

Electronic Supporting Information

Ligand-Regulated Oligomerisation of Allosterically Interacting Proteins

C. Schaefer,^{1,2,*} R. A. J. de Bruijn,^{1,3} and T. C. B. McLeish^{1,2}

¹*Department of Physics, Durham University,
South Road, Durham, DH1 3LE, UK*

²*Current address: Department of Physics,
University of York, Heslington, York, YO10 5DD, UK*

³*Current address: Theory of Polymers and Soft Matter,
Eindhoven University of Technology, P.O. Box 513,
5600 MB Eindhoven, The Netherlands*

Abstract

This Supplementary Information provides three calculations (section I-III) that we have used in the main text, as well as a calculation (section IV) that demonstrates the wider applicability of our approach. In section I, we show how the partition function of a chain of length N can be expressed as a function of the determinant of the Hessian matrix (of size $N \times N$) of the Hamiltonian. In section II, we provide an analytical expression for this determinant for arbitrary chain length N and an arbitrary number of ligand N_A bound to the chain, for any distribution of these ligands on the chain. In section III we calculate the configuration average of the determinant. Finally, in section IV we add additional degrees of freedom to the chain (representing shape fluctuation of the monomers and bond-length fluctuations between the monomers), and show that the resulting entropic penalty is additive to the penalty we already had in sections I-III.

*Electronic address: charley.schaefer@york.ac.uk

I. RELATION BETWEEN PARTITION FUNCTION AND DETERMINANT

The partition function in the main text, given by

$$Z_N(N_e) = \sum_{\eta_1=0}^1 \cdots \sum_{\eta_i=0}^1 \cdots \sum_{\eta_N=0}^1 \delta \left(\sum_{\eta_i=1}^N \eta_i - N_e \right) \int d\mathbf{x} \exp(-\beta\mathcal{H}(\mathbf{x}; \vec{\eta}, \alpha, K_d, \epsilon, \epsilon_e, \mathbf{x})), \quad (\text{S.1})$$

can be integrated, i.e., $\int d\mathbf{x} \exp(-\beta\mathcal{H}) = e^{-\beta\epsilon(N-1)} e^{-\beta\epsilon_e N_e} \sqrt{\pi^N / \beta^N \kappa^N \det \hat{H}_N(\vec{\eta}, \alpha, K_d)}$. We cast it in the usual form [1, 2]

$$Z_N(N_e) = Z_1^N(0) W_N(N_e) e^{-\beta\epsilon(N-1)} W_e^{N_e} e^{-\beta\epsilon_e N_e} \quad (\text{S.2})$$

with $Z_1(0) = \sqrt{\pi/\beta\kappa}$ the partition function of an inactive monomer. Further $W_e \equiv \alpha^{-1/2}$ measures the entropic penalty of monomer activation and

$$W_N(N_e) \equiv \alpha^{N_e/2} \binom{N}{N_e} \left\langle \left(\det \hat{H}_N(\vec{\eta}, \alpha, K_d) \right)^{-1/2} \right\rangle, \quad (\text{S.3})$$

measures the entropic penalty of polymerisation. In this expression, $\binom{N}{N_e} = N!/(N-N_e)!N_e!$ is the number of chain configurations and $\langle \cdot \rangle$ is the average over all chain configurations.

Below we show that this expression may be approximated by

$$W_N(N_e) \approx \alpha^{N_e/2} \binom{N}{N_e} \frac{1}{\sqrt{\sum_{k=1}^N \langle d_k^N(N_e, \alpha) \rangle K_d^{k-1}}}, \quad (\text{S.4})$$

where $\langle d_k^N(N_e, \alpha) \rangle$ are configuration-averaged polynomial coefficients that can be calculated analytically.

II. EXACT CALCULATION OF DETERMINANT

In this section, we calculate the determinant of the Hessian matrix in the theory section of the main text. This is a symmetric $N \times N$ tridiagonal matrix with arbitrary real (positive) values at the diagonal. The values at the diagonal are given by $\alpha_i + K_d$ for $i = 1, N$ and by $\alpha_i + 2K_d$ for $1 < i < N$. At the first upper and lower diagonals the matrix elements are given by $-K_d$. For general N the determinant may be calculated analytically using the method reported in Refs. [3, 4], where the solution is given in terms of products of the matrix elements. We are however primarily interested in the expansion of this determinant in the form of a polynomial of α and K_d , i.e., we seek the form

$$\det \hat{H}_N(\vec{\eta}, \alpha, K_d) \equiv \sum_{k=1}^N d_k^N(\vec{\eta}, \alpha) K_d^{k-1}, \quad \text{with } d_k^N \equiv \sum_{l=0}^N b_l \alpha^l \quad (\text{S.5})$$

Our task is hence to calculate the coefficients d_k^N . While we previously reported the solution for $\alpha = 1$ in Ref. [2], we now present the general case where α can take any real non-zero value.

