

Coarse-Grained Model of Entropic Allostery

Rhoda J. Hawkins* and Tom C.B. McLeish

*IRC in Polymer Science and Technology, Department of Physics and Astronomy,
University of Leeds, Leeds LS2 9JT, United Kingdom*

(Dated: June 9, 2004)

Many signalling functions in molecular biology require proteins to bind to substrates such as DNA in response to environmental signals such as the simultaneous binding to a small molecule. Examples are repressor proteins which may transmit information via a conformational change in response to the ligand binding. An alternative entropic mechanism of “allostery” suggests that the inducer ligand changes the intramolecular vibrational entropy, not just the mean static structure. We present a quantitative, coarse-grained model of entropic allostery, which suggests design rules for internal cohesive potentials in proteins employing this effect. It also addresses the issue of how the signal information to bind or unbind is transmitted through the protein. The model may be applicable to a wide range of repressors and also to signalling in trans-membrane proteins.

Introduction. It is becoming increasingly clear that dynamics, as well as static structure, are important in molecular biology. For example simulations of dynamical transitions in proteins [1] suggest that collective global modes are correlated with protein function. This letter focuses on repressor proteins which bind to DNA to “turn off” genes when the cell does not require their expression. The binding is “allosteric”: it is activated depending on the presence of inducer ligands, small molecules which themselves bind to the protein at a site distant from the active site. The “holorepressor” (“aporepressor”) is the protein with (without) a bound ligand. In allosteric repressor proteins the ligand binding site is distant from that of the DNA. For this reason ligand binding has often been assumed to cause a conformational change within the repressor protein, decreasing its affinity for DNA in one state compared to the other state. However, this is evidence that dynamically induced entropic changes may contribute to allostery [2–4].

A classic example of a repressor system is the E-coli lac repressor [5–9]. In this case the aporepressor binds to DNA, suppressing the genes for the metabolism of lactose. A second example is the E-coli trp repressor [10–12] that, on binding, prevents the expression of the gene for tryptophan synthesis. But in contrast to the lac, trp-type holorepressor proteins bind to DNA and the aporepressors do not. There are many such repressor systems but the lac and the trp will act as representative cases for this Letter. Our challenge is to explore whether the Brownian fluctuations in protein structure may carry information between the two binding sites, thereby producing cooperative lac-type or trp-type behaviour. This mechanism of cooperativity is one of the key questions in understanding protein function [13].

As generally true for the functional roles of protein dynamics, the lower-frequency softer modes will dominate [14]. Although higher modes are more numerous, they are spatially localised due to elastic disorder [15]. Ligand binding at sites where high frequency modes have significant amplitude will therefore generally have only

local effects; long distant allosteric signalling will be exponentially suppressed beyond the localisation length of the mode. Focusing on the slower, global modes additionally motivates a spatially coarse-grained model.

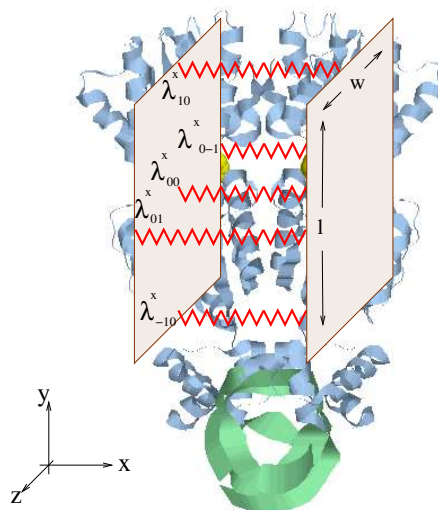


FIG. 1: Plates and springs model for the interaction of the two domains of a repressor dimer. The x-ray structure (PDB 1EFA [7]) of the lac repressor dimer (with DNA at the bottom) is shown behind the model.

A coarse-grained model. In a coarse-grained representation we model a repressor protein dimer as two rigid plates of length l and width w , representing the two protein monomers motivated by the common dimer motif (see Figure 1). We parameterise the relative motion of the plates by three relative translation displacements (x, y, z) and three rotation angles $(\theta_x, \theta_y, \theta_z)$. The stabilising contact interaction between the two dimers is characterised by 5 quadratic (3D) potential wells. Figure 1 shows 5 of the effective springs that arise, λ_i^x which are perpendicular to the plates. We find that a minimal model requires just 4 other springs in the plane of the plates (summations of diagonal springs between

the plates), which we label as $\lambda_{00}^y, \lambda_{0-1}^z, \lambda_{00}^z, \lambda_{01}^z$, representing the resolved y and z relative displacements. These local interactions represent a minimal set of “sticky patches” which could arise from hydrophobic, side chain, or electrostatic forces. The corresponding spring constants λ_i could be calculated in principle from the details of these interactions. We allow only local modifications of the contact interactions on binding. The springs local to the ligand binding site will be affected by the binding of an inducer or corepressor. (Lac actually binds two inducer ligands, which we simplify here with a single bound state.) Similarly springs local to the DNA binding site will be affected by binding to DNA. The other springs act as anchoring potentials.

