Variational Identification of Markovian Transition States

SUPPLEMENTARY INFORMATION

Linda Martini,[1,*] Adam Kells,[1,*] Roberto Covino,[2] Gerhard Hummer,[2,3] Nicolae-Viorel Buchete,[4] and Edina Rosta[1,**]

[1]*Department of Chemistry, King's College London, SE1 1DB, London, UK*
[2]*Department of Theoretical Biophysics, Max Planck Institute for Biophysics, 60438, Frankfurt am Main, Germany*
[3]*Institute of Biophysics, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany*
[4]*School of Physics and Institute for Discovery, University College Dublin, Dublin 4, Ireland*

Date: May 17, 2017
[*]*Equal contribution*
[**]*E-mail: edina.rosta@kcl.ac.uk*

I. FLUX THROUGH AND AROUND THE TS

The probability of the flux that goes from A to B via the TS or avoiding the TS can be considered using the following simple kinetic network:

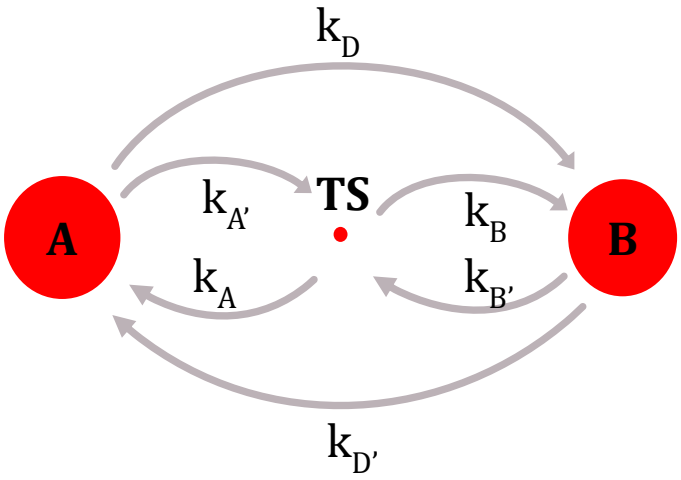


Figure S1. Kinetic scheme used to calculate the flux that flows through the TS (via k_A, $k_{A'}$, k_B, and $k_{B'}$) as compared to around the TS directly (k_D and $k_{D'}$).

It is easy to show that the probability to reach B directly from A, without going through the TS is $\frac{k_D\left(k_A+k_B\right)}{k_Dk_A+k_Dk_B+k_A,k_B}$ for the kinetic scheme of Fig. S1. Once an optimal discretization with a TS is found with our approach, the probability is calculated by subsequently coarse-graining the optimal Markov matrix into 3 states based on the ordering of the states and using e.g., the Hummer and Szabo coarse grained rate matrix (HS method) [1].

II. 3-STATE COARSE GRAINING

To demonstrate that a coarse graining that includes at least one TS-like state is a feasible alternative to finding only metastable states, we considered the theoretical case, for a linear chain of microstates (e.g., Fig. 1A), to obtain a 3-state reduced rate matrix by minimizing the magnitude of the resulting ν_2 second eigenvalue (with $\nu_1 = 0$), or alternatively, maximizing the relaxation time $t_2 = 1/|\nu_2|$. Note that we distinguish the eigenvalues of a rate matrix (ν_i) from the eigenvalues of the corresponding Markov transition probability matrix at lagtime τ (denoted by $\lambda_i = e^{-\tau\nu_i}$).

The nonzero second eigenvalue of the reduced matrix is given by

$$(1) \qquad -\nu_2 = \frac{1}{2}\left[K_{12} + K_{21} + K_{23} + K_{32} - \sqrt{\left(K_{12} + K_{21} - K_{23} - K_{32}\right)^2 + 4K_{21}K_{23}} \right] .$$

