

1 **Allosteric Regulatory Control in Dihydrofolate Reductase is Revealed by Dynamic**  
2 **Asymmetry**

3

4 I. Can Kazan<sup>1</sup>, Jeremy H. Mills<sup>2</sup>, S. Banu Ozkan<sup>1</sup> \*

5

6 <sup>1</sup>Center for Biological Physics and Department of Physics, Arizona State University, Tempe, AZ

7 <sup>2</sup>School of Molecular Sciences and The Biodesign Center for Molecular Design and Biomimetics,

8 Arizona State University, Tempe, AZ 85281, USA

9

10 \*Corresponding author:

11 Email: [Banu.Ozkan@asu.edu](mailto:Banu.Ozkan@asu.edu)

12

13 **Abstract**

14 We investigated the relationship between mutations and dynamics in *Escherichia coli*  
15 dihydrofolate reductase (DHFR) using computational methods. Our study focused on the M20 and  
16 FG loops, which are known to be functionally important and affected by mutations distal to the  
17 loops. We used Molecular Dynamics simulations and developed position-specific metrics,  
18 including the Dynamic Flexibility Index (DFI) and Dynamic Coupling Index (DCI), to analyze the  
19 dynamics of wild-type DHFR and compared our results with existing deep mutational scanning  
20 data. Our analysis showed a statistically significant association between DFI and mutational  
21 tolerance of the DHFR positions, indicating that DFI can predict functionally beneficial or  
22 detrimental substitutions. We also applied an asymmetric version of our DCI metric (DCI<sub>asym</sub>) to  
23 DHFR and found that certain distal residues control the dynamics of the M20 and FG loops,

24 whereas others are controlled by them. Residues that are suggested to control the M20 and FG  
25 loops by our DCI<sub>asym</sub> metric are evolutionarily non-conserved; mutations at these sites can enhance  
26 enzyme activity. On the other hand, residues controlled by the loops are mostly deleterious to  
27 function when mutated and are also evolutionary conserved. Our results suggest that dynamics-  
28 based metrics can identify residues that explain the relationship between mutation and protein  
29 function or can be targeted to rationally engineer enzymes with enhanced activity.

30

## 31 **1 Introduction**

32 The human-microbial antibiotic arms race has prompted extensive research efforts aimed  
33 at both developing new drugs and gaining a complete understanding of druggable enzymes.<sup>1-5</sup> One  
34 such enzyme is dihydrofolate reductase (DHFR), which has been investigated for its fundamental  
35 role in 5,6,7,8-tetrahydrofolate (THF) synthesis.<sup>6-10</sup> Due to an abundance of biophysical data<sup>8</sup>,  
36 DHFR from *Escherichia coli* represents an excellent model system for studying the relationship  
37 between protein dynamics and function. The catalytic activity of *E. coli* DHFR has been  
38 extensively studied. One major achievement in these studies was the crystallization of DHFR in  
39 conformations that represent intermediate steps of the enzymatic reaction pathway.<sup>11,12</sup> These  
40 experiments revealed that multiple loops in DHFR are implicated in its function. For example, the  
41 M20 loop (residues 9-24) controls access to the active site and the FG loop (residues 116-132)  
42 stabilizes the M20 loop through hydrogen bonding interactions.<sup>11-13</sup> Mutations on both of these  
43 loops have been reported to severely limit the activity of DHFR<sup>6,14</sup>, whereas positions distal to  
44 these sites can be altered to enhance activity.<sup>6,14-19</sup>

45 The dynamics of *E. coli* DHFR have also been thoroughly studied in an effort to gain insight  
46 into the impact of point mutations on its activity.<sup>20-24</sup> These studies revealed that mutations in

47 DHFR often modulate the enzyme's activity indirectly and at a distance.<sup>20,24</sup> Namely, in a series  
48 of computational and experimental studies, mutations distal to the active site of DHFR were shown  
49 to alter hydrogen bonding interactions and the rotamers of residues close to the active site through  
50 a network of interacting residues.<sup>21-27</sup> The critical role of dynamics in DHFR function has been  
51 studied previously<sup>20-27</sup> but a general relationship between dynamics of each position and their  
52 contribution to the activity has yet to be elucidated.

53 We hypothesize that the position-specific dynamic features of DHFR can shed light on the  
54 diverse impact of mutations on its activity. Therefore, we thoroughly examined the dynamics of  
55 DHFR utilizing three computational metrics: the Dynamic Flexibility Index (DFI)<sup>28-30</sup>, the  
56 Dynamic Coupling Index (DCI)<sup>27,28,31</sup>, and an asymmetric version of DCI, which we call  
57 DCI<sub>asym</sub>.<sup>27,31,32</sup> The DFI metric measures the normalized magnitude of response of a residue to  
58 perturbations applied on all other amino acids; a high DFI value indicates high flexibility,  
59 conversely, a low DFI score suggests that a residue is highly rigid. Our DCI metric reports on the  
60 dynamic coupling between residues.<sup>27,28,30,31,33-37</sup> A high DCI value indicates high dynamic  
61 coupling between residues *i* and *j*, while a low DCI score implies weak coupling between these  
62 residues. Due to the complex conformational dynamics of a protein, the DCI score between two  
63 distal, non-interacting residues is not necessarily symmetric. We therefore developed a new metric  
64 called DCI asymmetry (DCI<sub>asym</sub>), which reports the difference in fluctuation response of residue *i*  
65 when *j* is perturbed versus the response of residue *j* when *i* is perturbed (DCI<sub>ij</sub> – DCI<sub>ji</sub>). DCI<sub>asym</sub>  
66 can therefore be used to assess which of a pair of residues dominates the control of motion between  
67 them.

68 In this study, we applied DFI, DCI, and DCI<sub>asym</sub> to molecular dynamics (MD) simulations  
69 of DHFR. Because the M20 and FG loops of DHFR are highly important for its function<sup>6-9,11</sup>, we

70 used our DCI and DCI<sub>asym</sub> metrics to assess whether the distal regions of DHFR dynamically  
71 modulate these loops. We then compared these analyses to a published deep mutational scanning  
72 dataset<sup>6</sup>, which allowed us to link the activity of DHFR to its dynamics. Our dynamics metrics  
73 provided a link between the previously reported mutational data and collective motions of the  
74 enzyme. In particular, DCI<sub>asym</sub> allowed us to classify a given residue position as “controlled” (i.e.,  
75 dynamically controlled by the loops) if its fluctuation response to a perturbation on M20 & FG  
76 loops is considerably lower than the response of M20 & FG loops when that residue is perturbed.  
77 If the opposite is true, we classified that residue as “controller” (i.e., the residue is dynamically  
78 controlling the loop). When we analyzed the mutational outcome of the controlled and controller  
79 positions using previous deep sequencing data, we observed that “controller” positions act as  
80 allosteric hot spots (i.e., mutations at these positions modulate DHFR activity), whereas mutations  
81 on “controlled” positions are usually deleterious. Thus, dynamics based approaches (particularly  
82 the “controller” and “controlled” classification) could be used to better understand the relationship  
83 between protein dynamics and function in other proteins.

84

## 85 **2 Results and Discussion**

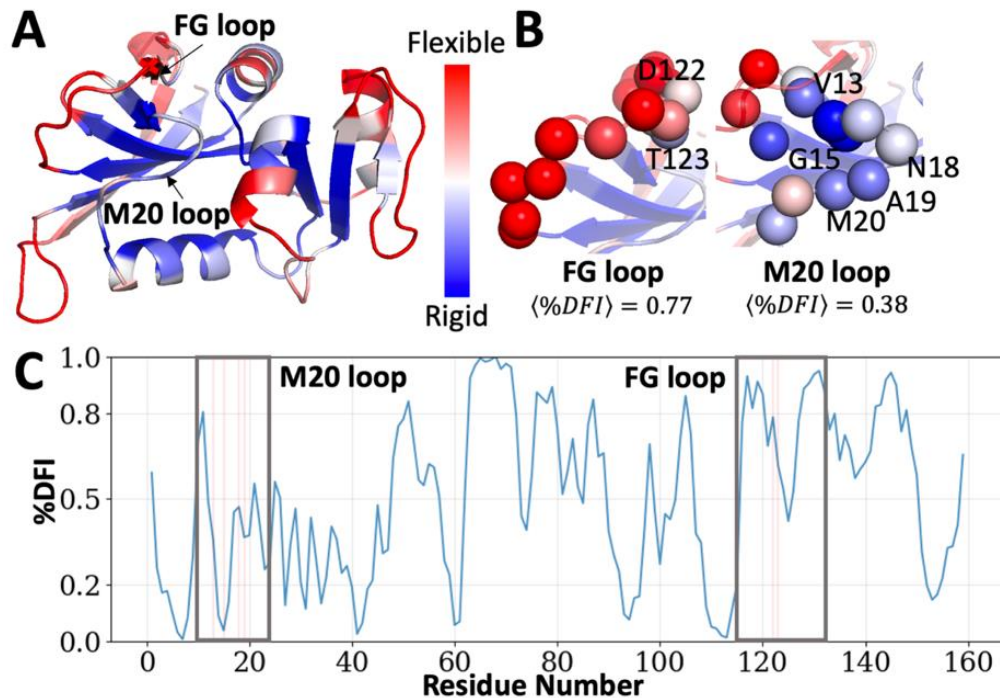
### 86 **2.1 Distinguishing Tolerant vs Non-Tolerant Mutations and Understanding Mutational** 87 **Outcomes using Dynamic Flexibility Analyses**

88 To gain insight into the impact of mutations on DHFR activity, we first investigated the  
89 enzyme using DFI (Figure 1). In our DFI analysis, we use Brownian force perturbation to capture  
90 each position’s response to random perturbations exerted on the protein chain. When a mutation  
91 creates a disturbance in the equilibrium dynamics of the protein, the local network of interactions  
92 surrounding the mutational site is often altered. Thus, the DFI value of a position can give a first

93 order approximation of the impact that a mutation at that site might have on the enzyme's activity.  
94 We previously demonstrated that a high correlation exists between DFI values and modulation of  
95 protein function by disease-related mutations.<sup>27,30,31,33</sup> Rigid locations identified by our DFI metric  
96 are often linked to disease related outcomes when mutated <sup>29,32,34,36,38</sup>; alternatively, flexible  
97 locations are less prone to these types of disadvantageous mutations.