We consider vibrations of the plates in 3 different planes: (i) in the x, y plane (translational vibrations along the x -axis and rotations about the z -axis); (ii) in the x, z plane (translation along the x -axis and rotation about the y -axis); and (iii) in the y, z plane (translation along the z -axis and rotation about the x -axis).

The Tirion potential [16] replaces the full MD potential with a simple pairwise quadratic potential of universal strength and this is found to be sufficient to describe the low frequency modes involving coherent motion of large groups of atoms. In our model we similarly look only at these low frequency modes and describe them by harmonic potentials between the protein domains, but allow the potentials between the protein monomers to acquire locally specific values.

As an example of detailed calculations we take the 2×2 system of motions in the y, z plane. We write $z_{01} = z_{00} - \frac{l\theta_x}{2}$ and $z_{0-1} = z_{00} + \frac{l\theta_x}{2}$ (where θ_x is the angle of rotation about the x -axis) to obtain the Hamiltonian in terms of the mutual translational vibration coordinate z_{00} and mutual rotational vibration coordinate θ_x .

$$H = \frac{1}{2} \mathbf{p} \mathbf{M}^{-1} \mathbf{p} + \frac{1}{2} \mathbf{x} \mathbf{K} \mathbf{x} \quad (1)$$

where the interaction matrix for the y, z plane

$$\mathbf{K} = \begin{pmatrix} -(\lambda_{01}^z + \lambda_{0-1}^z + \lambda_{00}^z) & \frac{1}{2}(\lambda_{01}^z - \lambda_{0-1}^z) \\ \frac{1}{2}(\lambda_{01}^z - \lambda_{0-1}^z) & -\frac{1}{4}(\lambda_{01}^z + \lambda_{0-1}^z) \end{pmatrix},$$

the inertial matrix

$$\mathbf{M} = \begin{pmatrix} m & 0 \\ 0 & I/l^2 \end{pmatrix}, \quad \text{and} \quad \mathbf{x} = \begin{pmatrix} z_{00} \\ l\theta_x \end{pmatrix}$$

where m is the reduced mass and I is the reduced moment of inertia of the dimer pair. This leads to the partition function in the relevant classical limit

$$Z = \int \dots \int e^{-\frac{H(x_{0_i}, \theta_i)}{kT}} dx_{0_i} d\theta_i = \frac{2\pi kT}{(|\mathbf{M}^{-1}| |\mathbf{K}|)^{1/2}}. \quad (2)$$

Finally from (2) the entropy of the protein dimer for a single plane is

$$S = Nk(\ln 2\pi kT \sqrt{mI}/l + 1 - 1/2 \ln(\lambda_1 \lambda_{-1} + \lambda_0(\lambda_1 + \lambda_{-1}))/4) \quad (3)$$

where $\lambda_{0_i}^z$ has been abbreviated to λ_i for convenience. We are interested in the difference between the change in entropy on binding at the DNA binding site of the two cases in which the protein is, and is not also bound to the inducer. We call this $\Delta\Delta S = \Delta S_{\text{holo}} - \Delta S_{\text{apo}}$. A result with $\Delta\Delta S \neq 0$ would signify cooperative behaviour i.e. the binding to DNA is affected by the binding to the inducer. We write $\Delta\Delta S$ in terms of dimensionless spring constants $\tilde{\lambda}_1 = \frac{\lambda_1}{\lambda_0}$ and $\tilde{\lambda}_{-1} = \frac{\lambda_{-1}}{\lambda_0}$ and bound to unbound ratios $\Lambda_1 = \frac{\lambda_{1B}}{\lambda_1}$ and $\Lambda_{-1} = \frac{\lambda_{-1B}}{\lambda_{-1}}$. This gives us

$$\Delta\Delta S = \frac{1}{2} Nk \ln \left(\frac{(4\Lambda_1 + \frac{\Lambda_1}{\tilde{\lambda}_{-1}} + \frac{1}{\tilde{\lambda}_1})(4\Lambda_{-1} + \frac{\Lambda_{-1}}{\tilde{\lambda}_1} + \frac{1}{\tilde{\lambda}_{-1}})}{(4\Lambda_1\Lambda_{-1} + \frac{\Lambda_1}{\tilde{\lambda}_{-1}} + \frac{\Lambda_{-1}}{\tilde{\lambda}_1})(4 + \frac{1}{\tilde{\lambda}_{-1}} + \frac{1}{\tilde{\lambda}_1})} \right). \quad (4)$$

The other modes can be modelled in the same way to give additional contributions to $\Delta\Delta S$. The 2-plate model generates a 3×3 form of \mathbf{M} and \mathbf{K} for coupled rotations about y and z and translations along x plus one simple y -translation.

