Supporting Text

Variation of the Structural Perturbation Method (δω-based method)

To assess that our major conclusion, namely functionally relevant low-frequency modes in biological nanomachines are robust to sequence variations, does not depend on the precise methodology, we have used another complementary approach that probes the response of a given mode to a residue-specific perturbation. In a previous article (1) we used this procedure to identify the network of dynamically linked residues, which relay allosteric signals in a number of DNA and RNA polymerases. Here we modify the scheme to take into account sequence variations thus exploiting evolutionary signals. Following the contact-based method, we define residue-similarity score for a substitution ($R_{i\alpha} \leftrightarrow R_{i\beta}$) at position *i* using

$$S(R_{i\alpha}, R_{i\beta}) = \log \frac{P(R_{i\alpha} \leftrightarrow R_{i\beta} \mid con(i))}{P(R_{i\alpha} \leftrightarrow R_{i\beta})}, \quad \text{(S1)}$$

which is essentially the log-probability of the substitution $R_{i\alpha} \leftrightarrow R_{i\beta}$ occurring if position *i* is conserved ("conserved" means position *i* maintains its interactions with its neighbors). As in the contact-based method we use the PAM250 score to evaluate residue similarity. The corresponding local energy variation at position *i* due to $R_{i\alpha} \leftrightarrow R_{i\beta}$ substitution is set to be $\delta\omega_1$ times the probability of the position *i* not being conserved if the substitution $R_{i\alpha} \leftrightarrow R_{i\beta}$ is allowed:

$$\begin{aligned} &\delta E_i(R_{i\alpha}, R_{i\beta}) \\ &= [1 - P(con(i) \mid R_{i\alpha} \leftrightarrow R_{i\beta})] \cdot \delta\omega_1 \quad \text{(S2)} \\ &= [1 - P(con(i)) \cdot e^{S(R_{i\alpha}, R_{i\beta})}] \cdot \delta\omega_1. \end{aligned}$$

Following the same procedure described in Eq. **6** we get

$$\delta E_i(R_{i1}, R_{i2}) = [1 - (1 - P_{rand}) \cdot e^{S(R_{i\alpha}, R_{i\beta}) - S_{rand}}] \cdot \delta\omega_1, \quad \text{(S3)}$$

where P_{rand} = 0.5. The variation of the local energy at position *i* is

$$\delta E_i \sim \left\langle \delta E_i(R_{i\alpha}, R_{i\beta}) \right\rangle_{MSA}, \quad \text{(S4)}$$

where the average is over all possible ($R_{i\alpha}, R_{i\beta}$) for the given position *i* in the MSA.

The robustness parameter $f_{\delta E}' = \frac{\delta E}{E}$ is computed using $\delta E = \sum_i \delta E_i$, $E \propto \sum_i \delta\omega_1$, where $\delta\omega_i$ is computed using Eq. **S3**.

Identification of Network of Residues ("Wiring Diagram") That Transmit Allosteric Signals

A byproduct of the SPM method is that it can be used to identify the network of distantly connected residues that cooperatively orchestrate the large-scale domain movements. By assessing the degree of response to site-specific perturbations the importance of a given residue in facilitating allosteric transitions can be gauged. Using the SPM, which can be implemented by the contact-based analysis or the δω-based method we have identified the mechanically "hot" residues for the three nanomachines.

***Thermus aquaticus* DNA polymerase I**: For the TAQ DNA polymerase I we have identified the network of residues that trigger the open/closed transition using the SPM. Several residues that span the fingers/palm domains are involved in the three modes that are required in the domain movements. Among the hot-spot residues, I614 (2) and F667 (3) were shown to be critical for fidelity of DNA replication. Suzuki *et al*. **(**4) screened 67 mutants in the O-helix (659-671) for altered fidelity of DNA synthesis. Among those, 13 mutations of residues R660, A661, A662, T664, N666, V669, L670 were found to cause decreased fidelity. In addition, R659, K663, F667, and Y671 were shown to be immutable or absolutely conserved. Kermekchiev *et al*. (5) reported four cold-sensitive mutations (E626K, W706R, I707L, E708D) and one deleterious mutation (Q690R). They are clustered on the outside surface of the enzyme's finger domain, which is far from the active site, but at the hinge point of the fingers domain. Remarkably, the SPM method identifies accurately many of the functionally relevant residues. The network of allostery transmitting residues are explicitly shown in Fig. 2.

***Dictyostelium* myosin II:** Many of the residues identified by SPM have been shown to be functionally relevant in myosin II. The amino acid sequence similarity between *Dictyostelium* myosin II and human beta-cardiac myosin have facilitated the study of the structural consequences of those hypertrophic cardiomyopathy (HCM) mutations in human beta-cardiac myosin heavy chain (6). Five HCM mutations are mapped to residues E492, F506, R695, A699, and F745 in *Dictyostelium* myosin II. The mutant A699R (7) exhibits the lowest level of force with their preserved actin-activated MgATPase activity. The E706K mutation causes a human myopathy in human skeletal myosin IIa. The homologous mutation, E683K, in the *Dictyostelium* myosin motor domain was also studied (8). The motor functions of the mutants I499A and F692A are highly compromised (9). Similarly, the rate of ATP cleavage is reduced 4-fold in the G691A mutant. The nucleotide-binding rate is 20- to 30-fold slower in G680A (10). The mutants Y494K, W501L, G740D produce a cold-sensitive phenotype in vivo (11). The mutant G680V myosin exhibits a substantially enhanced affinity for several nucleotides, decreased ATPase activity, and over-occupancy or creation of a novel strongly actin-binding state (12). The G691C mutant exhibits an elevated basal ATPase indicative of premature phosphate release (12). The mutants F481C/N483K and H484Q are functionally defective (13). *Dictyostelium* cells transformed with F487A or F506G (14) myosin were found to be unable to undergo processes that require myosin II function. The mutant W501Y's affinity for actin was shown (15) to be about 6-fold decreased, which confirmed that W501 is the primary ATP-sensitive tryptophanyl residue. Our method predicts that the above residues are critical in the functional dynamics of myosin II (Table 3). The residues, which have been found to be relevant for myosin allostery, form a sparse network (Fig. 4).