98 Our efforts to study the relationship between dynamics and function in DHFR began with  
99 molecular dynamics (MD) simulations of the enzyme using a model of apo DHFR (PDB ID: 1rx2)  
100 from the Protein Databank. We chose to focus our simulations on the apo protein because previous  
101 studies using NMR suggested that the apo enzyme also samples bound state dynamics.<sup>12,39</sup> Thus,  
102 we believed that use of an apo structure would provide insight into the dynamics of DHFR in  
103 conformations present in both the apo and bound forms. We then analyzed these MD simulations  
104 using our DFI metric, which revealed that previously known functionally important M20 and FG  
105 loops display dynamics profile different than each other (Figure 1). In our previous studies<sup>30,35,40</sup>,  
106 we observed that residues that directly interact with ligands are often more rigid, and maintenance  
107 of this rigidity is important for enzyme function. We observe a similar trend with residue M20 and  
108 its neighbors (N18 and A19) in the M20 loop (Figure 1C), which directly interact with DHFR's  
109 substrate. These positions are more rigid in our analysis than the remainder of residues in the loop.  
110 Moreover, residues D122 and T123 in the FG loop, and V13 and G15 on the M20 loop (Figure  
111 1C) are also found to be less flexible relative to other residues within these loops. These residues  
112 have previously been shown to stabilize the neighboring loops<sup>6,9-11</sup>. In agreement with our  
113 previous studies<sup>29,32,34,36,38</sup>, substitutions at M20 and FG loop positions with low DFI scores are  
114 experimentally shown to drastically diminish (if not abolish) the activity of DHFR.<sup>6</sup> Overall, the

115 M20 loop shows a lower average DFI score ( $\langle \%DFI \rangle = 0.38$ ) compared to the FG loop  
 116 ( $\langle \%DFI \rangle = 0.77$ ) implying modulation of DHFR activity by these loops are different (Figure 1B).  
 117

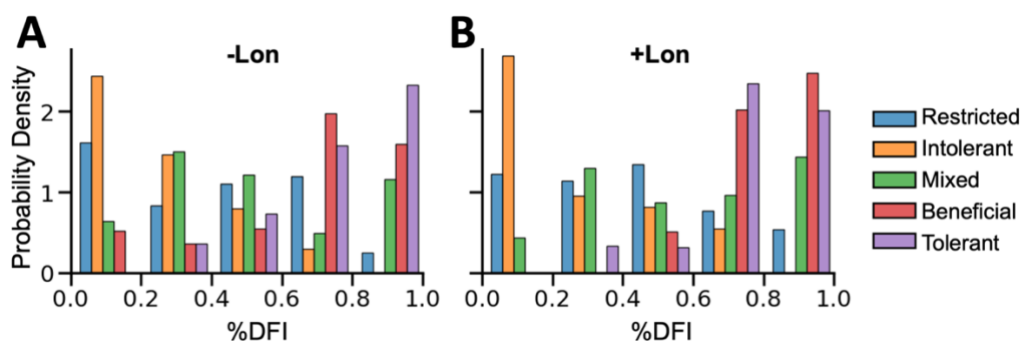


**Figure 1** A) DFI profile of DHFR projected on the crystal structure with PDB ID: 1rx2. Regions of high flexibility are colored red; regions of medium flexibility are colored white and highly rigid regions are colored blue. B) Functionally critical residues in the M20 (residues 9-24) and FG (residues 116-132) loops are shown as spheres colored by their DFI scores. C) The DFI profile of apo DHFR. D122 and T123 in the FG loop; and V13, G15, N18, and A19 on the M20 loop are highlighted with red colored vertical lines.

118

119 To further understand the implication of the conformational dynamics of residues related  
 120 to function in DHFR, we sought to relate data obtained using our DFI metric to previously reported  
 121 experimental data.<sup>6</sup> Namely, we used the per residue functional classification defined by

122 Thompson et al.<sup>6</sup>, in which positions with advantageous mutations (named “Beneficial”), positions  
 123 with WT-like behavior (named “Tolerant”), positions that possess both advantageous and  
 124 disadvantageous mutations (named “Mixed”), residues with mostly disadvantageous mutations  
 125 (named “Restricted”), and locations that exhibited almost no activity when mutated away from the  
 126 wild-type amino acid (named “Intolerant”) are described. When we analyzed positions belonging  
 127 to the aforementioned categories from the perspective of their DFI values, we observed the  
 128 following trends both in the absence (-Lon ) and the presence (+Lon) of Lon protease (Figure 2A):  
 129 We found that "Tolerant" and "Beneficial" mutations were more commonly found in residues with  
 130 high %DFI scores, suggesting that more flexible residues are better able to accommodate  
 131 mutations without negatively impacting protein function. In contrast, "Intolerant" and "Restricted"  
 132 mutations were more commonly found in residues with lower %DFI scores, suggesting that more  
 133 rigid residues are less able to tolerate mutations without negatively impacting protein function.  
 134 The difference between "Tolerant" and "Intolerant" distributions are statistically significant  
 135 ( $p=0.0002$ ). As well as the difference in distributions of "Beneficial" and "Restricted" with a  $p$   
 136 value of  $2e-05$ . The "Mixed" residues do not show a particular trend towards either rigid or flexible.  
 137



**Figure 2.** DFI score distributions for the five previously defined functional classes with and without the in the presence of Lon protease in Thompson et al.<sup>6</sup> A) In the absence of Lon protease

the “Intolerant” labeled residues almost always display very low DFI values, followed by the residues labeled as “Restricted” showing an overall rigid behavior (i.e., %DFI<0.6). Conversely, “Beneficial” and “Tolerant” residues are more commonly found in high DFI regions of the protein (i.e., % DFI≥0.6). The differences in these distributions are statistically significant, with p-values 0.0002 and 2e-05 respectively, calculated by Fisher's exact test using 0.6 %DFI as the threshold value. Residues labeled “Mixed” are distributed across different DFI ranges. B) In the presence of Lon protease, DFI scores of the residues distributed among functional classes are similar to those when the Lon protease is absent.

138

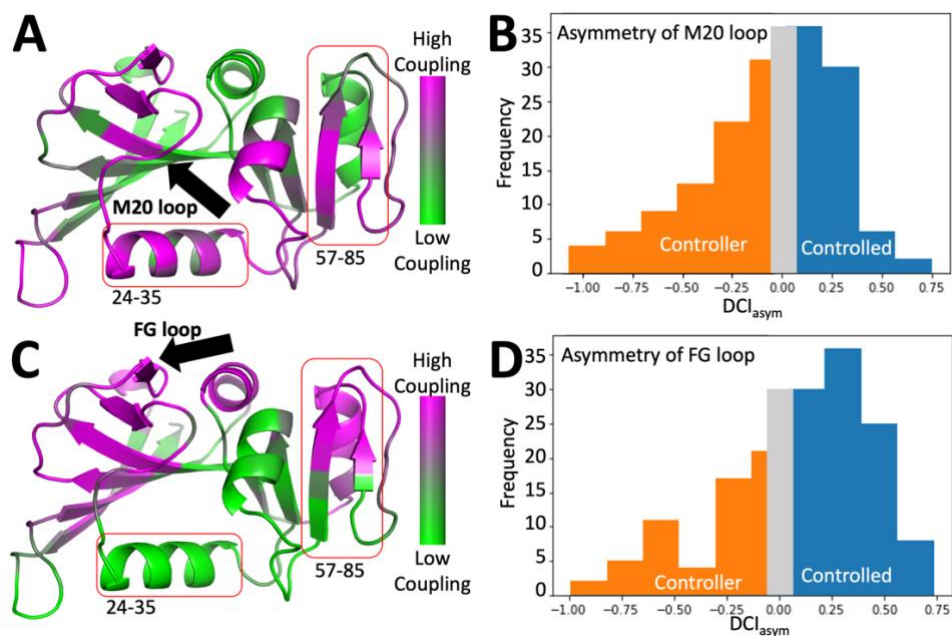
139           Moreover, a broad range of mutations, including many advantageous ones, face penalties  
140 due to Lon protease activity. To understand the relationship between DFI and protease sensitivity  
141 of residues, we investigated the distribution of DFI values. Our analysis focused on two distinct  
142 categories: residues that exhibit tolerance to Lon (i.e., residues denoted as “Beneficial” when Lon  
143 is absent that remain “Beneficial” when Lon is present), and residues that are susceptible to Lon  
144 (i.e., those originally labeled as “Beneficial”, but which become “Restricted” in the presence of  
145 Lon) (Figure S1). The box plots shows that residues that are susceptible to the presence of Lon  
146 protease overall have a slightly higher DFI value compared to residues that are tolerant suggesting  
147 that enhanced flexibility, low rigidity might play a role with stability, as our earlier studies showed  
148 that rigid sites contribute to overall folding stability of a protein.<sup>34,41,41</sup> In summary, these results  
149 support the strength of the DFI metric in assessing the functional outcome of mutations on the  
150 protein, regardless of whether they proximal to or distal from functionally important loops.