We take $N = 2$ for the lac since it is a tetramer of two dimers and similarly for the trp since it represses as a dimer of dimers so also has two dimers.

$\Delta\Delta S > 0$ gives the trp case whereby the affinity for the holorepressor binding to DNA is greater than the apo repressor. $\Delta\Delta S < 0$ however gives the lac case since the apo lac repressor is the one with the higher affinity for DNA. Applying these inequalities to equation (4) gives the following rule determining which case arises.

$$\Lambda_1 \Lambda_{-1} + 1 \begin{cases} > \\ < \end{cases} \Lambda_1 + \Lambda_{-1} \quad \begin{cases} \text{trp} \\ \text{lac} \end{cases}. \quad (5)$$

The trp case occurs when both spring constants increase ($\Lambda_1 > 1, \Lambda_{-1} > 1$) or decrease ($\Lambda_1 < 1, \Lambda_{-1} < 1$) on ligand binding. The lac case occurs when one spring constant increases and the other decreases ($\Lambda_1 > 1$ and $\Lambda_{-1} < 1$ or $\Lambda_1 < 1$ and $\Lambda_{-1} > 1$). Figure 2 plots the function $\Delta\Delta S(\Lambda_1, \Lambda_{-1})$ (equation 4). Biologically relevant values for the original spring constants were chosen using protein B -factor data (related to the RMS positions of the atoms $B = 8\pi^2 \langle u^2 \rangle > [17]$) and steered simulations (see later) giving the case for a potential which is stronger at the inducer binding site ($\tilde{\lambda}_1 > \tilde{\lambda}_{-1}$). It can be seen that the negative $\Delta\Delta S$ lac-type effect is maximised when the ligand binding at the inducer binding site decreases the spring constant $\tilde{\lambda}_1$ but $\tilde{\lambda}_{-1}$ increases on DNA binding. The trp-type effect (positive $\Delta\Delta S$) is maximised when both spring constants decrease as much as possible on binding. This reduction will however be limited by the physical requirement for overall stability of the complex. Requiring that the RMS displacement of the monomers be less than the average separation of the atoms leads to the estimation that $\lambda_i > 0.1kT\text{\AA}^{-2}$.

Physically, “entropic allostery” allows the lac inducer binding to communicate via the large amplitudes of the

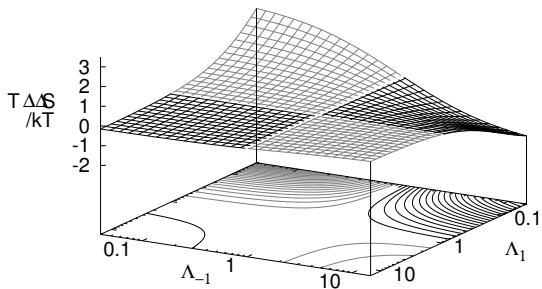


FIG. 2: Graph showing $\Delta\Delta S$ against Λ_1 and Λ_{-1} . The contour plot on the base shows regions where lac-type (black) and trp-type (grey) behaviour is optimised.

internal modes of the protein to the “read-head” binding regions near the DNA which as a result move too much to be inserted into the DNA.

Estimating the spring constants. We evaluate our model in the real example of the lac repressor as an illustration. Firstly we converted the protein data for the B -factors [7, 8, 18] into RMS vibration values and estimated the spring constants λ_i and Λ_i (in this case averaged over the vibrations in the different planes) from the expression $\lambda \sim 1 / \langle \bar{u}^2 \rangle$ giving $\tilde{\lambda}_1 \approx 1.2$, $\tilde{\lambda}_{-1} \approx 0.1$ (estimated from [19]), $\Lambda_1 \approx 0.07$ and $\Lambda_{-1} \approx 6.7$ (so supporting our prediction that the lac case has $\Lambda_1 < 1$ and $\Lambda_{-1} > 1$). We then calculated an estimate of $\Delta\Delta S$ using equation (4). Including a factor of three due to the three planes of vibration, we obtain for our plate-dimer model a value of $T\Delta\Delta S \sim -1.4kT$. Since the experimental values for the change in binding energy between holo and aporepressor binding to DNA are $\Delta\Delta G \sim 6kT$ [20–22] this indicates that the entropic contribution is likely to be significant since the crystal dynamics can only be a lower bound for the amplitudes of vibration.