***E. coli* GroEL Chaperonin**: Our analysis correctly identifies, Y360 and D361, as being functionally relevant. Mutant Y360E has reduced ATPase activity while D361K lacks the ability to bind to the cochaperonin GroES (16). More recently, using genetic experiments, Klein and coworkers **(**17) identified two mutants (GroEL44 mutant: E191G) and (GroEL673 mutant: G173D and G337D). They found intragenic suppressor mutations at 13 residues (G173S, V174F, V174I, T176N, V189E, V190I, V190L, Q194P, Q194H, Q194H, R322G, T331S, T331S, T331A, T331N, D359Y, D359N, D359G, Q366R, K371N, G375C, V378G, and V378A) that restore growth at elevated temperatures to the mutant *E. coli* GroEL44 bacteria (17). Surprisingly, all 13 amino acids are positioned in such a way as to allow for large *en bloc* allosteric movement in GroEL so that it can correctly interact with GroEL. Twelve of these 13 residues are also identified using our method as being allosterically relevant. Murai *et al*. cross-linked the equatorial and apical domains through a disulfide in the D83C/K327C double mutant (with additional mutations C138S, C458S, and C519S) (18). This mutant can still bind ATP, which remains bound without hydrolysis. However, it fails to process further reaction steps such as ATP hydrolysis, binding of GroES, dissociation of substrate protein from GroEL, and facilitating protein folding. Xu *et al.* (19) observed that the G375 residue plays an important role in the rotation of

the apical domain, as does G192. Furthermore, G192 and G375 are absolutely conserved. The wiring diagram for allostery is explicitly shown in Fig. 6.

References

1**.** Zheng, W., Brooks, B. R., Doniach, S. & Thirumalai, D. (2005) *Structure (London)* **13**, 565-577.

2. Patel, P. H., Kawate, H., Adman, E., Ashbach, M. & Loeb, L. A. (2000) *J. Biol. Chem*. **276**, 5044-51.

3. Suzuki, M., Yoshida, S., Adman, E. T., Blank, A. & Loeb, L. A. (2000) *J. Biol. Chem*. **275**, 32728-32735.

4. Suzuki, M., Avicola, A. K., Hood, L. & Loeb, L. A. (1997) *J. Biol. Chem*. **272**, 11228-11235.

5. Kermekchiev, M.B., Tzekov, A. & Barnes, W. M. (2003) *Nucleic Acids Res*. **31**, 6139-6147.

6. Hypertrophic cardiomyopathy mutations from the mutation database at www.angis.org.au/Databases/Heart.

7. Fujita, H., Sugiura, S., Momomura, S., Sugi, H. & Sutoh, K. (1998) *Adv. Exp. Med. Biol*. **453**, 131-137.

8. Zeng, W., Conibear, P. B., Dickens, J. L., Cowie, R.A., Wakelin, S., Malnasi-Csizmadia, A. & Bagshaw, C. R. (2004) *Philos. Trans. R. Soc. London Ser. B* **359**, 1843-1855.

9. Patterson, B. & Spudich, J. A. (1996) *Genetics* **143**, 801-810.

10. Sasaki, N., Ohkura, R. & Sutoh, K. (2003) *Biochemistry* **42**, 90-95.

11. Batra, R., Geeves, M. A. & Manstein, D. J. (1999) *Biochemistry*. **38**, 6126-6134.

12 Patterson, B., Ruppel, K. M., Wu, Y. & Spudich, J. A. (1997) *J. Biol. Chem*. **272**, 27612-27617.

13. Ruppel, K. M. & Spudich, J. A. (1996) *Annu. Rev. Cell Dev. Biol*. **12**, 543-573.

14. Tsiavaliaris, G., Fujita-Becker, S., Batra, R., Levitsky, D. I., Kull, F. J., Geeves, M. A. & Manstein, D. J. (2002) *EMBO Rep*. **3**, 1099-1105.

15. Batra, R. & Manstein, D. J. (1999) *J. Biol. Chem*. **380**, 1017-23.

16. Fenton, W. A., Kashi, Y., Furtak, K. & Horwich, A. L. (1994) *Nature* **371**, 614-619.

17. Klein, G. & Georgopoulos, C. (2001) *Genetics* **158**, 507-517.

18. Murai, N., Makino, Y. & Yoshida, M. (1996) *J. Biol. Chem.* **271**, 28229-28234.

19. Xu, Z., Horwich, A. L. & Sigler, P. B**.** (1997) *Nature* **388**, 741-750