The magnitude of ν_2 can be minimized by minimizing all K_{ij} rates (i.e., the first four terms in Eq. (1)) using appropriate boundaries in the full dimensional microstates, which corresponds to metastable states with slow interconversion between states (leading to solutions as e.g., in Fig. 2c-e). Alternatively, one can also maximize the square root term, which corresponds to large K_{21} and K_{23} rates (as e.g., in Fig. 2a-b). In this case, the relaxation rate can be approximated as:

$$(2) \qquad -\nu_2 \approx \frac{1}{2}\left[K_{12} + K_{21} + K_{23} + K_{32} - \left(K_{21} + K_{23}\right)\left(1 + \frac{\left(K_{21} - K_{23}\right)\left(K_{12} - K_{32}\right)}{\left(K_{21} + K_{23}\right)^2} \right) \right]$$

$$(3) \qquad -\nu_2 \approx \frac{1}{2}\left[K_{12} + K_{32} - \frac{\left(K_{21} - K_{23}\right)\left(K_{12} - K_{32}\right)}{\left(K_{21} + K_{23}\right)} \right]$$

This latter solution can thus provide an optimal choice that minimizes the magnitude of ν_2, leading to Eq. (9) in the main text. This solution for the minimization problem to obtain optimal boundary positions thus defines a 3-state coarse-grained system with two metastable states 1 and 3, as well as a short-lived, TS-like state 2 with large out-going rates to both directions.

III. 1D ANALYTICAL POTENTIAL

We analyzed the 1D analytical potential (Fig. 1) of the form:

(4) $$F(x) = \alpha \cdot 2\sin\left(x - \pi\right) + (1-\alpha) \cdot 2\sin\left[\left(x - \pi\right)/2\right]$$

with $F(x)$ in kcal/mol units, $x \in \left[0, 5\pi\right]$. The free energy profile was discretized into $N_\mu = 150$ bins (unless otherwise stated). The kinetic rates of the full system correspond to nearest neighbor jumps between the discrete bins, as also described in the main text, with the rate matrix elements given by:

(6a) $$K_{i,i+1} = A\exp\left[\frac{F\left(x_i\right) - F\left(x_{i+1}\right)}{2k_B T}\right]$$

(5b) $$K_{i,i} = -\sum_{\substack{j=1 \\ j\neq i}}^{N_\mu} K_{i,j} \ .$$

And all other rates are zero. The Arrhenius prefactor is A=10, k_B is Boltzmann's constant, and T is the absolute temperature with $1/\beta = k_B T = 0.596$ kcal/mol corresponding to T=300K.

Analysis of the Optimal Two-State Coarse-Grained System

The optimal boundary positions were calculated for an asymmetric two-state system, the 1D free energy profile corresponding to $\alpha = 0$ (Fig. S2) as described above. Identical boundary positions were obtained for the optimal Markov matrices using long lagtimes ($\tau \geq 100$), or, equivalently, the HS discretization. Therefore, here we show the boundary positions for the optimal two- and three-state discretizations, calculated with the HS coarse-grained rates (Fig. S2A). Additionally, we also show how the second eigenvalue varies at different boundary positions for a two-state discretization (Fig. S2B). Note that the exact eigenvalue is approximated better and better at long lagtimes, and will match it at infinitely long lagtimes.

To test how the optimal width of the TS changes with respect to various parameters, we performed several benchmarking calculations. We systematically changed the lagtime and determined the optimal three-state discretization using both the LE approximation at a wide range of lagtimes (0.01-1000), and also using the lagtime-independent HS discretization (Fig. S3). We found that the optimal width of the TS converged to a well-defined value at long lagtimes, which coincides exactly with the optimal TS obtained using the HS discretization.

We also calculated the optimal TS using different extent of discretizations of the 1D profile by changing N_μ (Fig. S4). We found that as long as the discretization is fine enough, the optimal width of the TS is invariant of the discretization level.

The only parameters that does have a significant effect on the width of the TS is the temperature (Fig. 5S). As expected, as the temperature increases, the role of a barrier is less important in the dynamics of the system, and the optimal TS broadens. Therefore, our algorithm leads to an optimal coarse graining invariant of most of the parameters, except for the shape of the free energy profile in reduced units.