151 **2.2 Asymmetry in Dynamic Coupling Reveals Allosteric Mutational Sites**

152           Assessment of DHFR with our DFI metric also raised other important questions: First, how  
153 do residues distal to functionally important loops impact the overall enzymatic activity and second,  
154 what role do dynamic networks play in the control of DHFR function? Mutations found far from  
155 (but dynamically coupled to) functionally important loops, which we term “allosteric mutations”,  
156 have previously been shown to substantially affect enzyme activity.<sup>6,14,24,31</sup> The manner in which  
157 such this long-range dynamic communication propagates through proteins can be highlighted with  
158 our dynamic coupling metric DCI.<sup>32,41,42</sup> A high DCI value indicates high dynamic coupling  
159 between residues  $i$  and  $j$ , suggests that strong communication between these residues exists. A low  
160 DCI score implies weak coupling between residues and suggests the absence of strong  
161 communication between them.<sup>27,28,31,33,35</sup>

162           We applied DCI analysis to the M20 and FG loops to explore how these loops affect protein  
163 activity. Dynamic coupling analyses reveal that the M20 and FG loops, despite being close to each  
164 other, exhibit different long-distance interactions with the rest of the protein (Figure 3A). We also  
165 discovered that each loop is dynamically coupled to specific regions within DHFR. For example,  
166 helix B, which spans residues 24 to 35, is more coupled to the M20 Loop (Figure 3A), while the  
167 FG loop is highly coupled to  $\beta$  sheets C and D and the helical “E region” (residues 57 to 85, Figure  
168 3C).

169



**Figure 3** An analysis of DHFR using  $DCI_{asym}$ . Panels A and C show DCI profiles measuring the dynamic coupling of the M20 (A), and FG (C) loops projected onto the DHFR structure; high coupling is shown in purple, low coupling is shown in green. Panels B and D show the distribution of  $DCI_{asym}$  values of all residues calculated by targeting the M20 (B) and FG (D) loops. Positive  $DCI_{asym}$  values indicate that the residues within the loop control interactions with other residues while negative values represent residues that control the dynamics of the loop.

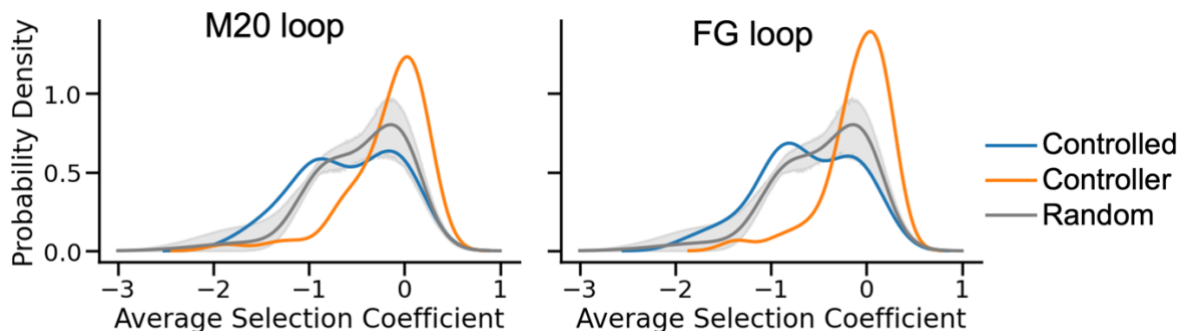
170

171 The complex network of the protein immediately suggested a disparity in dynamic  
 172 coupling between positions that could be understood by an asymmetry in communication (Figure  
 173 3). Since each residue directly contacts a distinct set of neighboring residues, each position in a  
 174 protein has a unique coupling network. Moreover, the dynamic coupling for position  $i$  with respect  
 175 to  $j$  is not necessarily symmetric to the dynamic coupling of  $j$  to  $i$ . Thus, changes at position  $i$  may  
 176 have larger effect on the flexibility of position  $j$  and vice versa. To capture this asymmetry, we  
 177 created a novel metric  $DCI_{asym}$ .<sup>31</sup> If the magnitude of difference between dynamic coupling scores

178 of positions  $i$  to  $j$  ( $DCI_{ji}$ ) vs the coupling of  $j$  to  $i$  ( $DCI_{ij}$ ) is significant, an asymmetry in  
179 communication between the two residues will exist. This asymmetry in communication can be  
180 informative on why certain amino acid substitutions at particular positions are more deleterious or  
181 beneficial to activity, vice versa.

182 To explore how dynamic coupling between the M20 loop and the B helix or the FG loop  
183 and the E helix and the C and D sheets affected enzyme function, dynamic coupling was considered  
184 in both directions using our  $DCI_{asym}$  metric. We first applied  $DCI_{asym}$  to the M20 loop. We consider  
185  $DCI_{asym}$  values between -0.05 and 0.05 to suggest that both positions are symmetric in their  
186 coupling; in other words, neither position has dominance. Alternatively, a residue would be  
187 defined as an “M20 controlled position” when its average  $DCI_{asym}$  score (calculated by taking the  
188 average of all M20 loop positions; i.e.,  $\langle DCI_{asym} \rangle$ ) is positive and larger than 0.05, and as an “M20  
189 controller position” when  $\langle DCI_{asym} \rangle$  is negative and lower than -0.05 (Figure 3B). The same  
190 analysis is repeated on the FG loop (Figure 3D). The “controlled”/”controller” categorization of  
191 residues is then compared with average selection coefficient values from the work of Thompson  
192 et al.<sup>6</sup> (Figure 4). Selection coefficient values represent the impact of a mutation to DHFR activity  
193 relative to wild type. A mutation with a selection coefficient value around zero ( $\pm 0.2$ ) is considered  
194 as neutral. Values higher than 0.2 are beneficial to function, and conversely values lower than -0.2  
195 are deleterious.

196



**Figure 4** Analyses of the “controller” and “controlled” classified average selection coefficient value distributions (+Lon) for the M20 and FG loops. For both M20 and FG loops the average selection coefficient value distributions are different for “controller” and “controlled” labels. The residues with “controller” labels are commonly distributed near either neutral/enhanced (near zero, or positive) region while “controlled” residues display a distribution among negative values (deleterious) (M20 loop:  $p= 0.005$ , and FG loop:  $p= 0.008$ , Student's t-test). The gray distribution (line as the mean and shade as the variance) is generated by randomly selecting a different subset of residues (excluding “controller” / “controlled” residues) five times. Comparison of the randomly selected positions’ average selection coefficient distributions with those of “controller” and “controlled”, distributions of both M20 loop and FG loop shows that randomly selecting residues fails to capture the selection coefficient distribution of the “controller” residues (average  $p$  values over 5 random samples are 0.028, and 0.001, respectively) and “controlled” residues ( $p < 0.043$  and  $< 0.0425$ , respectively).

197

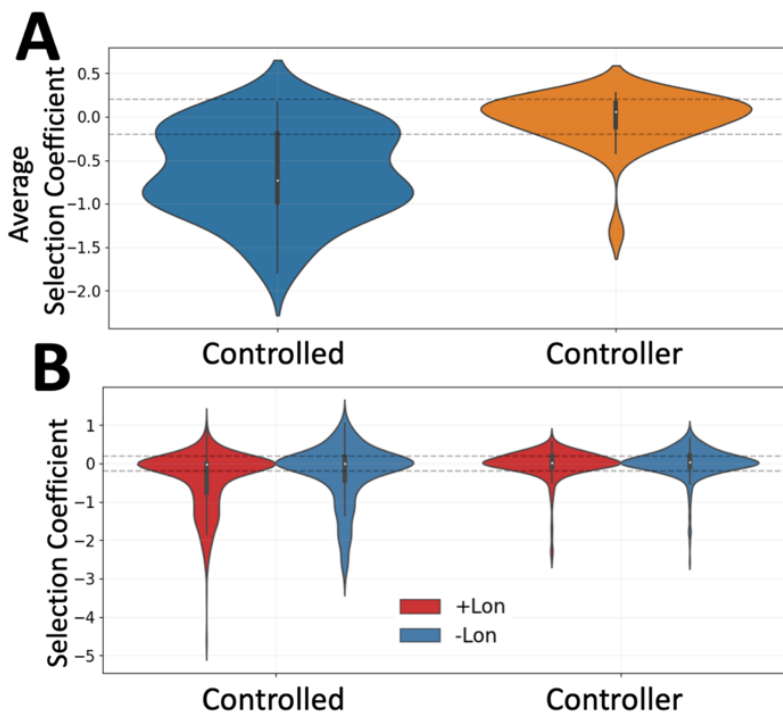
198           When investigated, the distribution of average selection coefficient values of “controlled”  
 199 and “controller” residues displayed a different pattern (Figures 4 and S2A). When residues that  
 200 control the M20 loop are considered, the peak of the distribution is observed to be above zero,  
 201 which indicates that mutations at these positions have, on average, a positive impact on the activity.

