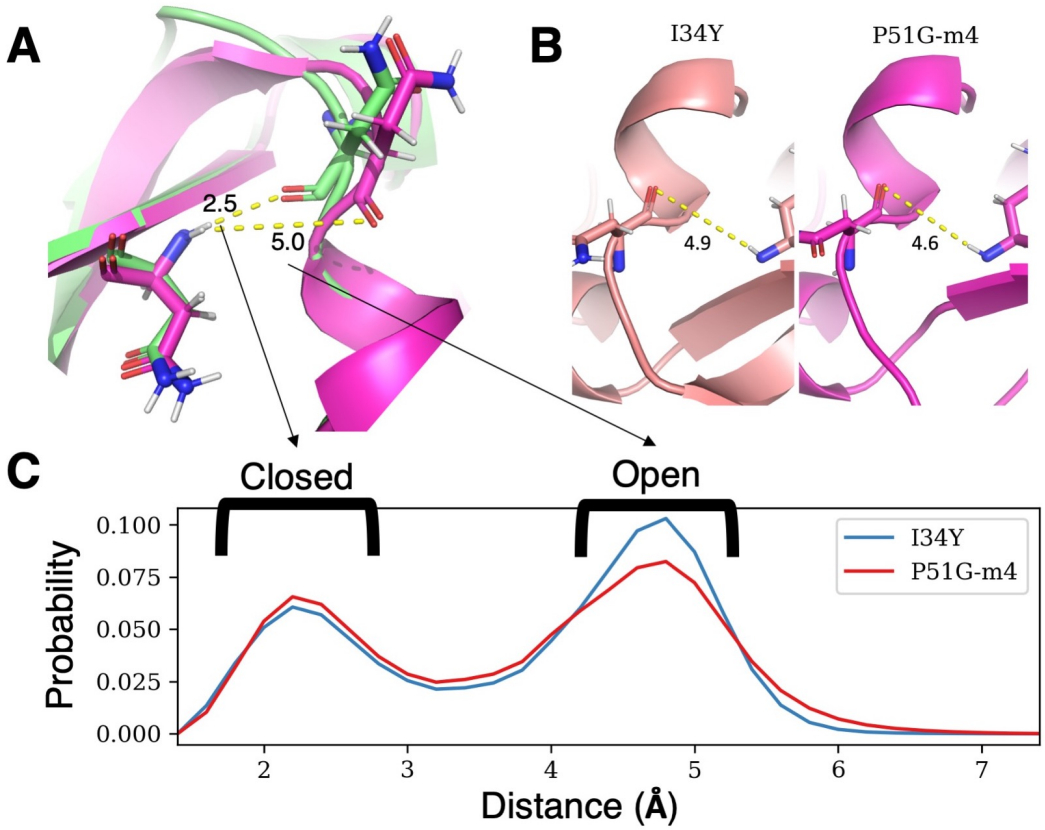

Fig S4 A) Structural difference of open vs closed conformation based on the hydrogen bond distance between residue N42 and N53 B) Hydrogen bond distance between residue N42 and N53 from crystal structures of P51G-m4, and P51G-m4-I34Y C) Frequencies of hydrogen bond distance between residue N42 and N53 from GROMACS production runs showing I34Y variant sampling more open conformation compared to P51G-m4.