The determinant of the $N \times N$ symmetric tridiagonal matrix, $\det \hat{H}_N$, can be calculated using a set of recurrence relations given by

$$p_1 = \alpha_1 + K_d, \quad (\text{S.6})$$

$$p_2 = (\alpha_1 \alpha_2) + (2\alpha_1 + \alpha_2)K_d + K_d^2, \quad (\text{S.7})$$

$$p_n = (\alpha_n + 2K_d)p_{n-1} - K_d^2 p_{n-2}, \text{ for } 2 < n < N, \quad (\text{S.8})$$

$$\det \hat{H}_N = (\alpha_N + K_d)p_{N-1} - K_d^2 p_{N-2}. \quad (\text{S.9})$$

For a monomer ($N = 1$) and a dimer ($N = 2$) the determinant is $\det \hat{H}_1 = \alpha_1$ and $\det \hat{H}_2 = \alpha_1 \alpha_2 + (\alpha_1 + \alpha_2)K_d^2$, respectively.

For oligomers ($N > 2$), the polynomial form in Eq. (S.5) can be obtained by writing $p_n = \sum_{l=1}^{n+1} c_l^n K_d^{l-1}$ and inserting that into the recurrence relations above, to give

$$p_n = \alpha_n c_1^{n-1} + (\alpha_n c_2^{n-1} + 2c_1^{n-1})K_d + \sum_{l=3}^n (\alpha_n c_l^{n-1} + 2c_{l-1}^{n-1} - c_{l-2}^{n-2})K_d^{l-1} + (2c_n^{n-1} - c_{n-1}^{n-2})K_d^n, \quad (\text{S.10})$$

for $2 < n < N$. In general, c_l^n is a linear combination of all permutations of the product $[\alpha_{j_1} \alpha_{j_2} \dots \alpha_{j_k}]$. This product consists of $k = n + 2 - i$ different monomers selected from the chain of length N .

Finally, from the recurrence relations it now follows that the determinant of the Hessian matrix is given by

$$\det \hat{H}_N = \alpha_N c_1^{N-1} + (\alpha_N c_2^{N-1} + c_1^{N-1})K_d + \sum_{l=3}^N (\alpha_N c_l^{N-1} + c_{l-1}^{N-1} - c_{l-2}^{N-2})K_d^{l-1}. \quad (\text{S.11})$$

Note that $c_{n+1}^n = 1$ for all n and hence that the term of order K_d^N equals zero. Comparison of Eq. (S.11) to Eq. (S.5) yields the polynomial coefficients d_k^N we were after.

For long-chain polymers we approximate this determinant by

$$\det \hat{H}_N = w^{(1-\phi)(N-1)} (K_d) \alpha^{\phi N} w^{\phi(N-1)} (K_d/\alpha), \quad (\text{S.12})$$

where $\phi \equiv N_e/N$ is the fraction of activated monomers in the chain and $w(x) \equiv \frac{1}{2} + x + \frac{1}{2}\sqrt{1+4x}$ describes the reduction of internal degrees of freedom of a homopolymer

upon monomer addition in the long-chain limit [2]. In this case, Eq. (S.4) reduces to the approximation

$$\Delta S/k_B(N-1) = -\phi \ln \phi - (1-\phi) \ln(1-\phi) + \frac{1}{2}(1-\phi) \ln w(K_d) + \frac{1}{2}\phi \ln w(K_d/\alpha) \quad (\text{S.13})$$

given in the theory section in the main text.

III. APPROXIMATION: CONFIGURATION AVERAGE OF DETERMINANT

As shown above, the coefficient $d_k^N(\vec{\eta}, \alpha)$ is a weighted sum of all possible permutations of a product of $n = N + 1 - k$ different α_j s from the diagonal of the Hessian matrix \hat{H} . In this section, we will approximate this coefficient by the average $d_k^N(\vec{\eta}) \approx \langle d_k^N(\vec{\eta}) \rangle \equiv d_k^N(N_e)$ that can be obtained from the analytical solution for $N_e = 0$, provided in Ref. [2] and implicitly summarised in this Supplementary Information. In general, these coefficients can be written as

$$d_k^N(\vec{\eta}, \alpha) = \sum_{j_1=1}^k \alpha_{j_1} \sum_{j_2=j_1+1}^{k+1} \alpha_{j_2} \cdots \sum_{j_n=j_{n-1}+1}^N \alpha_{j_n} \xi_{j_1 j_2 \dots j_n}(\vec{\eta}). \quad (\text{S.14})$$

The weight, $\xi_{j_1 j_2 \dots j_n}$, of each permutation is independent of the values of any α_j , but in principle depends on the configuration, $\vec{\eta}$, of the full polymer chain.