202 Conversely, sites “controlled” by residues in the M20 loop display a broad distribution with high  
203 density around very negative values. This suggests that mutations at these residues have a  
204 deleterious effect on protein activity. A similar trend is observed when the FG loop is targeted with  
205 DCI and DCI<sub>asym</sub> analyses (Figure 4) (Figure S2B). Furthermore, comparison of the average  
206 selection coefficient distributions of “controlled” and “controller” residue positions with that of  
207 randomly selected positions reveals the statistical significance of the distribution of our  
208 classifications in distinguishing the impact of mutations on activity (Figure 4).

### 209 **2.3 Beneficial Mutations are Enriched at Controller Sites**

210 To further assess the impact of amino acid substitutions on residues with variety of control  
211 over the M20 and FG loops, we combined the “controlled” and “controller” designations from  
212 each loop. Namely, in this analysis, a residue was defined as a “controller” if it exerts control over  
213 both the M20 and FG loops simultaneously and is considered a “controlled” residue otherwise.  
214 The average selection coefficient value distributions of “controlled” and “controller” residues  
215 differ from each other when viewed in this way (Figure 5A). Controller residues generally present  
216 more activity-enhancing amino acid substitutions compared to residues in the “controlled”  
217 category. Interestingly, the peak of the distribution of “controlled” residue mutations is below the  
218 neutral range (near -1.0). This indicates that mutations to “controlled” positions more often yield  
219 deleterious outcomes with respect to function. On the other hand, mutation of “controller”  
220 positions could gradually modulate function both positively and negatively and could therefore act  
221 like rheostatic switches.<sup>31</sup> This ultimately suggests that the M20 and FG loops themselves are  
222 highly conserved due to functional constraints. However, those residues that control the loops can  
223 affect the overall enzyme function by distally altering functionally important residues.

224



**Figure 5** Experimentally measured selection coefficient values of “controller” and “controlled” residues of the M20 and FG loops. A) A violin plot of average selection coefficient values of the residues controlling both loops suggests that these residues have a peak on positive values compared to those residues that are controlled by the loops ( $p=3e-07$ ). This suggests that mutations to residues that are controller of the M20, and FG loops can potentially enhance the activity of DHFR, while mutations to residues controlled by these loops are mostly deleterious. B) A violin plot generated using the selection coefficient of all amino acid substitutions per position. The distribution of selection coefficient values for “controller” residues falls primarily in the neutral to positive range. Alternatively, a broader distribution is observed for residues controlled by both loops; mutations at these residues often have a drastic negative impact on activity.

226 To remove any bias that arose from averaging, we also obtained the distributions using  
227 selection coefficient values for every mutation (as opposed to average values for all mutational  
228 outcomes per position). When all selection coefficient values are considered, the differences in  
229 asymmetry between “controlled” and “controller” residues is more pronounced (Figure 5B).  
230 Additionally, when other functionally important sites, GH loop (spanning through residues 142 to  
231 149) and Adenosine Binding Domain (residues 63, 64, and 65) are investigated, the results are  
232 similar to those found with M20 and FG loops (Figure S3). This striking difference in the  
233 distribution of functional outcomes of mutations on “controlled” versus “controller” residues  
234 illustrate the importance of dynamic allosteric control.<sup>7,43</sup> Previously, we explored asymmetry in  
235 dynamic coupling by analyzing 591 pathogenic missense variants in 144 human enzymes.<sup>32</sup> We  
236 showed that many mutations, far from the active site, exhibit deleterious behavior (sometimes  
237 leading to pathogenicity) due to their high coupling with the active site. Furthermore, we also  
238 observed that these mutations are coupled to the active site, but the coupling strength (DCI score)  
239 of the mutation sites back to active site is not as high, showcasing an indifference in coupling  
240 strength (asymmetry). The “controller” and “controlled” classification developed in this present  
241 work highlights the importance of dynamic coupling to active sites, in agreement with previous  
242 study. In addition, it highlights the degree to which asymmetry in this coupling can modulate  
243 function in a positive or negative direction.

#### 244 **2.4 Leveraging Asymmetry in Dynamic Coupling for Fine-Tuning Function: A Comparative** 245 **Analysis of Other Metrics and Functional Outcomes**

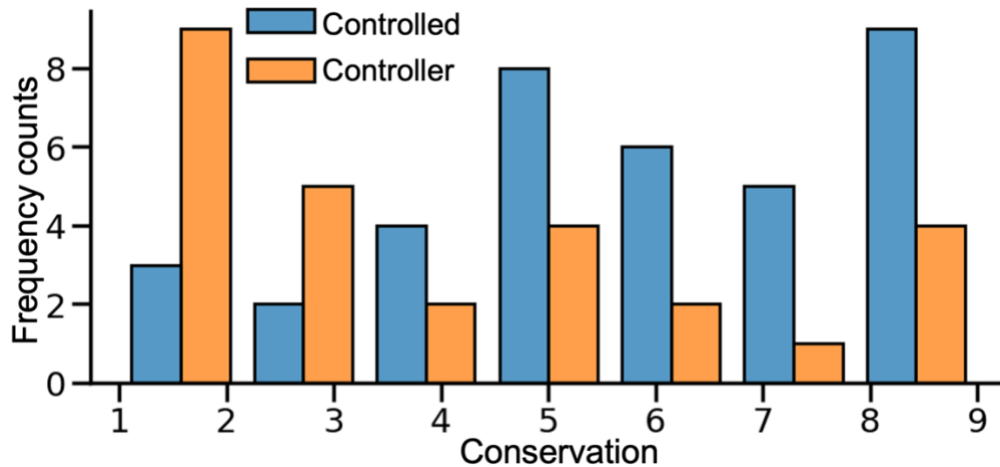
246 To evaluate the effectiveness of the DCI<sub>asym</sub> metric in identifying residue positions with  
247 diverse activities upon mutations, we compared it with several other metrics that are commonly  
248 used to identify functionally critical sites such as solvent-accessible surface area (SASA)<sup>34,44–46</sup>,

249 average # of contacts as well as network metrics including betweenness, closeness, and eigenvector  
250 centrality values (Figure S4).<sup>47-49</sup> (See methods) After computing these metrics for each residue,  
251 we grouped positions sharing similar values using histograms and analyzed the average and  
252 variance of the experimental average selection coefficients of the positions residing in each  
253 bin/group. The average of SASA values in each bin correlates with average experimental values  
254 ( $R=0.88$ ), indicating that highly accessible residues are more likely to have neutral outcomes upon  
255 mutation. The average number (#) of contacts shows a correlation of  $-0.84$  with experimental  
256 fitness, but the deviation at medium ranges suggests that, on average, most of these residues are  
257 deleterious to function when mutated. The betweenness scores show a relatively low correlation  
258 with the experimental values ( $R=-0.65$ ). Despite its strong correlation with fitness ( $R=-0.85$ ), the  
259 observed average negative fitness values with all eigenvector centrality ranges suggest that  
260 underlying factors beyond altered functional outcomes upon substitution are not fully captured by  
261 this metric. The closeness measure identifies residue positions with high experimental fitness  
262 scores ( $>0.2$ ) but fails largely to identify residues with deleterious behavior. In contrast, when we  
263 analyzed dynamic-based metrics (e.g.,  $DCI_{asym}$ ) of the M20 and FG loops in the same manner, the  
264 results show that the  $DCI_{asym}$  metric is the most effective in capturing the trend of changing fitness  
265 for both the M20 loop and FG loop, with high correlations of  $-0.98$  and  $-0.97$ , respectively. This  
266 indicates that as a position becomes more controlled, mutations at that site are more deleterious;  
267 alternatively, “controller” residues yield more neutral or beneficial outcomes when mutated. These  
268 findings suggest that the “controlled” and “controller” classification based on the  $DCI_{asym}$  metric  
269 can not only provide high accuracy in identifying and characterizing residues, put can also help  
270 identify “controller” sites that may subtly tune the enzyme’s function when mutated.  
271

272 **2.5. Examining the Interplay of Asymmetry in Dynamic Coupling and Evolutionary**  
273 **Conservation**

274 To gain insight into how our “controller”/“controlled” categorization relates to a position’s  
275 conservation, we utilized the ConSurf database<sup>50,51</sup> to evaluate conservation of each site. Namely,  
276 the distribution of conservation of each residue in DHFR was considered with respect to its  
277 asymmetry categorization (Figure 6). Previous studies have shown that positions of M20 and FG  
278 loops are structurally and evolutionarily conserved.<sup>5,22,52</sup> When the conservation of residues  
279 controlled by the M20 and FG loops was investigated, the “controlled” residues were found to be  
280 highly conserved, which agrees with our analysis. This suggests that mutations of “controlled”  
281 residues often yield deleterious outcomes and are therefore filtered out by natural selection.  
282 Alternatively, “controller” sites are found to be non-conserved, indicating that they can  
283 accommodate a diverse number of mutations. This behavior is observed in the deep mutational  
284 scanning data<sup>6</sup> where mutations on “controller” residues enable enhancement or modular changes  
285 in the activity of DHFR. Conservation analyses showed that “controlled” residues are highly  
286 conserved, while “controller” sites are non-conserved, allowing for a diverse number of mutations  
287 and enabling enhancement or modular changes in the activity of DHFR. Indeed, these results agree  
288 with the deep sequencing data where the mutations on controlled sites are usually deleterious;  
289 therefore, they are also eliminated during evolution.