623

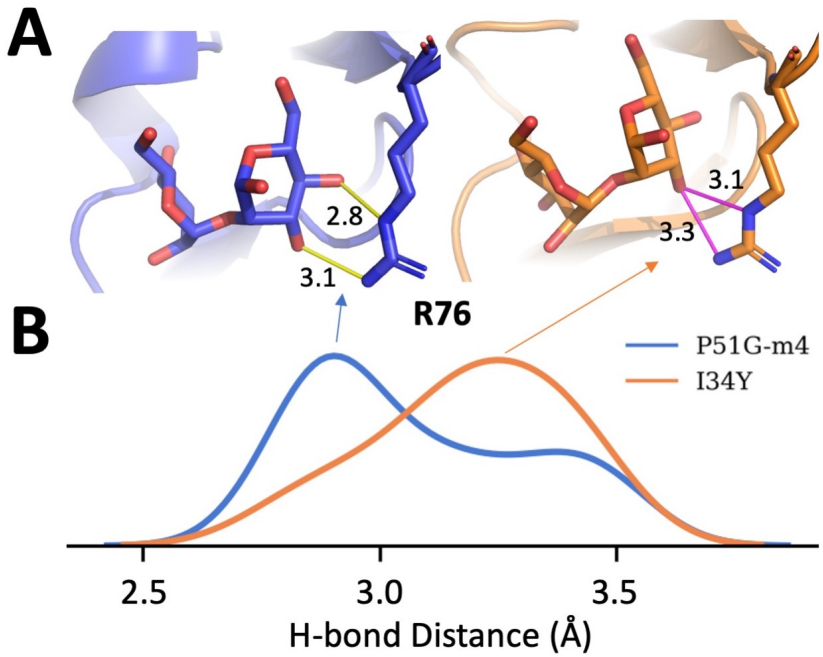

Fig S5 We sampled 2000 different conformations from MD simulations for P51G-m4 CV-N and I34Y mutant and performed dimannose docking to obtained docked poses and then analyzed hydrogen bond patterns A) Hydrogen bonds (representing the peak of the distribution on panel B) and their distances are shown between dimannose and residue R76 for P51G-m4 (blue) and I34Y (orange) B) H-bond distance distribution between dimannose and residue R76.

624

625

626

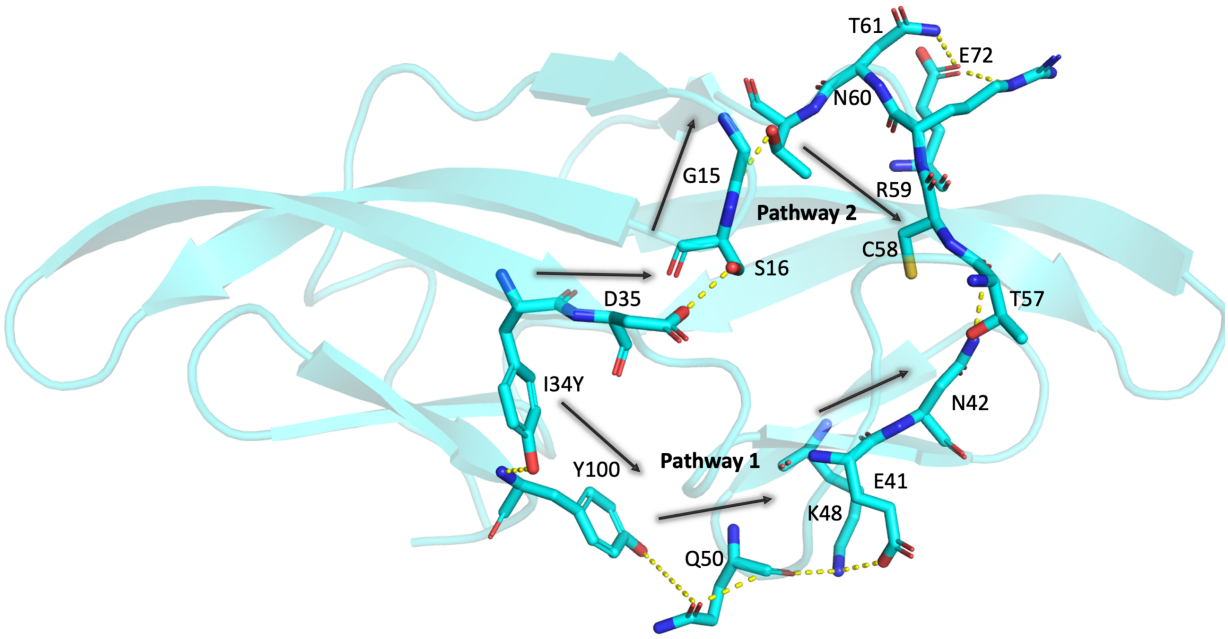

Fig S6 Network of hydrogen bond interactions connecting residue location 34 to T57 is investigated in I34Y variant and P51G-m4 CV-N. Two Hydrogen bond pathways are found connecting residue 34 to 57. Pathway 1 is unique to I34Y. Pathway 2 is also observed in P51G-m4 CV-N but sampled much more frequently in I34Y variant.

Additional files

Supplementary file 1

This file contains Supplementary Tables 1 and 2 (referenced in the main text).