In the following, we neglect these configuration-dependent variations and assume $\xi_{j_1 j_2 \dots j_n} \approx \langle \xi_{j_1 j_2 \dots j_n} \rangle$, where $\langle \cdot \rangle$ is the average over all chain configurations. We discuss the validity and implications of this approximation at the end of this section. The analytical solution for the configuration average can be calculated by considering a homopolymer, i.e., for $N_e = 0$,

$$d_k^N(\vec{0}, \alpha) = \sum_{j_1=1}^k \sum_{j_2=j_1+1}^{k+1} \cdots \sum_{j_n=j_{n-1}+1}^N \xi_{j_1 j_2 \dots j_n} = \frac{N!}{n!(N-n)!} \langle \xi_{j_1 j_2 \dots j_n} \rangle. \quad (\text{S.15})$$

Hence, the result $\langle \xi_{j_1 j_2 \dots j_n} \rangle$ is given in terms of the known homopolymer coefficients $d_{k,0}^N \equiv d_k^N(\vec{0})$.

Using this result for the homopolymer we generate the polynomial coefficients for copolymer of active and inactive monomers,

$$\langle d_k^N(N_e) \rangle \approx d_{k,0}^N \frac{n!(N-n)!}{N!} \sum_{j_1=1}^k \cdots \sum_{j_n=j_{n-1}+1}^N [\alpha_{j_1} \dots \alpha_{j_n}] = d_{k,0}^N \langle [\alpha_{j_1} \dots \alpha_{j_n}] \rangle. \quad (\text{S.16})$$

The average $\langle [\alpha_{j_1} \dots \alpha_{j_n}] \rangle$ is a polynomial in α of order $\min\{n, N_e\}$. Its polynomial coefficient for a term of order $n_e = 0, 1, \dots, n$ represents the probability, $P(n, n_e; N, N_e)$, that a product $[\alpha_{j_1} \dots \alpha_{j_n}]$ is of order α^{n_e} . Hence, it follows that Eq. (S.16) can be rephrased as

$$\langle d_k^N(N_e) \rangle = d_{k,0}^N \sum_{n_e=n_{A,\min}(k)}^{n_{A,\max}(k)} P(N+1-k, n_e; N, N_e) \alpha^{n_e}, \quad (\text{S.17})$$

with $n_{A,\min}(k) = \max\{1, N_e + 1 - k\}$ and with $n_{A,\max}(k) = \min\{N + 1 - k, N_e\}$. The probability is given by the probability that if n monomers are randomly, but uniquely, selected from $N - N_e$ inactive and N_e active monomers. Hence, the probability is given by

$$P(n, n_e; N, N_e) = \binom{n}{n_e} \binom{N-n}{N_e-n_e} \left[\binom{N}{N_e} \right]^{-1}. \quad (\text{S.18})$$

Note that this probability is proportional to the number of permutations of the n_e active monomers in a sequence of n monomers, and to the number of permutations of the remaining $N_e - n_e$ in a sequence of $N - n$ monomers. The total number of permutations $\binom{N_e}{N}$ of the N_e monomers in a sequence of length N normalises the probability distribution.

We have verified the validity of the approximate configuration average by comparing it to the numerical configuration average. The latter we have obtained by averaging over various randomly generated permutations of copolymers. In Figure 2 of the main text, we have presented the exact solutions for the individual chains (black lines) and the approximate ensemble average (red lines) for chains of $N = 10$ monomers with all possible numbers of ligands bound to them ($N_e = 0, \dots, 10$). For weak ($K_d \ll 1$) and strong ($K_d \gg \sqrt{N}$) allosteric coupling the approximation is exact, implying that, in absence of energetic preferences, the ligands are randomly distributed for length scales smaller than the allosteric interaction range \sqrt{N} . This implies that our approximation is accurate for chains with ligand-distributions that are uniform at length scales of $> \sqrt{N}$ monomers, but is inaccurate for ‘block copolymers’ with significant deviations in ligand distribution at those length scales. The latter type of chains is, due to the absence of enthalpic allostery, rare in the present work, and do therefore not affect the results presented in the main text.

IV. EFFECT OF (ANISOTROPIC) SHAPE FLUCTUATIONS

In this work we have used the allosteron model to describe the direct coupling of the internal modes of monomers, giving rise to an entropic penalty. What has been excluded in

this model, is the indirect coupling of the internal modes because of shape fluctuations. In Figure S.1 we show that anisotropic shape fluctuations are in principle coupled to the internal modes of the monomers. The entropies of direct and indirect coupling of the internal modes turn out to be additive, which implies that indirect couplings do not alter our conclusions in the main text.