290



**Figure 6** Conservation distribution of DHFR positions designated either controlled by or controllers of the M20 and FG loops. Conservation values are obtained using ConSurf database.<sup>50,51</sup> Residues that are “controller” attain lower values (non-conserved) compared to “controlled” residues which are more distributed on higher (conserved) values. The Student's t-test showed that the difference in distribution was statistically significant ( $p= 0.003$ ).

291 To gain a deeper insight into why a distinction between “controlled” and “controller”  
 292 residue conservation exists, we examined the flexibility of these positions; our previous studies  
 293 highlighted a strong correlation between conservation and flexibility<sup>31,32,36,37,41</sup> The analyses  
 294 demonstrate that "controlled" residues are often highly rigid with an average %DFI score of 0.2  
 295 (Figure S5). Mutations occurring at these rigid sites typically have detrimental effects on function.  
 296 In contrast, the "controller" residues exhibit a higher average %DFI value of 0.8 (Figure S5),  
 297 indicating high flexibility. This flexibility enables these positions to tolerate a broader range of  
 298 amino acid changes. Consequently, selectively targeting "controller" residues holds promise for  
 299 random mutagenesis or rational design approaches aiming to finely adjust the activity of DHFR.  
 300 We believe our dynamics metrics DFI and DCI<sub>asym</sub> could uncover these positions in other proteins  
 301 as well.

302

### 303 **3 Conclusion**

304           In this study, we show that dynamic based metrics can be utilized to better understand the  
305 functional outcomes of mutations in DHFR. DFI scores displayed great promise in differentiating  
306 positions that might lead to beneficial or deleterious functional changes. The DCI metric revealed  
307 that the long-distance dynamic coupling between the M20 loop and other residues in DHFR differs  
308 significantly differs from that of FG loop. These diverse allosteric features are further investigated  
309 with our novel DCI<sub>asym</sub> metric. The observed differences between residues that are controlled by  
310 or control two important loops in DHFR highlight how mutation of “controller” residues can fine  
311 tune enzyme activity through dynamic allostery. In addition, the evaluation of evolutionary  
312 conservation of “controlled” versus “controller” positions indicated that the “controller” sites are  
313 more amenable to mutations. On the other hand, “controlled” sites are more conserved since  
314 mutations to these sites often results in loss of function. Although our study was carried out using  
315 DHFR, the conclusions drawn in this work display great promise for using dynamics metrics to  
316 gain a better understanding of how residues distal from functional portions of proteins can  
317 potentially modulate protein activity without compromising the fold.

318

### 319 **4 Methods**

#### 320 4.1 Molecular dynamics simulations

321           We used the AMBER molecular dynamics software<sup>53</sup> to study the dynamics of *E. coli*  
322 DHFR. The protein system is parametrized with ff14SB force field<sup>54</sup> and solvated with TIP3P  
323 explicit water model using minimum 16Å distance from the protein to define the box size. The  
324 solvated protein is neutralized by sodium and chlorine ions and the energy is minimized with a

325 steepest descent algorithm by ten thousand steps. The production trajectory was simulated with  
326 Isothermal, isobaric, constant number of particles ensemble (NPT) at 300K and 1 bar pressure.  
327 Langevin thermostat was utilized to maintain the kinetic temperature of 300K, and the pressure is  
328 regulated by the Berendsen barostat. Additionally, SHAKE algorithm was used constrain the  
329 hydrogens. The simulation is run for 2 $\mu$ s until convergence is achieved. We considered the  
330 simulation converged when the root mean square deviation (RMSD) between the highest sampled  
331 conformation in consecutive time windows (i.e., the last 300ns windows and the 300ns window  
332 sequentially before it) is lower than 1Å. Similar to the procedure described by Sawle and Ghosh<sup>55</sup>,  
333 we used window sizes ranging from 100ns up to 1 $\mu$ s to determine convergence.

#### 334 4.2 Dynamic Flexibility Index (DFI)

335 Our DFI metric calculates the relative flexibility/rigidity of individual residues in an  
336 protein.<sup>28–30,42,56</sup> The DFI algorithm, which is developed using Linear Response Theory (LRT) and  
337 Perturbation Response Scanning (PRS), calculates the average response of a residue as a result of  
338 a perturbation on every other residue in a protein.<sup>40</sup> Taking advantage of the residue covariances,  
339 DFI provides position specific flexibility profiles.

$$340 \quad [\Delta\mathbf{R}]_{3N \times 1} = [\mathbf{H}]_{3N \times 3N}^{-1} [\mathbf{F}]_{3N \times 1} \quad (1)$$

341 A Hessian matrix,  $\mathbf{H}$ , is compiled from the second derivatives of potentials. The inverse of the  
342 Hessian matrix,  $\mathbf{H}^{-1}$ , contains residue covariances. The covariance matrix can be generated from a  
343 protein structure by utilizing an Elastic Network Model (ENM) or gathered from a MD simulation  
344 of the protein, which implicitly accounts for amino-acid side chain interactions and solvent  
345 interactions. We used the latter to calculate the dynamic metrics in this study. The residue response  
346 vector,  $\Delta\mathbf{R}$ , is the resultant vector containing the magnitude of responses from multiplying the  
347 covariance matrix by the force vector,  $\mathbf{F}$ . The dynamic flexibility index (DFI) for position  $i$ , which

348 computes the normalized fluctuation response of a position upon perturbation on the chain is  
349 calculated as

$$DFI_i = \frac{\sum_{j=1}^N |\Delta R^j|_i}{\sum_{i=1}^N \sum_{j=1}^N |\Delta R^j|_i} \quad (2)$$

350  
351 where  $|\Delta R^j|_i = \sqrt{\langle(\Delta R)^2\rangle}$  is the magnitude of fluctuation response at position  $i$  due to a  
352 perturbation at position  $j$ .

353 The DFI score yields position specific information about the conformational dynamics of  
354 a protein system. Positions displaying low DFI scores are highly rigid. These sites often make  
355 more than an average number of interactions with their neighbors, which suggests that they  
356 represent crucial dynamic hubs in a protein. Conversely, positions with high DFI scores are often  
357 highly mobile regions of a protein. These sites do not contribute to the collective motion of a  
358 protein as substantially as the rigid regions.

#### 359 4.3 Dynamic Coupling Index (DCI) and DCI asymmetry (DCI<sub>asym</sub>)

360 The DCI metric stems from the same fundamental analysis method that is used to carry out  
361 a DFI calculation.<sup>28,31</sup> DCI measures the allosteric coupling between residue pairs. To carry out  
362 DCI analysis of DHFR, a random unit force was applied to residues contained in M20 & FG loops  
363 and was allowed to propagate through the protein until it reached a residue distal from the initial  
364 perturbation location. After probing all active site residues, we calculate a “magnitude of response”  
365 to other residues in the protein, which ultimately represents the strength of coupling between each  
366 active site residue and all other residues in the protein. A calculated DCI of position  $i$  suggests its  
367 response to a perturbation on position  $j$  and is calculated as follows:

$$DCI_{ij} = \frac{|\Delta R^j|_i}{\sum_{j=1}^N |\Delta R^j|_i / N} \quad (3)$$

369 where,  $|\Delta R^j|_i = \sqrt{\langle(\Delta R)^2\rangle}$  is the magnitude of the fluctuation response at position  $i$  due to  
370 perturbations at position  $j$  normalized over the average response of position  $i$  when any position in  
371 the protein is perturbed by a random Brownian force. Thus,  $DCI_{ij} > 1$  indicates that position  $i$  is  
372 more sensitive to perturbations occurring on position  $j$ . Alternatively, a position with a  $DCI_{ij}$  value  
373 lower than 1 is regarded as weakly coupled to the site  $j$ . Moreover, the dominance in dynamic  
374 control can be determined by calculating the asymmetry between residue locations  $i$  and  $j$ .  $DCI_{ij}$  is  
375 defined as the response of residue  $i$  when residue  $j$  is perturbed and  $DCI_{ji}$  represents the response  
376 of residue  $j$  when residue  $i$  is perturbed. DCI asymmetry ( $DCI_{asym}$ )<sup>27,31,32</sup> of location  $i$  is calculated  
377 as follows:

$$378 \quad DCI_{asym} = DCI_{ij} - DCI_{ji}$$

379 Given this definition,  $DCI_{asym}$  can take both positive and negative value. Accordingly, we  
380 consider residues with  $DCI_{asym}$  values around zero (between negative -0.05 to 0.05) to be  
381 dynamically coupled with the M20 and FG loops in a symmetric fashion. The residues with  
382  $DCI_{asym}$  values higher than 0.05 are considered as “controlled” (e.g., M20 loop controlled) and the  
383 ones with  $DCI_{asym}$  values lower than -0.05 as “controller” (e.g., M20 loop controller).