Supplementary Table 1 contains the complete DFI, DCI, RaptorX, Evcoupling, and MISTIC metric data used in this study. Supplementary Table 2 contains the sequence variations with

636 binding residues from MSA used in this study to identify possible amino acid variants on
637 location I34. Previous mutational study (Bolia et al., 2014b) on binding residues shown that, N42
638 is highly sequentially conserved as N. Position E41 has two high frequency variants E and G that
639 do not abolish binding. Residue T57 has high frequency variants S and T, and a rare case V.
640 From the amino acid variation data with binding sites, I34 variants L and Y are selected based on
641 their high frequency and K as a rare case.
642

643 **References**

644

645 Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. 2015. GROMACS:
646 High performance molecular simulations through multi-level parallelism from laptops to
647 supercomputers. *SoftwareX* **1–2**:19–25. doi:10.1016/j.softx.2015.06.001
648 Acar B, Rose J, Aykac Fas B, Ben-Tal N, Lewinson O, Haliloglu T. 2020. Distinct Allosteric
649 Networks Underlie Mechanistic Speciation of ABC Transporters. *Structure* **28**:651-
650 663.e5. doi:10.1016/j.str.2020.03.014
651 Anishchenko I, Ovchinnikov S, Kamisetty H, Baker D. 2017. Origins of coevolution between
652 residues distant in protein 3D structures. *PNAS* **114**:9122–9127.
653 doi:10.1073/pnas.1702664114
654 Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. 2001. Anisotropy of
655 Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophysical Journal*
656 **80**:505–515. doi:10.1016/S0006-3495(01)76033-X

657 Atilgan C, Atilgan AR. 2009. Perturbation-Response Scanning Reveals Ligand Entry-Exit
658 Mechanisms of Ferric Binding Protein. *PLoS Comput Biol* **5**.
659 doi:10.1371/journal.pcbi.1000544
660 Atilgan C, Gerek ZN, Ozkan SB, Atilgan AR. 2010. Manipulation of Conformational Change in
661 Proteins by Single-Residue Perturbations. *Biophysical Journal* **99**:933–943.
662 doi:10.1016/j.bpj.2010.05.020
663 Balzarini J. 2007. Targeting the glycans of glycoproteins: a novel paradigm for antiviral therapy.
664 *Nature Reviews Microbiology* **5**:583–597. doi:10.1038/nrmicro1707
665 Barrientos LG, Matei E, Lasala F, Delgado R, Gronenborn AM. 2006. Dissecting carbohydrate–
666 Cyanovirin-N binding by structure-guided mutagenesis: functional implications for viral
667 entry inhibition. *Protein Eng Des Sel* **19**:525–535. doi:10.1093/protein/gzl040
668 Barrientos LG, O'Keefe BR, Bray M, Sanchez A, Gronenborn AM, Boyd MR. 2003.
669 Cyanovirin-N binds to the viral surface glycoprotein, GP1,2 and inhibits infectivity of
670 Ebola virus. *Antiviral Research* **58**:47–56. doi:10.1016/S0166-3542(02)00183-3
671 Bewley CA. 2001. Solution Structure of a Cyanovirin-N:Manα1-2Manα Complex: Structural
672 Basis for High-Affinity Carbohydrate-Mediated Binding to gp120. *Structure* **9**:931–940.
673 doi:10.1016/S0969-2126(01)00653-0
674 Bolia A, Gerek ZN, Ozkan SB. 2014a. BP-Dock: A Flexible Docking Scheme for Exploring
675 Protein–Ligand Interactions Based on Unbound Structures. *J Chem Inf Model* **54**:913–
676 925. doi:10.1021/ci4004927
677 Bolia A, Ozkan SB. 2016. Adaptive BP-Dock: An Induced Fit Docking Approach for Full
678 Receptor Flexibility. *J Chem Inf Model* **56**:734–746. doi:10.1021/acs.jcim.5b00587

