**SUPPORTING INFORMATION S1**

**Rate theory**

In this section we will show more explicitly the calculations to obtain the expression for the rates in Eqs. 6. We start by substituting Eq. (5) into the expression in Eq. (3) and the equivalent for B we obtain

\[
\begin{align*}
r_{A \rightarrow B} &= \frac{\int e^{-\beta E(A,s)} \theta \left[ \frac{P(B,s)}{P(A,s)} - 1 \right] ds}{\int \int dsdc e^{-\beta E(c,s)}} \\
r_{B \rightarrow A} &= \frac{\int e^{-\beta E(B,s)} \theta \left[ \frac{P(A,s)}{P(B,s)} - 1 \right] ds}{\int \int dsdc e^{-\beta E(c,s)}}
\end{align*}
\]

which can be rewritten by further substituting the expression of the probabilities \(P(A,s)\) and \(P(B,s)\) from Eq. (5), where we have substituted the sum with an integral over the sequence space.

\[
\begin{align*}
r_{A \rightarrow B} &= \frac{\int e^{-\beta E(A,s)} \theta \left[ e^{-\beta(E(B,s) - E(A,s))} - 1 \right] ds}{\int \int dsdc e^{-\beta E(c,s)}} = \frac{\int e^{-\beta E(A,s)} \theta \left[ -E(B,s) + E(A,s) \right] ds}{Z} \\
r_{B \rightarrow A} &= \frac{\int e^{-\beta E(B,s)} \theta \left[ e^{-\beta(E(A,s) - E(B,s))} - 1 \right] ds}{\int \int dsdc e^{-\beta E(c,s)}} = \frac{\int e^{-\beta E(B,s)} \theta \left[ -E(A,s) + E(B,s) \right] ds}{Z}
\end{align*}
\]

**Detail Balance**

At this stage we can check that we obey detailed balance to do that we will use an approximate form for the Heaviside function that is particularly good for large \(\beta\) or small temperatures

\[
\theta(\beta x) \approx \frac{1}{1 + e^{-\beta x}}
\]

If we substitute Eq. (S5) into Eq. (S3)

\[
\begin{align*}
r_{A \rightarrow B} &= \frac{\int e^{-\beta E(A,s)} \frac{1}{1 + e^{\beta E(B,s) - E(A,s)}} ds}{Z} \\
r_{B \rightarrow A} &= \frac{\int e^{-\beta E(B,s)} \frac{1}{1 + e^{\beta E(A,s) - E(B,s)}} ds}{Z}
\end{align*}
\]

if we rearrange the exponentials

\[
\begin{align*}
r_{A \rightarrow B} &= \frac{\int e^{-\beta E(A,s)} e^{-\beta E(B,s)} \frac{1}{e^{-\beta E(B,s)} + e^{-\beta E(A,s)}} ds}{Z} \\
r_{B \rightarrow A} &= \frac{\int e^{-\beta E(B,s)} e^{-\beta E(A,s)} \frac{1}{e^{-\beta E(A,s)} + e^{-\beta E(B,s)}} ds}{Z}
\end{align*}
\]

It is easy to see that Eq. (S8) and Eq. (S9) are identical and hence detailed balance is respected.
Metropolis acceptance

If we want to sample the distribution $P(A)P(B)$ then we can use the Metropolis method. We start from the detailed balance equation:

$$P(\text{Old})\Pi(\text{Old} \rightarrow \text{New}) = P(\text{New})\Pi(\text{New} \rightarrow \text{Old})$$  \hspace{1cm} (S10)

where $P(\text{Old}) = P_A(O)P_B(O)$ and $P(\text{New}) = P_A(N)P_B(N)$

$$\frac{\Pi(\text{Old} \rightarrow \text{New})}{\Pi(\text{New} \rightarrow \text{Old})} = \frac{P_A(N)P_B(N)}{P_A(O)P_B(O)}$$  \hspace{1cm} (S11)

If we now remember that $P_A$ and $P_B$ are Boltzmann distributed then we can calculate the transition probabilities and hence our acceptance rule

$$\frac{\Pi(\text{Old} \rightarrow \text{New})}{\Pi(\text{New} \rightarrow \text{Old})} = \frac{e^{-\beta E_A(N)}/Q_A e^{-\beta E_B(N)}/Q_B}{e^{-\beta E_A(O)}/Q_A e^{-\beta E_B(O)}/Q_B}$$  \hspace{1cm} (S12)

So if we group the exponents together we get that the final acceptance rule depends on the variation of the sum of Hamiltonians at every step.

$$\frac{\Pi(\text{Old} \rightarrow \text{New})}{\Pi(\text{New} \rightarrow \text{Old})} = e^{-\beta \Delta E_{\text{All}}}.$$  \hspace{1cm} (S13)