#### 384 4.4 Solvent Accessible Surface Area (SASA)

385 The SASA calculation is employed by using Naccess.<sup>57</sup> Naccess algorithm first creates a  
386 sphere with the radius of a water molecule and then rolls the sphere on the surface of the protein.  
387 The accessible surface area is calculated per residues by measuring the fraction of residue that is  
388 accessible to the solvent.

#### 389 4.5 Network Features

390 Network Analysis of Protein Structures (NAPS) webserver is utilized to calculate the  
391 network features betweenness, closeness, and eigenvector centrality.<sup>47</sup> Betweenness measures how

392 often an amino acid lies on the shortest path between two other amino acids in the protein. High-  
393 betweenness nodes have been previously shown as important residues for protein structure and  
394 function.<sup>48</sup> These residues are crucial in proteins, as the shortest paths between nodes (i.e., distal  
395 sites and active sites) pass through these nodes. Closeness metric shows how easily an amino acid  
396 can be reached by other amino acids in the protein. Eigenvector centrality measures how well an  
397 amino acid is connected to other important amino acids in the protein. Amino acids that are more  
398 easily reached by others and well connected to other important amino acids are important for  
399 maintaining the overall stability and function of the protein.<sup>43,49,58</sup>

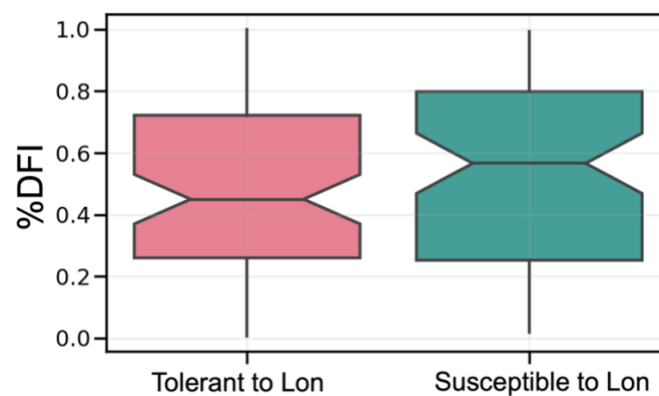
#### 400 4.6 Number of Contacts

401 To determine the average number of contacts, we analyzed the MD simulation trajectory  
402 by counting the C $\alpha$  contacts within 10Å for each residue that appeared in over 80% of the frames  
403 in the trajectory sampled every 1ns.

404

#### 405 5 Supplementary Figures

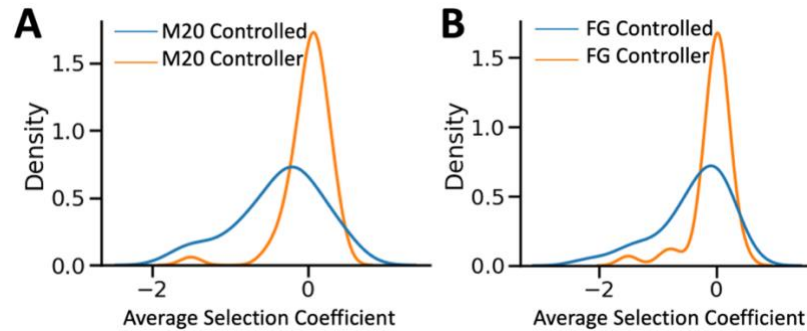
406



**Figure S1.** Box plot of DFI values for two sets of residues related to their protease sensitivity. Residues that are tolerant to Lon protease have slightly lower DFI scores compared to the one

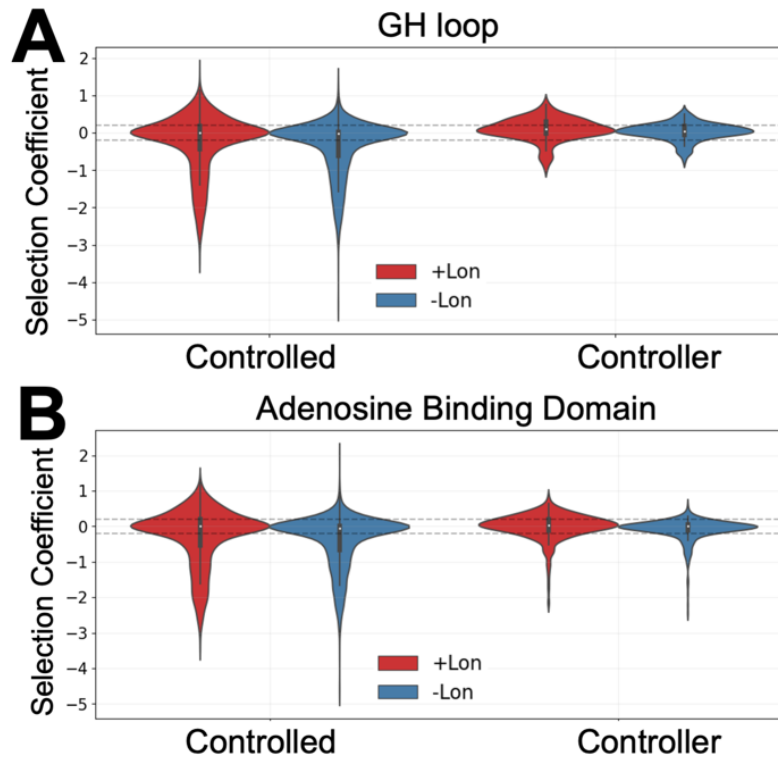
that are susceptible ( $p < 0.23$ ). This shows that the susceptible residues have a higher degree of flexibility than the tolerant residues. This observation is interesting because it may suggest that the degree of flexibility of a residue play a role in its susceptibility to protease activity and its overall stability.

407

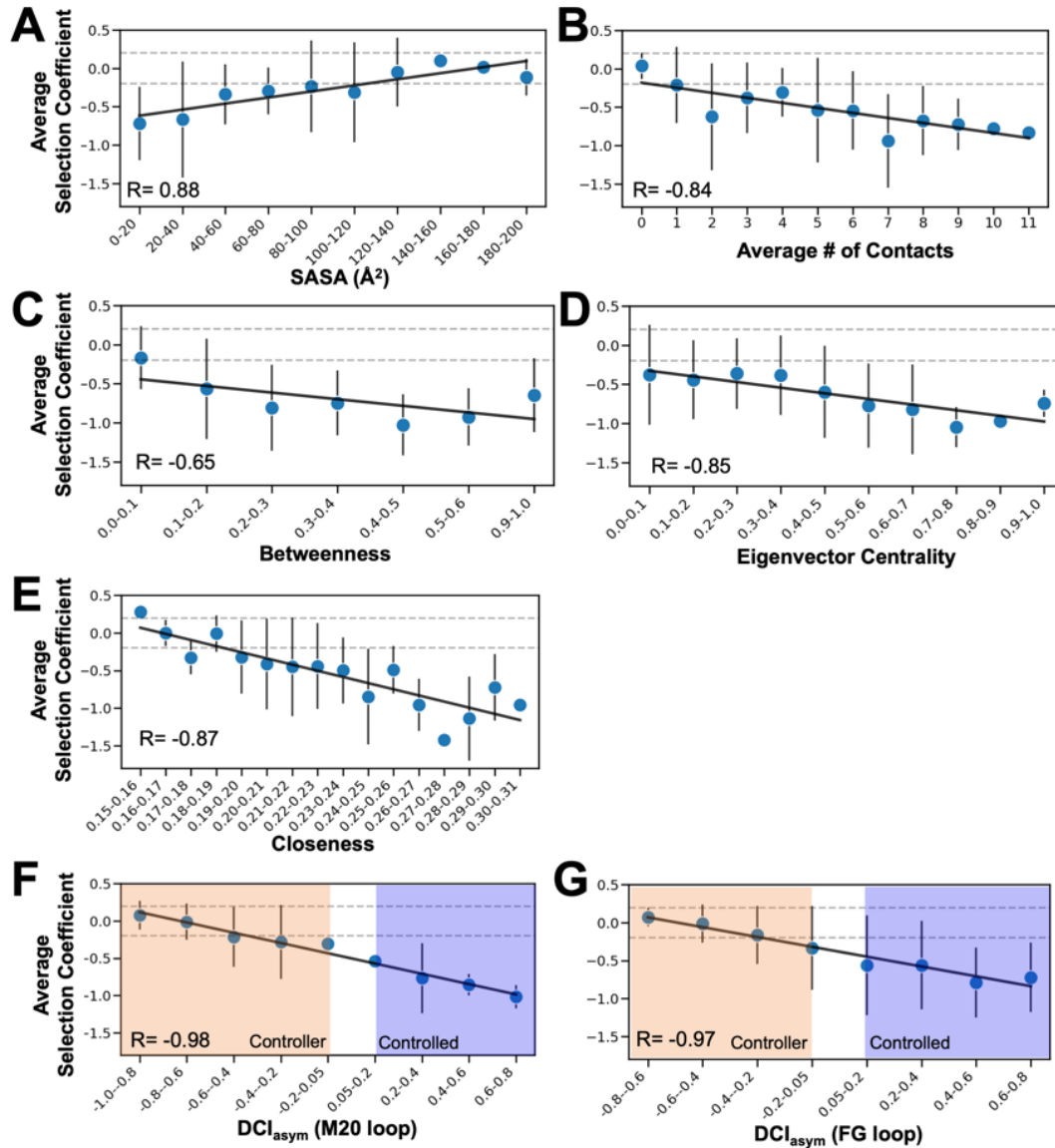


**Figure S2** The asymmetry labeled average selection coefficient value (in the absence of Lon) distributions for the M20 and FG loops. A) Distribution for M20 loop plot shows “controller” and “controlled” labeled distributions are different ( $p= 0.005$ , Student's t-test). B) FG loop value distribution show “controller” and “controlled” labeled distributions are different. ( $p= 0.008$ , Student's t-test).