679 Bolia A, Woodrum BW, Cereda A, Ruben MA, Wang X, Ozkan SB, Ghirlanda G. 2014b. A
680 Flexible Docking Scheme Efficiently Captures the Energetics of Glycan-Cyanovirin
681 Binding. *Biophysical Journal* **106**:1142–1151. doi:10.1016/j.bpj.2014.01.040
682 Bolmstedt AJ, O'Keefe BR, Shenoy SR, McMahon JB, Boyd MR. 2001. Cyanovirin-N Defines
683 a New Class of Antiviral Agent Targeting N-Linked, High-Mannose Glycans in an
684 Oligosaccharide-Specific Manner. *Mol Pharmacol* **59**:949–954.
685 doi:10.1124/mol.59.5.949
686 Botos I, Wlodawer A. 2005. Proteins that bind high-mannose sugars of the HIV envelope.
687 *Progress in Biophysics and Molecular Biology*, Structure-guided design of AIDs
688 Antivirals **88**:233–282. doi:10.1016/j.pbiomolbio.2004.05.001
689 Botos I, Wlodawer A. 2003. Cyanovirin-N: a sugar-binding antiviral protein with a new twist.
690 *CMLS, Cell Mol Life Sci* **60**:277–287. doi:10.1007/s000180300023
691 Butler BM, Gerek ZN, Kumar S, Ozkan SB. 2015. Conformational dynamics of nonsynonymous
692 variants at protein interfaces reveals disease association. *Proteins: Structure, Function,*
693 *and Bioinformatics* **83**:428–435. doi:10.1002/prot.24748
694 Campbell E, Kaltenbach M, Correy GJ, Carr PD, Porebski BT, Livingstone EK, Afriat-Jurnou L,
695 Buckle AM, Weik M, Hollfelder F, Tokuriki N, Jackson CJ. 2016. The role of protein
696 dynamics in the evolution of new enzyme function. *Nature Chemical Biology* **12**:944–
697 950. doi:10.1038/nchembio.2175
698 Campbell EC, Correy GJ, Mabbitt PD, Buckle AM, Tokuriki N, Jackson CJ. 2018. Laboratory
699 evolution of protein conformational dynamics. *Current Opinion in Structural Biology*,
700 Carbohydrates • Sequences and topology **50**:49–57. doi:10.1016/j.sbi.2017.09.005

701 Campitelli P, Guo J, Zhou H-X, Ozkan SB. 2018. Hinge-Shift Mechanism Modulates Allosteric
702 Regulations in Human Pin1. *J Phys Chem B* **122**:5623–5629.
703 doi:10.1021/acs.jpcb.7b11971
704 Campitelli P, Modi T, Kumar S, Ozkan SB. 2020a. The Role of Conformational Dynamics and
705 Allostery in Modulating Protein Evolution. *Annual Review of Biophysics* **49**:267–288.
706 doi:10.1146/annurev-biophys-052118-115517
707 Campitelli P, Ozkan SB, Swint-Kruse L. 2020b. Asymmetry in Dynamic Allosteric Residue
708 Coupling (DARC) Interactions Captures Evolutionary Landscape. *Biophysical Journal*
709 **118**:52a. doi:10.1016/j.bpj.2019.11.464
710 Chang LC, Bewley CA. 2002. Potent Inhibition of HIV-1 Fusion by Cyanovirin-N Requires
711 Only a Single High Affinity Carbohydrate Binding Site: Characterization of Low Affinity
712 Carbohydrate Binding Site Knockout Mutants. *Journal of Molecular Biology* **318**:1–8.
713 doi:10.1016/S0022-2836(02)00045-1
714 Davis IW, Baker D. 2009. RosettaLigand Docking with Full Ligand and Receptor Flexibility.
715 *Journal of Molecular Biology* **385**:381–392. doi:10.1016/j.jmb.2008.11.010
716 Davis IW, Raha K, Head MS, Baker D. 2009. Blind docking of pharmaceutically relevant
717 compounds using RosettaLigand. *Protein Science* **18**:1998–2002. doi:10.1002/pro.192
718 de Juan D, Pazos F, Valencia A. 2013. Emerging methods in protein co-evolution. *Nature*
719 *Reviews Genetics* **14**:249–261. doi:10.1038/nrg3414
720 DeLuca S, Khar K, Meiler J. 2015. Fully Flexible Docking of Medium Sized Ligand Libraries
721 with RosettaLigand. *PLoS One* **10**. doi:10.1371/journal.pone.0132508