Coarse-grained Monte Carlo simulations

The caterpillar model [13] is a 5-bead model with the $C_\alpha$ augmented by the full main atomic positions to introduce directional hydrogen bonds. The degrees of freedom of the model are the torsional angles $\phi$ and $\psi$; all other structural parameters are kept fixed at values from the literature [1]. The C, O, N, H positions were determined from the $C_\alpha$ (see Fig.1 in Ref [13]). The side chain interactions are represented by and effective $C_\alpha$-$C_\alpha$ sphere-sphere interaction energy given by

$$E_{ij}(r_{ij}) = \alpha \varepsilon_{ij} \left[ 1 - \frac{1}{1.0 + e^{2.5[1\text{Å]}(r_{\text{max}}-r_{ij})}} \right]$$  \hspace{1cm} (S14)

where $r_{ij}$ is the distance between the $C_\alpha$ atoms at the centers of spheres $i$ and $j$ and $r_{\text{max}}$ ($r_{\text{max}} = 12\text{Å}$) is the distance at which $E_{ij} = \varepsilon_{ij}/2$; $\alpha = 1/4$ is a scale factor. This expression provides a continuous square well form for the sphere-sphere interaction energy. To determine the parameter $\varepsilon_{ij}$ we made use of the model of Betancourt and Thirumalai (BT) [2], in which the interaction energies were derived from a calculation of the contact frequency in the PDB. This potential had been used primarily for lattice proteins, but it is
also appropriate for the caterpillar model, which employs a square-well-like potential. Backbone hydrogen bonds were modeled with a 10-12 Lennard-Jones type potential using the expression \[ E_{H} = -\varepsilon_{H} (\cos \theta_1 \cos \theta_2)^{\nu} \left[ 5 \left( \frac{\sigma}{r_{OH}} \right)^{12} - 6 \left( \frac{\sigma}{r_{OH}} \right)^{10} \right] \] where \( r_{OH} \) is the distance between the hydrogen atom of the amide group (NH) and the oxygen atom of the carboxyl group (CO) of the main chain. We set \( \sigma = 2.0 \, \text{Å}, \, \varepsilon_{H} = -3.1 \, k_{B} T, \) and \( \nu = 2; \) the values are given in [3].

**VIRTUAL-MOVE PARALLEL TEMPERING**

The VMPT scheme is a combination of the adaptive parallel tempering algorithm [4] and the waste recycling method developed by Frenkel [23]. This method consists in using the information about the reject moves to generate a bias potential that helps the system visit regions of the phase space that are otherwise difficult to reach. In the conventional MCMC method all information about rejected trial moves is discarded. Recently Frenkel has proposed a scheme that makes it possible to include the contributions of rejected configurations in the sampling of averages [23]. In the standard parallel tempering (PT) scheme, we only retain information about PT moves that have been accepted. However, in the spirit of Ref. [23], we can include the contribution of all PT trial moves, irrespective of whether they are accepted. The weight of the contribution of such a virtual move is directly related to its acceptance probability. For instance, if we use the symmetric acceptance rule for MC trial moves, then the weights of the original and new (trial) state in the sampling of virtual moves are given by

\[
P_{N} = \frac{e^{\Delta \beta \Delta E_{O-N} + \Delta W_{O-N}}}{1 + e^{\Delta \beta \Delta E_{O-N} + \Delta W_{O-N}}},
\]

\[
P_{O} = \frac{1}{1 + e^{\Delta \beta \Delta E_{O-N} + \Delta W_{O-N}}},
\]

where \( W \) is a bias potential. We are not limited to a single trial swap of state \( i \) with a given state \( j \). Rather, we can include all possible trial swaps between the temperature state \( i \) and all \( N - 1 \) remaining temperatures. Our estimate for the contribution to the probability distribution \( P_{i} \) corresponding to temperature \( i \) is then given by the following sum

\[
P_{i}(Q) = \sum_{j=1}^{N-1} \left( \frac{1}{1 + e^{\Delta \beta_{ij} \Delta E_{ij} + \Delta W_{ij}}} \right) \delta(Q_{j} - Q) + \sum_{j=1}^{N-1} \left( \frac{e^{\Delta \beta_{ij} \Delta E_{ij} + \Delta W_{ij}}}{1 + e^{\Delta \beta_{ij} \Delta E_{ij} + \Delta W_{ij}}} \right) \delta(Q - Q_{j}),
\]
where the delta functions select the configurations with order parameter $Q$. The combination with the parallel tempering is particularly efficient because the information about the sampled states is shared between all the simulations running at different temperatures, that naturally will tend to explore different regions of the phase space. In the case of proteins it means that at low temperature it is possible to know what bias is needed to reach the unfolded state because that is sampled more often at higher temperatures.

REFERENCES

[1] Creighton (1992) Proteins. W.H.Freeman & Co Ltd 2nd Revised edition.

[2] Betancourt M, Thirumalai D (1999) Pair potentials for protein folding: Choice of reference states and sensitivity of predicted native states to variations in the interaction schemes. Protein Science 8: 361369.

[3] Irbäck A, Sjunnesson F, Wallin S (2000) Three-helix-bundle protein in a ramachandran model. P Natl Acad Sci USA 97: 13614-13618.

[4] Faller R, Yan Q, de Pablo J (2002) Multicanonical parallel tempering. J Chem Phys 116: 54195423.