408



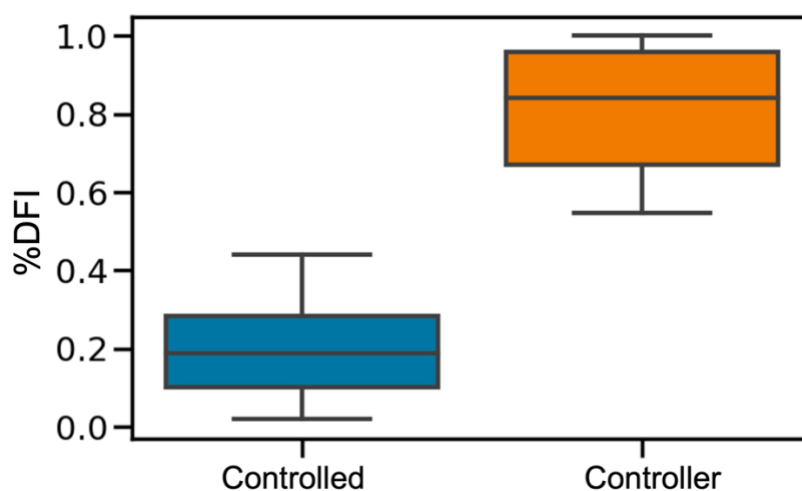
**Figure S3** Violin plots of experimentally measured selection coefficient values of “controller” and “controlled” residues of the A) GH loop and B) Adenosine Binding Domain. The distribution of selection coefficient values for “controller” residues follows an overall neutral trend. On the other hand, “controlled” residues show a diverse distribution spreading to negative (deleterious) ranges. The difference observed in the distributions of “controlled” and “controller” are statistically significant, with  $p$  values 0.003 and  $3e-07$ , for GH loop and Adenosine Binding Domain, respectively.



**Figure S4** Correlation plots of binned structural and dynamic features with average selection coefficients A) Structural features SASA and B) average number (#) of contacts are binned and compared with experimental data. Both SASA and average # of contacts values of residues correlate well with the experimental data. Structural metrics betweenness, closeness, and eigenvector centrality are compared with average selection coefficient. C) Betweenness metric binned every 0.1 range shows that ranges from zero to 0.2 and 0.9 to 1.0 have higher fitness values relative to others ( $R=-0.65$ ). D) Eigenvector centrality metric is binned every 0.1 range.

The eigenvector centrality metric overall shows a good correlation, but all the bins have an experimental value lower than the neutral range ( $R=-0.85$ ). E) Closeness metric binned every 0.01 range shows that residues with values from 0.15 to 0.19 shows great promise in enhancing the activity ( $R=-0.87$ ). F) M20 loop and G) FG loop  $DCI_{asym}$  value binned every 0.2 window shows that  $DCI_{asym}$  values lower than zero yield higher activity compared to those positions in the positive ranged bins. This correlation fits well with the definitions of “controlled” and “controller”.

410



**Figure S5** A box plot showing %DFI distribution of “controlled” and “controller” residues. These distributions show that “controller” sites attain high %DFI values on average; conversely “controlled” positions are generally found to be rigid.

411

## 412 **6 Supplementary File**

413 An Excel spreadsheet containing the data calculated or utilized in this study is provided as a  
414 supplementary file.

415

## 416 **7 Acknowledgements**

417 This research was supported by the Gordon and Betty Moore Foundations and National Science  
418 Foundation (1901709).

419

## 420 **8 Conflict of interest**

421 The authors declare no conflict of interest.

422

## 423 **9 References**

424 1. Davies J, Davies D (2010) Origins and Evolution of Antibiotic Resistance. *Microbiology and*  
425 *Molecular Biology Reviews* 74:417–433.

426 2. Aminov RI (2010) A Brief History of the Antibiotic Era: Lessons Learned and Challenges for the  
427 Future. *Front Microbiol* 1:134.

428 3. Laxminarayan R, Duse A, Wattal C, Zaidi AK, Wertheim HF, Sumpradit N, Vlieghe E, Hara GL,  
429 Gould IM, Goossens H (2013) Antibiotic resistance—the need for global solutions. *The Lancet*  
430 *infectious diseases* 13:1057–1098.

431 4. Martínez JL (2008) Antibiotics and antibiotic resistance genes in natural environments.  
432 *Science* 321:365–367.

433 5. Weinreich DM, Delaney NF, DePristo MA, Hartl DL (2006) Darwinian evolution can follow only  
434 very few mutational paths to fitter proteins. *science* 312:111–114.

435 6. Thompson S, Zhang Y, Ingle C, Reynolds KA, Kortemme T (2020) Altered expression of a  
436 quality control protease in *E. coli* reshapes the in vivo mutational landscape of a model enzyme.  
437 *Elife* 9:e53476.

438 7. McCormick JW, Russo MA, Thompson S, Blevins A, Reynolds KA (2021) Structurally  
439 distributed surface sites tune allosteric regulation. *elife* 10:e68346.

440 8. Schnell JR, Dyson HJ, Wright PE (2004) Structure, dynamics, and catalytic function of  
441 dihydrofolate reductase. *Annual review of biophysics and biomolecular structure* 33:119–140.

442 9. Reynolds KA, McLaughlin RN, Ranganathan R (2011) Hot spots for allosteric regulation on  
443 protein surfaces. *Cell* 147:1564–1575.

- 444 10. Luk LYP, Javier Ruiz-Pernía J, Dawson WM, Roca M, Loveridge EJ, Glowacki DR, Harvey JN,  
445 Mulholland AJ, Tuñón I, Moliner V, et al. (2013) Unraveling the role of protein dynamics in  
446 dihydrofolate reductase catalysis. *Proceedings of the National Academy of Sciences* 110:16344–  
447 16349.
- 448 11. Sawaya MR, Kraut J (1997) Loop and subdomain movements in the mechanism of  
449 *Escherichia coli* dihydrofolate reductase: crystallographic evidence. *Biochemistry* 36:586–603.
- 450 12. Boehr DD, McElheny D, Dyson HJ, Wright PE (2006) The Dynamic Energy Landscape of  
451 Dihydrofolate Reductase Catalysis. *Science* 313:1638–1642.
- 452 13. Cammarata MB, Thyer R, Rosenberg J, Ellington A, Brodbelt JS (2015) Structural  
453 Characterization of Dihydrofolate Reductase Complexes by Top-Down Ultraviolet  
454 Photodissociation Mass Spectrometry. *J. Am. Chem. Soc.* 137:9128–9135.
- 455 14. Benkovic SJ, Fierke CA, Naylor AM (1988) Insights into enzyme function from studies on  
456 mutants of dihydrofolate reductase. *Science* 239:1105–1110.
- 457 15. Bhabha G, Lee J, Ekiert DC, Gam J, Wilson IA, Dyson HJ, Benkovic SJ, Wright PE (2011) A  
458 dynamic knockout reveals that conformational fluctuations influence the chemical step of  
459 enzyme catalysis. *Science* 332:234–238.
- 460 16. Wang L, Tharp S, Selzer T, Benkovic SJ, Kohen A (2006) Effects of a distal mutation on active  
461 site chemistry. *Biochemistry* 45:1383–1392.
- 462 17. Agarwal PK, Billeter SR, Rajagopalan PTR, Benkovic SJ, Hammes-Schiffer S (2002) Network of  
463 coupled promoting motions in enzyme catalysis. *Proceedings of the National Academy of*  
464 *Sciences* 99:2794–2799.
- 465 18. Wong KF, Selzer T, Benkovic SJ, Hammes-Schiffer S (2005) Impact of distal mutations on the  
466 network of coupled motions correlated to hydride transfer in dihydrofolate reductase.  
467 *Proceedings of the National Academy of Sciences* 102:6807–6812.
- 468 19. Rod TH, Radkiewicz JL, Brooks CL (2003) Correlated motion and the effect of distal  
469 mutations in dihydrofolate reductase. *Proceedings of the National Academy of Sciences*  
470 100:6980–6985.
- 471 20. Tamer YT, Gaszek IK, Abdizadeh H, Batur TA, Reynolds KA, Atilgan AR, Atilgan C, Toprak E  
472 (2019) High-Order Epistasis in Catalytic Power of Dihydrofolate Reductase Gives Rise to a  
473 Rugged Fitness Landscape in the Presence of Trimethoprim Selection. *Molecular Biology and*  
474 *Evolution* 36:1533–1550.
- 475 21. Rodrigues JV, Bershtein S, Li A, Lozovsky ER, Hartl DL, Shakhnovich EI (2016) Biophysical  
476 principles predict fitness landscapes of drug resistance. *Proceedings of the National Academy of*  
477 *Sciences* 113:E1470–E1478.