722 Dunn SD, Wahl LM, Gloor GB. 2008. Mutual information without the influence of phylogeny or
723 entropy dramatically improves residue contact prediction. *Bioinformatics* **24**:333–340.
724 doi:10.1093/bioinformatics/btm604
725 Fromme R, Katiliene Z, Fromme P, Ghirlanda G. 2008. Conformational gating of dimannose
726 binding to the antiviral protein cyanovirin revealed from the crystal structure at 1.35 Å
727 resolution. *Protein Science* **17**:939–944. doi:10.1110/ps.083472808
728 Fromme R, Katiliene Z, Giomarelli B, Bogani F, Mc Mahon J, Mori T, Fromme P, Ghirlanda G.
729 2007. A Monovalent Mutant of Cyanovirin-N Provides Insight into the Role of Multiple
730 Interactions with gp120 for Antiviral Activity,. *Biochemistry* **46**:9199–9207.
731 doi:10.1021/bi700666m
732 Fujimoto YK, Green DF. 2012. Carbohydrate Recognition by the Antiviral Lectin Cyanovirin-N.
733 *J Am Chem Soc* **134**:19639–19651. doi:10.1021/ja305755b
734 Gerek ZN, Kumar S, Ozkan SB. 2013. Structural dynamics flexibility informs function and
735 evolution at a proteome scale. *Evolutionary Applications* **6**:423–433.
736 doi:10.1111/eva.12052
737 Gerek ZN, Ozkan SB. 2011. Change in Allosteric Network Affects Binding Affinities of PDZ
738 Domains: Analysis through Perturbation Response Scanning. *PLOS Computational*
739 *Biology* **7**:e1002154. doi:10.1371/journal.pcbi.1002154
740 Gouveia-Oliveira R, Pedersen AG. 2007. Finding coevolving amino acid residues using row and
741 column weighting of mutual information and multi-dimensional amino acid
742 representation. *Algorithms for Molecular Biology* **2**:12. doi:10.1186/1748-7188-2-12
743 Gunasekaran K, Ma B, Nussinov R. 2004. Is allostery an intrinsic property of all dynamic
744 proteins? *Proteins: Structure, Function, and Bioinformatics* **57**:433–443.

745 Hopf TA, Green AG, Schubert B, Mersmann S, Schärfe CPI, Ingraham JB, Toth-Petroczy A,
746 Brock K, Riesselman AJ, Palmedo P, Kang C, Sheridan R, Draizen EJ, Dallago C,
747 Sander C, Marks DS. 2019. The EVcouplings Python framework for coevolutionary
748 sequence analysis. *Bioinformatics* **35**:1582–1584. doi:10.1093/bioinformatics/bty862
749 Hopf TA, Schärfe CPI, Rodrigues JPGLM, Green AG, Kohlbacher O, Sander C, Bonvin AMJJ,
750 Marks DS. 2018. Sequence co-evolution gives 3D contacts and structures of protein
751 complexes. *eLife* **3**. doi:10.7554/eLife.03430
752 Ikeguchi M, Ueno J, Sato M, Kidera A. 2005. Protein Structural Change Upon Ligand Binding:
753 Linear Response Theory. *Phys Rev Lett* **94**:078102. doi:10.1103/PhysRevLett.94.078102
754 Jana B, Morcos F, Onuchic JN. 2014. From structure to function: the convergence of structure
755 based models and co-evolutionary information. *Phys Chem Chem Phys* **16**:6496–6507.
756 doi:10.1039/C3CP55275F
757 Jones DT, Buchan DWA, Cozzetto D, Pontil M. 2012. PSICOV: precise structural contact
758 prediction using sparse inverse covariance estimation on large multiple sequence
759 alignments. *Bioinformatics* **28**:184–190. doi:10.1093/bioinformatics/btr638
760 Kamisetty H, Ovchinnikov S, Baker D. 2013. Assessing the utility of coevolution-based residue–
761 residue contact predictions in a sequence- and structure-rich era. *PNAS* **110**:15674–
762 15679. doi:10.1073/pnas.1314045110
763 Keeffe JR, Gnanapragasam PNP, Gillespie SK, Yong J, Bjorkman PJ, Mayo SL. 2011. Designed
764 oligomers of cyanovirin-N show enhanced HIV neutralization. *PNAS* **108**:14079–14084.
765 doi:10.1073/pnas.1108777108