To improve upon this rough estimate we calculated the interaction energy between the lac monomers in a fully atomistic computation using the software “cns” [23] and steering the relative positions of the two monomers. We used the x-ray crystal structure coordinates PDB ID 1LBI [7] for the aporepressor and 1TLF [18] for the holorepressor. By relative translation along the 3 axes, and rotation about axes at the extremities of the protein dimer, and recalculating the total interaction energy at each increment, we were able to build up curves for the potential energy wells (see Figure 3). By curve fitting the bottom of these wells to a quadratic (to fit with our harmonic approximation) we were able to extract the curvature and therefore the spring constants for each global mode of both the holo and aporepressors. We then used these to calculate $\Delta S_{\text{inducerbind}}$ which is an

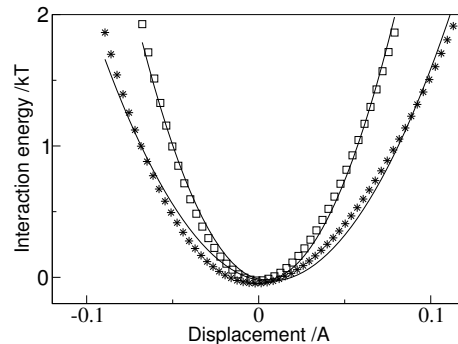


FIG. 3: Graph showing an example of a change in spring constant with and without the inducer for the lac repressor. The mode shown is rotation about the y axis (rotation point $(0,0,20)\text{\AA}$). The squares are cns data for 1LBI the aporepressor and the stars are cns data for 1TLF the holorepressor. The lines are quadratic fits.

estimate for $-\Delta\Delta S$ in the case of large Λ_{-1} . For the motion in the x, y and x, z planes (translation along x and rotation about y and z) this gave $T\Delta\Delta S \sim -1.66kT$, for the y, z plane $-0.42kT$ and for the relative translation along y (which does not have rotation modes coupled to it) $-0.28kT$. Interestingly the softest mode, contributing most to the allostery, is the one which shifts the DNA read heads (which point in the plane perpendicular to the core) away from the DNA perpendicular to it (the x, z plane). Therefore in total we have an estimate for $\Delta\Delta S \sim -2.36kT$.

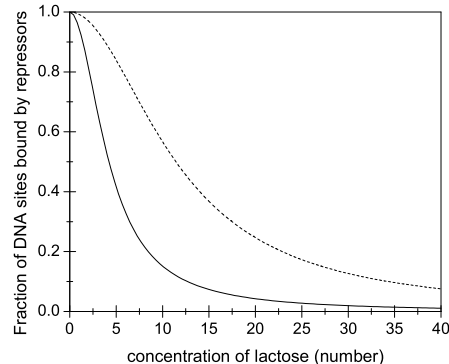


FIG. 4: Graph showing the fraction of DNA sites bound by repressors against the concentration of inducer lactose. The solid curve includes the entropic component and the dashed curve is without the entropic component.

To check how significant this entropy contribution is to the total $\Delta\Delta G \sim 6kT$ we compared the probability of genes being repressed (bound by repressor) against lactose concentration with and without this entropic contribution (following Yildirim [24])(see Figure 4). For 95%

activation (operators not bound) 18 lactose molecules are required with the entropic contribution but 50 would be required if there was no entropic contribution. Note there are only of order 100 repressors in the cell [5]. This would imply that the entropic contribution to allostery is significant in controlling the lactose level at which the gene expression is turned on.

Conclusions. We conclude that inducer binding affects the Brownian motion within the repressor protein and this entropic effect contributes to the allosteric mechanism in DNA binding proteins alongside static conformational changes. We can relate the communication of the signal across the protein to “design rules” for the potentials within it.

Several extensions of this approach suggest themselves. A discrete many-springs model naturally extends to the case of a continuous potential between the plates. Secondly, the case of multiple, sequential ligand binding will lead to additional structure. To make the model even more realistic we should also include bending modes of the protein monomers themselves. Significantly, addition of such bending modes must increase the predicted $\Delta\Delta S$ value if the ligand binding changes the bending rigidity. For lac any such increase might give values of $\Delta\Delta S$ that actually dominate the binding free energy.