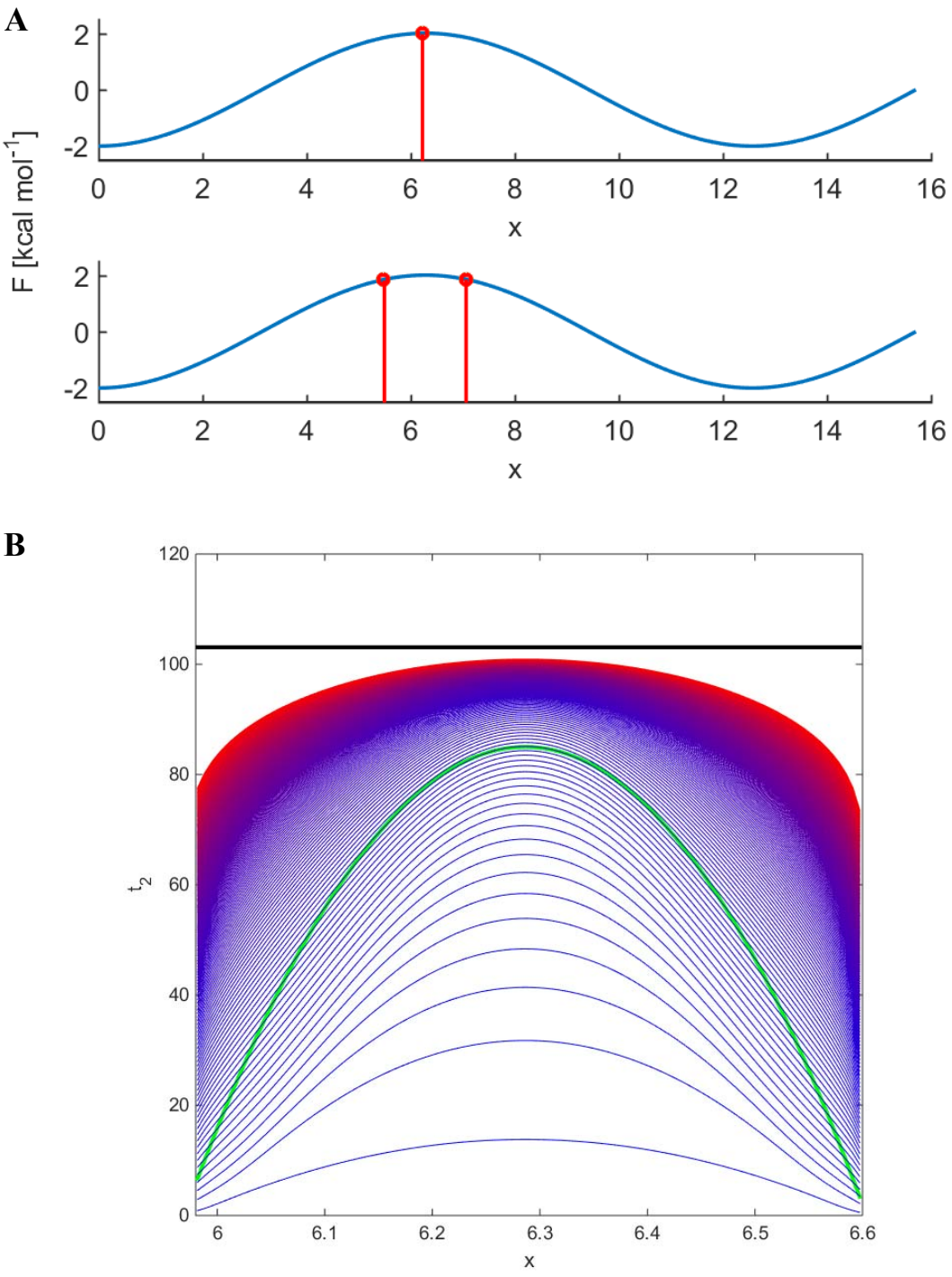


Figure S2. Optimal discretization of an asymmetrical two-state system. **A.** The red lines and symbols indicate the locations of optimal boundaries for $M = 2$ and 3 states (from top to bottom) using the HS coarse-grained rates. **B.** Variation of the relaxation time t_2 from the coarse-grained two-state systems ($M = 2$) as a function of the boundary position along x. The relaxation time was calculated from the second eigenvalues of the HS coarse rate matrix (green), or from the

reduced Markov matrix using the LE method at various lagtimes ranging from 1 (blue) to 1000 (red). The exact relaxation time of the full system is also shown as an upper bound (black line).