766 Kelley BS, Chang LC, Bewley CA. 2002. Engineering an Obligate Domain-Swapped Dimer of
767 Cyanovirin-N with Enhanced Anti-HIV Activity. *J Am Chem Soc* **124**:3210–3211.
768 doi:10.1021/ja025537m
769 Koharudin LMI, Furey W, Gronenborn AM. 2009. A designed chimeric cyanovirin-N homolog
770 lectin: Structure and molecular basis of sucrose binding. *Proteins: Structure, Function,*
771 *and Bioinformatics* **77**:904–915. doi:10.1002/prot.22514
772 Koharudin LMI, Gronenborn AM. 2013. Sweet entanglements—protein: Glycan interactions in
773 two HIV-inactivating lectin families. *Biopolymers* **99**:196–202. doi:10.1002/bip.22106
774 Koharudin LMI, Viscomi AR, Jee J-G, Ottonello S, Gronenborn AM. 2008. The Evolutionarily
775 Conserved Family of Cyanovirin-N Homologs: Structures and Carbohydrate Specificity.
776 *Structure* **16**:570–584. doi:10.1016/j.str.2008.01.015
777 Kryshtafovych A, Schwede T, Topf M, Fidelis K, Moult J. 2019. Critical assessment of methods
778 of protein structure prediction (CASP)—Round XIII. *Proteins: Structure, Function, and*
779 *Bioinformatics* **87**:1011–1020. doi:10.1002/prot.25823
780 Kumar A, Butler BM, Kumar S, Ozkan SB. 2015a. Integration of structural dynamics and
781 molecular evolution via protein interaction networks: a new era in genomic medicine.
782 *Current Opinion in Structural Biology*, Catalysis and regulation • Protein-protein
783 interactions **35**:135–142. doi:10.1016/j.sbi.2015.11.002
784 Kumar A, Glembo TJ, Ozkan SB. 2015b. The Role of Conformational Dynamics and Allostery
785 in the Disease Development of Human Ferritin. *Biophysical Journal* **109**:1273–1281.
786 doi:10.1016/j.bpj.2015.06.060
787 Kuriyan J, Eisenberg D. 2007. The origin of protein interactions and allostery in colocalization.
788 *Nature* **450**:983–990. doi:10.1038/nature06524

789 Larrimore KE, Kazan IC, Kannan L, Kendle RP, Jamal T, Barcus M, Bolia A, Brimijoin S, Zhan
790 C-G, Ozkan SB. 2017. Plant-expressed cocaine hydrolase variants of
791 butyrylcholinesterase exhibit altered allosteric effects of cholinesterase activity and
792 increased inhibitor sensitivity. *Scientific reports* **7**:10419.
793 Li Z, Bolia A, Maxwell JD, Bobkov AA, Ghirlanda G, Ozkan SB, Margulis CJ. 2015. A Rigid
794 Hinge Region Is Necessary for High-Affinity Binding of Dimannose to Cyanovirin and
795 Associated Constructs. *Biochemistry* **54**:6951–6960. doi:10.1021/acs.biochem.5b00635
796 Liu J, Nussinov R. 2017. Energetic redistribution in allostery to execute protein function. *Proc*
797 *Natl Acad Sci U S A* **114**:7480–7482. doi:10.1073/pnas.1709071114
798 Lu H-M, Liang J. 2009. Perturbation-based Markovian transmission model for probing allosteric
799 dynamics of large macromolecular assembling: a study of GroEL-GroES. *PLoS Comput*
800 *Biol* **5**:e1000526. doi:10.1371/journal.pcbi.1000526
801 Ma J, Wang S, Wang Z, Xu J. 2015. Protein contact prediction by integrating joint evolutionary
802 coupling analysis and supervised learning. *Bioinformatics* **31**:3506–3513.
803 doi:10.1093/bioinformatics/btv472
804 Marks DS, Hopf TA, Sander C. 2012. Protein structure prediction from sequence variation.
805 *Nature Biotechnology* **30**:1072–1080. doi:10.1038/nbt.2419
806 Matei E, Basu R, Furey W, Shi J, Calnan C, Aiken C, Gronenborn AM. 2016. Structure and
807 Glycan Binding of a New Cyanovirin-N Homolog. *J Biol Chem* **291**:18967–18976.
808 doi:10.1074/jbc.M116.740415
809 Matei E, Furey W, Gronenborn AM. 2008. Solution and Crystal Structures of a Sugar Binding
810 Site Mutant of Cyanovirin-N: No Evidence of Domain Swapping. *Structure* **16**:1183–
811 1194. doi:10.1016/j.str.2008.05.011