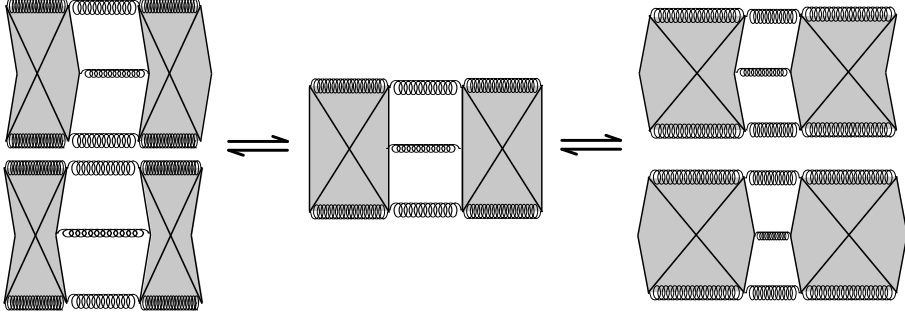


FIG. S.1: Schematic examples of the coupling between coarse-grained bending modes in the allosteron model for proteins.

In order to show this, we describe shape changes by extending the Hamiltonian in the main text with additional degrees of freedom $x_{n,n+1}$ that represent the indirect shape coupling between monomers n and $n + 1$. In general, the monomers are anisotropic and the two different monomer-monomer binding sites may be characterised with two different form factors γ_1 and γ_2 . Depending on γ_1 and γ_2 , which may be either be positive or negative, the internal modes of the monomers may be correlated or anti correlated. The contribution by indirect coupling to the Hamiltonian is given by

$$\mathcal{H}_i = \kappa_i \sum_{n=1}^{N-1} (x_{n,n+1} - \gamma_1 x_n - \gamma_2 x_{n+1})^2. \quad (\text{S.19})$$

Hence, the entropic part of the Hamiltonian becomes $\mathbf{x}'^T \kappa \hat{H}'_N \mathbf{x}'$, with $\mathbf{x}' = [x_1, x_2, \dots, x_N, x_{1,2}, \dots, x_{(N-1)N}]^T$ and

$$\hat{H}'_N = \left(\begin{array}{c|c} \hat{H}_N + A & B \\ \hline B^T & D \end{array} \right) \quad (\text{S.20})$$

a block matrix within which \hat{H}_N is the original $N \times N$ square matrix in the main text, and A represents the indirect coupling between the internal degrees of freedom of the monomers. This is again a tridiagonal matrix, with at the main diagonal $A_{1,1} = \gamma_1^2 K_i$, $A_{N,N} = \gamma_2^2 K_i$ and

$A_{n,n} = (\gamma_1^2 + \gamma_2^2)K_i$ for $1 < n < N$. Here, $K_i \equiv \kappa_i/\kappa$ is the dimensionless indirect-coupling constant. At the first upper and lower diagonals, $A_{m,n} = \gamma_1\gamma_2K_i$. The matrix D provides the contribution by the $x_{m,m+1}x_{n,n+1}$ terms. Since only square terms are present, D is a $(N-1) \times (N-1)$ diagonal matrix with K_i as diagonal elements. Finally, B generates the $x_mx_{n,n+1}$ cross terms. Here, $B_{m,m} = -\gamma_1K_i$ and $B_{m-1,m} = -\gamma_2K_i$.

The determinant of the dimensionless Hessian matrix is therefore $\det(\hat{H}'_N) = \det(D)\det(\hat{H}_N + A - BD^{-1}B^T)$, which we find equals $\det(\hat{H}'_N) = \det(D)\det(\hat{H}_N) = K_i^{N-1}\det(\hat{H}_N)$. Hence, the (anisotropic) shape fluctuations provide an entropic penalty $\Delta S = (1/2)k_B \ln(\beta\kappa_i/\pi)$ that is independent of both the chain length and the binding of ligands.

Acknowledgments

This work was funded by the Engineering and Physical Sciences Research Council [grant number EP/N031431/1].

-
- [1] P. van der Schoot, *Theory of supramolecular polymerization* (CRC Press, Boca Raton, 2005), 2nd ed.
- [2] T. McLeish, C. Schaefer, and A. C. von der Heydt, *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **373** (2018), ISSN 0962-8436, <http://rstb.royalsocietypublishing.org/content/373/1749/20170186.full.pdf>, URL <http://rstb.royalsocietypublishing.org/content/373/1749/20170186>.
- [3] H. A. Yamani and M. S. Abdelmonem, *J. Phys. A: Math. Gen.* **30**, 2889 (1997).
- [4] G. Y. Hu and R. F. O'Connell, *J. Phys. A: Math. Gen.* **29**, 1511 (1996).