- 478 22. Bhabha G, Ekiert DC, Jennewein M, Zmasek CM, Tuttle LM, Kroon G, Dyson HJ, Godzik A,  
479 Wilson IA, Wright PE (2013) Divergent evolution of protein conformational dynamics in  
480 dihydrofolate reductase. *Nat Struct Mol Biol* 20:1243–1249.
- 481 23. Epstein DM, Benkovic SJ, Wright PE (1995) Dynamics of the Dihydrofolate Reductase-Folate  
482 Complex: Catalytic Sites and Regions Known To Undergo Conformational Change Exhibit Diverse  
483 Dynamical Features. *Biochemistry* 34:11037–11048.
- 484 24. Gekko K, Kamiyama T, Ohmae E, Katayanagi K (2000) Single Amino Acid Substitutions in  
485 Flexible Loops Can Induce Large Compressibility Changes in Dihydrofolate Reductase1. *The*  
486 *Journal of Biochemistry* 128:21–27.
- 487 25. Mauldin RV, Carroll MJ, Lee AL (2009) Dynamic Dysfunction in Dihydrofolate Reductase  
488 Results from Antifolate Drug Binding: Modulation of Dynamics within a Structural State.  
489 *Structure* 17:386–394.
- 490 26. Mauldin RV, Lee AL (2010) Nuclear Magnetic Resonance Study of the Role of M42 in the  
491 Solution Dynamics of Escherichia coli Dihydrofolate Reductase. *Biochemistry* 49:1606–1615.
- 492 27. Campitelli P, Ozkan SB (2020) Allostery and Epistasis: Emergent Properties of Anisotropic  
493 Networks. *Entropy* 22:667.
- 494 28. Larrimore KE, Kazan IC, Kannan L, Kendle RP, Jamal T, Barcus M, Bolia A, Brimijoin S, Zhan C-  
495 G, Ozkan SB (2017) Plant-expressed cocaine hydrolase variants of butyrylcholinesterase exhibit  
496 altered allosteric effects of cholinesterase activity and increased inhibitor sensitivity. *Scientific*  
497 *reports* 7:10419.
- 498 29. Kumar A, Glembo TJ, Ozkan SB (2015) The Role of Conformational Dynamics and Allostery in  
499 the Disease Development of Human Ferritin. *Biophysical Journal* 109:1273–1281.
- 500 30. Modi T, Risso VA, Martinez-Rodriguez S, Gavira JA, Mebrat MD, Van Horn WD, Sanchez-Ruiz  
501 JM, Banu Ozkan S (2021) Hinge-shift mechanism as a protein design principle for the evolution  
502 of  $\beta$ -lactamases from substrate promiscuity to specificity. *Nat Commun* 12:1852.
- 503 31. Campitelli P, Swint-Kruse L, Ozkan SB (2021) Substitutions at Nonconserved Rheostat  
504 Positions Modulate Function by Rewiring Long-Range, Dynamic Interactions. *Mol Biol Evol*  
505 38:201–214.
- 506 32. Ose NJ, Butler BM, Kumar A, Kazan IC, Sanderford M, Kumar S, Ozkan SB (2022) Dynamic  
507 coupling of residues within proteins as a mechanistic foundation of many enigmatic pathogenic  
508 missense variants. *PLoS Comput Biol* 18:e1010006.
- 509 33. Modi T, Ozkan SB (2018) Mutations Utilize Dynamic Allostery to Confer Resistance in TEM-1  
510  $\beta$ -lactamase. *Int J Mol Sci [Internet]* 19. Available from:  
511 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6321620/>

- 512 34. Butler BM, Gerek ZN, Kumar S, Ozkan SB (2015) Conformational dynamics of  
513 nonsynonymous variants at protein interfaces reveals disease association. *Proteins: Structure,*  
514 *Function, and Bioinformatics* 83:428–435.
- 515 35. Kolbaba-Kartchner B, Kazan IC, Mills JH, Ozkan SB (2021) The Role of Rigid Residues in  
516 Modulating TEM-1  $\beta$ -Lactamase Function and Thermostability. *International Journal of*  
517 *Molecular Sciences* 22:2895.
- 518 36. Butler BM, Kazan IC, Kumar A, Ozkan SB (2018) Coevolving residues inform protein  
519 dynamics profiles and disease susceptibility of nSNVs. *PLoS computational biology*  
520 14:e1006626.
- 521 37. Kazan IC, Sharma P, Rahman MI, Bobkov A, Fromme R, Ghirlanda G, Ozkan SB (2022) Design  
522 of novel cyanovirin-N variants by modulation of binding dynamics through distal mutations  
523 Faraldo-Gómez JD, Ben-Tal N, editors. *eLife* 11:e67474.
- 524 38. Gerek ZN, Kumar S, Ozkan SB (2013) Structural dynamics flexibility informs function and  
525 evolution at a proteome scale. *Evolutionary Applications* 6:423–433.
- 526 39. Beach H, Cole R, Gill ML, Loria JP (2005) Conservation of  $\mu$ s–ms Enzyme Motions in the Apo-  
527 and Substrate-Mimicked State. *J. Am. Chem. Soc.* 127:9167–9176.
- 528 40. Gerek ZN, Ozkan SB (2011) Change in Allosteric Network Affects Binding Affinities of PDZ  
529 Domains: Analysis through Perturbation Response Scanning. *PLOS Computational Biology*  
530 7:e1002154.
- 531 41. Modi T, Campitelli P, Kazan IC, Ozkan SB (2021) Protein folding stability and binding  
532 interactions through the lens of evolution: a dynamical perspective. *Current Opinion in*  
533 *Structural Biology* 66:207–215.
- 534 42. Campitelli P, Modi T, Kumar S, Ozkan SB (2020) The Role of Conformational Dynamics and  
535 Allostery in Modulating Protein Evolution. *Annual Review of Biophysics* 49:267–288.
- 536 43. van den Bedem H, Bhabha G, Yang K, Wright PE, Fraser JS (2013) Automated identification  
537 of functional dynamic contact networks from X-ray crystallography. *Nat Methods* 10:896–902.
- 538 44. Chan HS, Dill KA (1990) Origins of structure in globular proteins. *Proc Natl Acad Sci U S A*  
539 87:6388–6392.
- 540 45. Cao H, Wang J, He L, Qi Y, Zhang JZ (2019) DeepDDG: Predicting the Stability Change of  
541 Protein Point Mutations Using Neural Networks. *J. Chem. Inf. Model.* 59:1508–1514.
- 542 46. Wei Q, Xu Q, Dunbrack RL (2013) Prediction of phenotypes of missense mutations in human  
543 proteins from biological assemblies. *Proteins* 81:199–213.

544 47. Chakrabarty B, Parekh N (2016) NAPS: Network Analysis of Protein Structures. *Nucleic Acids*  
545 *Research* 44:W375–W382.

546 48. del Sol A, O’Meara P (2005) Small-world network approach to identify key residues in  
547 protein–protein interaction. *Proteins: Structure, Function, and Bioinformatics* 58:672–682.

548 49. del Sol A, Fujihashi H, Amoros D, Nussinov R (2006) Residue centrality, functionally  
549 important residues, and active site shape: Analysis of enzyme and non-enzyme families. *Protein*  
550 *Science* 15:2120–2128.

551 50. Ben Chorin A, Masrati G, Kessel A, Narunsky A, Sprinzak J, Lahav S, Ashkenazy H, Ben-Tal N  
552 (2020) ConSurf-DB: An accessible repository for the evolutionary conservation patterns of the  
553 majority of PDB proteins. *Protein Sci* 29:258–267.

554 51. Goldenberg O, Erez E, Nimrod G, Ben-Tal N (2009) The ConSurf-DB: pre-calculated  
555 evolutionary conservation profiles of protein structures. *Nucleic Acids Res* 37:D323–D327.

556 52. Thorpe IF, Brooks III CL (2004) The coupling of structural fluctuations to hydride transfer in  
557 dihydrofolate reductase. *Proteins: Structure, Function, and Bioinformatics* 57:444–457.

558 53. Salomon-Ferrer R, Gotz AW, Poole D, Le Grand S, Walker RC (2013) Routine microsecond  
559 molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald.  
560 *Journal of chemical theory and computation* 9:3878–3888.

561 54. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C (2015) ff14SB:  
562 improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of*  
563 *chemical theory and computation* 11:3696–3713.

564 55. Sawle L, Ghosh K (2016) Convergence of Molecular Dynamics Simulation of Protein Native  
565 States: Feasibility vs Self-Consistency Dilemma. *J. Chem. Theory Comput.* 12:861–869.

566 56. Stevens AO, Kazan IC, Ozkan B, He Y (2022) Investigating the allosteric response of the PICK1  
567 PDZ domain to different ligands with all-atom simulations. *Protein Science* 31:e4474.

568 57. Hubbard SJ, Thornton JM (1993) Naccess, Computer Program; Department of Biochemistry  
569 and Molecular Biology, University College London,.

570 58. Chakrabarty B, Parekh N (2014) PRIGSA: Protein repeat identification by graph spectral  
571 analysis. *J. Bioinform. Comput. Biol.* 12:1442009.

572