812 Matei E, Louis JM, Jee J, Gronenborn AM. 2011. NMR solution structure of a cyanovirin
813 homolog from wheat head blight fungus. *Proteins: Structure, Function, and*
814 *Bioinformatics* **79**:1538–1549. doi:10.1002/prot.22981
815 Meiler J, Baker D. 2006. ROSETTALIGAND: Protein–small molecule docking with full side-
816 chain flexibility. *Proteins: Structure, Function, and Bioinformatics* **65**:538–548.
817 doi:10.1002/prot.21086
818 Modi T, Campitelli P, Kazan IC, Ozkan SB. 2021. Protein folding stability and binding
819 interactions through the lens of evolution: a dynamical perspective. *Current Opinion in*
820 *Structural Biology* **66**:207–215. doi:10.1016/j.sbi.2020.11.007
821 Modi T, Ozkan SB. 2018. Mutations Utilize Dynamic Allostery to Confer Resistance in TEM-1
822 β-lactamase. *Int J Mol Sci* **19**. doi:10.3390/ijms19123808
823 Morcos F. 2020. Protein conformations à la carte, a step further in de novo protein design. *PNAS*
824 **117**:8674–8676. doi:10.1073/pnas.2004188117
825 Morcos F, Jana B, Hwa T, Onuchic JN. 2013. Coevolutionary signals across protein lineages
826 help capture multiple protein conformations. *PNAS* **110**:20533–20538.
827 doi:10.1073/pnas.1315625110
828 Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, Sander C, Zecchina R, Onuchic JN, Hwa
829 T, Weigt M. 2011. Direct-coupling analysis of residue coevolution captures native
830 contacts across many protein families. *PNAS* **108**:E1293–E1301.
831 doi:10.1073/pnas.1111471108
832 Morcos F, Schafer NP, Cheng RR, Onuchic JN, Wolynes PG. 2014. Coevolutionary information,
833 protein folding landscapes, and the thermodynamics of natural selection. *PNAS*
834 **111**:12408–12413. doi:10.1073/pnas.1413575111

835 Mori T, Boyd MR. 2001. Cyanovirin-N, a Potent Human Immunodeficiency Virus-Inactivating
836 Protein, Blocks both CD4-Dependent and CD4-Independent Binding of Soluble gp120
837 (sgp120) to Target Cells, Inhibits sCD4-Induced Binding of sgp120 to Cell-Associated
838 CXCR4, and Dissociates Bound sgp120 from Target Cells. *Antimicrobial Agents and*
839 *Chemotherapy* **45**:664–672. doi:10.1128/AAC.45.3.664-672.2001
840 O'Keefe BR, Smee DF, Turpin JA, Saucedo CJ, Gustafson KR, Mori T, Blakeslee D, Buckheit
841 R, Boyd MR. 2003. Potent Anti-Influenza Activity of Cyanovirin-N and Interactions with
842 Viral Hemagglutinin. *Antimicrobial Agents and Chemotherapy* **47**:2518–2525.
843 doi:10.1128/AAC.47.8.2518-2525.2003
844 Ose N, Butler BM, Kumar A, Ozkan SB, Kumar S. 2020. Dynamic Allosteric Residue Coupling
845 Reveals Disease Mechanism for Gaucher Disease and NSNVS Across the Proteome.
846 *Biophysical Journal* **118**:53a. doi:10.1016/j.bpj.2019.11.472
847 Otten R, Liu L, Kenner LR, Clarkson MW, Mavor D, Tawfik DS, Kern D, Fraser JS. 2018.
848 Rescue of conformational dynamics in enzyme catalysis by directed evolution. *Nature*
849 *Communications* **9**:1314. doi:10.1038/s41467-018-03562-9
850 Patsalo V, Raleigh DP, Green DF. 2011. Rational and Computational Design of Stabilized
851 Variants of Cyanovirin-N That Retain Affinity and Specificity for Glycan Ligands.
852 *Biochemistry* **50**:10698–10712. doi:10.1021/bi201411c
853 Percudani R, Montanini B, Ottonello S. 2005. The anti-HIV cyanovirin-N domain is
854 evolutionarily conserved and occurs as a protein module in eukaryotes. *Proteins:*
855 *Structure, Function, and Bioinformatics* **60**:670–678. doi:10.1002/prot.20543