We expect this model can be also applied to trans-membrane proteins that transmit signals across membranes into cells and organelles [25, 26]. These systems are similar in that an inducer ligand (e.g. adrenaline) binds to the receptor trans-membrane protein which in turn allows it to bind or unbind proteins on the interior of the membrane (e.g. to bind to a G-protein in the control cycle for glycogen). Within the restricted environment of the membrane we expect entropic allostery to play an important role in the transmission of the signal through the receptor protein.

The calculated values of the contribution to the free energy change from the change in intramolecular vibrational entropy of the protein easily reach the order of a few kT per molecule, within the experimentally observed range for these systems.

We thank the EPSRC for funding, and P. G. Stockley, S. E. V. Phillips, S. W. Homans, R. F. Bruinsma and D. T. F. Dryden for stimulating discussions.

* Electronic address: rhoda.hawkins@physics.org

[1] A. L. Tournier and J. C. Smith, Phys. Rev. Lett. **91**, 208106 (2003).

- [2] C. E. Bell and M. Lewis, Current Opinion in Structural Biology **11**, 19 (2001).
- [3] S. Jusuf, P. J. Loll, and P. H. Axelsen, Journal of the American Chemical Society **125**, 3988 (2003).
- [4] A. Cooper and D. T. F. Dryden, European Biophysics Journal **11**, 103 (1984).
- [5] R. F. Bruinsma, Physica A **313**, 211 (2002).
- [6] B. Alberts, *Molecular Biology of the cell* (Garland Science, 2002), 4th ed.
- [7] M. Lewis, G. Chang, N. C. Horton, M. A. Kercher, H. C. Pace, M. A. Schumacher, R. G. Brennan, and P. Lu, Science **271**, 1247 (1996), -PDB ID: 1LBI.
- [8] C. E. Bell and M. Lewis, Nature Structural Biology **7**, 209 (2000), -PDB ID: 1EFA.
- [9] A. L. Lehninger, D. L. Nelson, and M. M. Cox, *Lehninger Principles of Biochemistry* (Worth, 2000), 3rd ed., biochemistry in 3D Tutorial website www.worthpublishers.com/lehninger3D.
- [10] L. Otwinowski, R. W. Schevitz, R.-G. Zhang, C. L. Lawson, A. Joachimak, R. Q. Marmorstein, B. F. Luisi, and P. Sigler, Nature **335**, 321 (1988).
- [11] C. L. Lawson and J. Carey, Nature **366**, 178 (1993), -PDB ID 1TRR.
- [12] D. Zhao, C. H. Arrowsmith, X. Jia, and O. Jardetzky, Journal of molecular biology **229**, 735 (1993), -PDB ID: 1WRS and 1WRT.
- [13] K. S. Matthews, Science **271**, 1245 (1996).
- [14] I. Bahar, A. R. Atilgan, M. C. Demirel, and B. Erman, Phys. Rev. Lett. **80**, 2733 (1998).
- [15] C. Micheletti, G. Lattanzi, and A. Maritan, J. Mol. Biol. **321**, 909 (2002).
- [16] M. M. Tirion, Phys. Rev. Lett. **77**, 1905 (1996).
- [17] G. Rhodes, *Crystallography made crystal clear* (Academic Press Inc., San Diego, California, 1993).
- [18] A. M. Friedman, T. O. Fischmann, and T. A. Steitz, Science **268**, 1721 (1995), -PDB ID: 1TLF.
- [19] M. Slijper, R. Boelens, A. L. Davis, R. N. H. Konings, G. A. van der Marel, J. H. van Boom, and R. Kaptein, Biochemistry **36**, 249 (1997), -PDB ID: 1LQC.
- [20] M. D. Barkley and S. Bourgeois, *The Operon* (Cold Spring Harbour, 1980), chap. 7.
- [21] N. Horton, M. Lewis, and P. Lu, Journal of molecular biology **265**, 1 (1997).
- [22] M. M. Levandoski, O. V. Tsodikov, D. E. Frank, S. E. Melcher, R. M. Saecker, and M. T. Record Jr, Journal of Molecular Biology **260**, 697 (1996).
- [23] A. Brunger, P. Adams, G. Clore, W. Delano, P. Gros, R. Grosse-Kunstleve, J.-S. Jiang, J. Kuszewski, N. Nilges, N. Pannu, et al., ACTA CRYST. **D54**, 905 (1998).
- [24] N. Yildirim and M. C. Mackey, Biophys. J. **84**, 2841 (2003).
- [25] A. M. Spiegel, P. S. Backlund Jr, J. E. Butrynski, T. L. Jones, and W. F. Simonds, TIBS **16** (1991).
- [26] L. Swint-Kruse, H. L. Zhan, B. M. Fairbanks, A. Maheshwari, and K. S. Matthews, Biochemistry **42**, 14004 (2003).