856 Qi X, Yang Y, Su Y, Wang T. 2009. Molecular Cloning and Sequence Analysis of Cyanovirin-N
857 Homology Gene in Ceratopteris thalictroides. *amfj* **99**:78–92. doi:10.1640/0002-8444-
858 99.2.78
859 Ramadugu SK, Li Z, Kashyap HK, Margulis CJ. 2014. The Role of Glu41 in the Binding of
860 Dimannose to P51G-m4-CVN. *Biochemistry* **53**:1477–1484. doi:10.1021/bi4014159
861 Risso VA, Sanchez-Ruiz JM, Ozkan SB. 2018. Biotechnological and protein-engineering
862 implications of ancestral protein resurrection. *Current Opinion in Structural Biology*,
863 Engineering and design: New applications • Membranes **51**:106–115.
864 doi:10.1016/j.sbi.2018.02.007
865 Romero PA, Arnold FH. 2009. Exploring protein fitness landscapes by directed evolution.
866 *Nature Reviews Molecular Cell Biology* **10**:866–876. doi:10.1038/nrm2805
867 Salinas VH, Ranganathan R. 2018. Coevolution-based inference of amino acid interactions
868 underlying protein function. *eLife* **7**:e34300. doi:10.7554/eLife.34300
869 Simonetti FL, Teppa E, Chernomoretz A, Nielsen M, Marino Buslje C. 2013. MISTIC: mutual
870 information server to infer coevolution. *Nucleic Acids Res* **41**:W8–W14.
871 doi:10.1093/nar/gkt427
872 Spoel DVD, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. 2005. GROMACS:
873 Fast, flexible, and free. *Journal of Computational Chemistry* **26**:1701–1718.
874 doi:10.1002/jcc.20291
875 Tokuriki N, Jackson CJ, Afriat-Jurnou L, Wyganowski KT, Tang R, Tawfik DS. 2012.
876 Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme.
877 *Nature Communications* **3**:1257. doi:10.1038/ncomms2246

878 Tripathi S, Waxham MN, Cheung MS, Liu Y. 2015. Lessons in Protein Design from Combined
879 Evolution and Conformational Dynamics. *Scientific Reports* **5**:14259.
880 doi:10.1038/srep14259
881 Vorontsov II, Miyashita O. 2009. Solution and Crystal Molecular Dynamics Simulation Study of
882 m4-Cyanovirin-N Mutants Complexed with Di-Mannose. *Biophysical Journal* **97**:2532–
883 2540. doi:10.1016/j.bpj.2009.08.011
884 Wagner JR, Sørensen J, Hensley N, Wong C, Zhu C, Perison T, Amaro RE. 2017. POVME 3.0:
885 Software for Mapping Binding Pocket Flexibility. *J Chem Theory Comput* **13**:4584–
886 4592. doi:10.1021/acs.jctc.7b00500
887 Wang R, Lai L, Wang S. 2002. Further development and validation of empirical scoring
888 functions for structure-based binding affinity prediction. *J Comput Aided Mol Des* **16**:11–
889 26. doi:10.1023/A:1016357811882
890 Wang S, Li W, Zhang R, Liu S, Xu J. 2016. CoinFold: a web server for protein contact
891 prediction and contact-assisted protein folding. *Nucleic Acids Res* **44**:W361–W366.
892 doi:10.1093/nar/gkw307
893 Wang S, Sun S, Li Z, Zhang R, Xu J. 2017. Accurate De Novo Prediction of Protein Contact
894 Map by Ultra-Deep Learning Model. *PLOS Computational Biology* **13**:e1005324.
895 doi:10.1371/journal.pcbi.1005324
896 Wodak SJ, Paci E, Dokholyan NV, Berezovsky IN, Horovitz A, Li J, Hilser VJ, Bahar I,
897 Karanicolas J, Stock G, Hamm P, Stote RH, Eberhardt J, Chebaro Y, Dejaegere A,
898 Cecchini M, Changeux J-P, Bolhuis PG, Vreede J, Faccioli P, Orioli S, Ravasio R, Yan
899 L, Brito C, Wyart M, Gkeka P, Rivalta I, Palermo G, McCammon JA, Panecka-Hofman
900 J, Wade RC, Di Pizio A, Niv MY, Nussinov R, Tsai C-J, Jang H, Padhorny D, Kozakov

901 D, McLeish T. 2019. Allostery in Its Many Disguises: From Theory to Applications.
902 *Structure* **27**:566–578. doi:10.1016/j.str.2019.01.003
903 Woodrum BW, Maxwell JD, Bolia A, Ozkan SB, Ghirlanda G. 2013. The antiviral lectin
904 cyanovirin-N: probing multivalency and glycan recognition through experimental and
905 computational approaches. *Biochem Soc Trans* **41**:1170–1176.
906 doi:10.1042/BST20130154
907 Yuan L, Kurek I, English J, Keenan R. 2005. Laboratory-Directed Protein Evolution. *Microbiol*
908 *Mol Biol Rev* **69**:373–392. doi:10.1128/MMBR.69.3.373-392.2005
909 Zhang X, Gureasko J, Shen K, Cole PA, Kuriyan J. 2006. An allosteric mechanism for activation
910 of the kinase domain of epidermal growth factor receptor. *Cell* **125**:1137–1149.
911 doi:10.1016/j.cell.2006.05.013
912
913
914
915