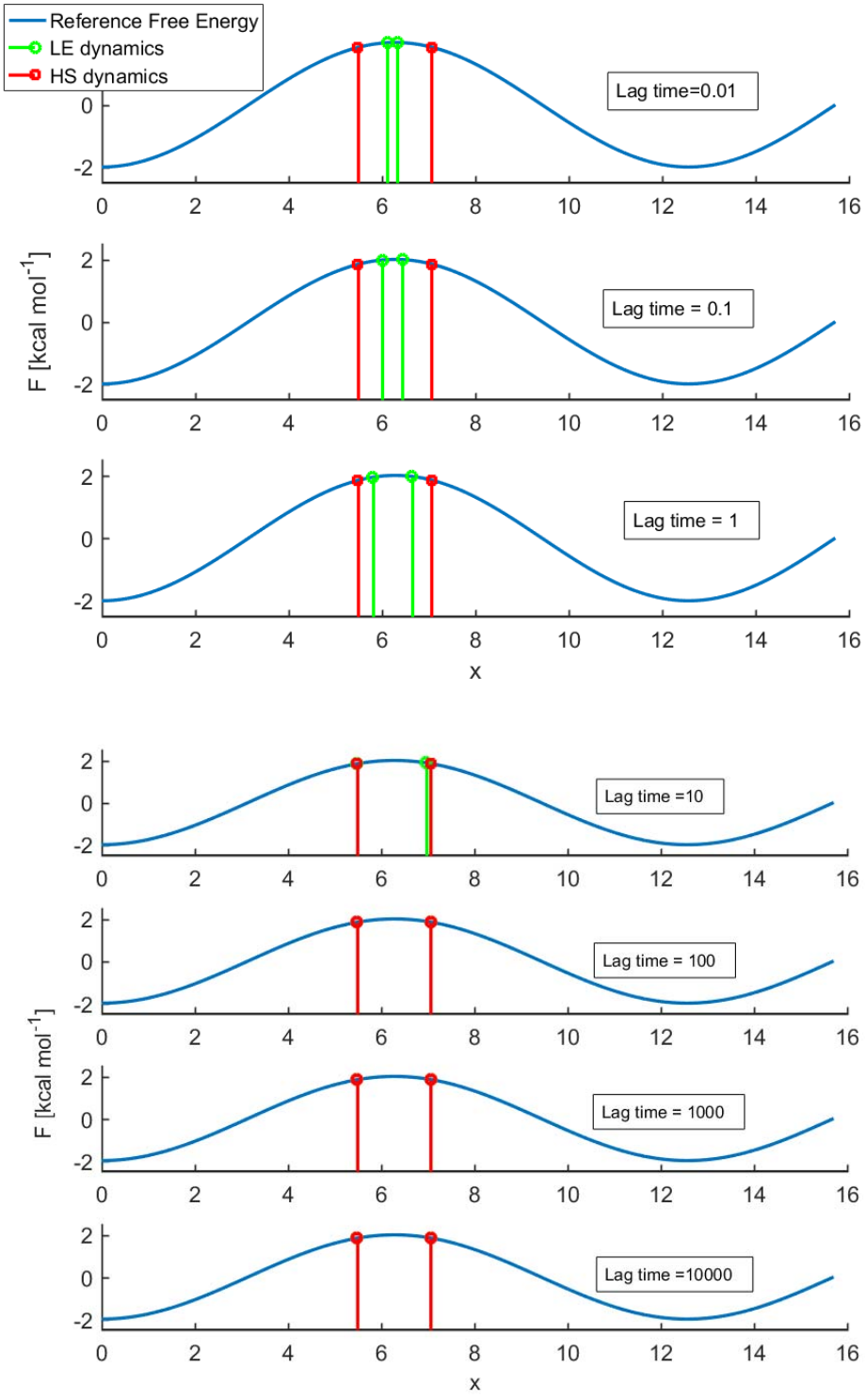


Figure S3. Optimal discretization of an asymmetrical two-state system into three states. The optimal boundary locations were determined using the HS coarse grained rates (red lines and symbols), and the LE method (green lines and symbols) for various lagtimes (0.01, 0.1, 1, 10, 100, 1000, and 10000, from top to bottom, respectively).

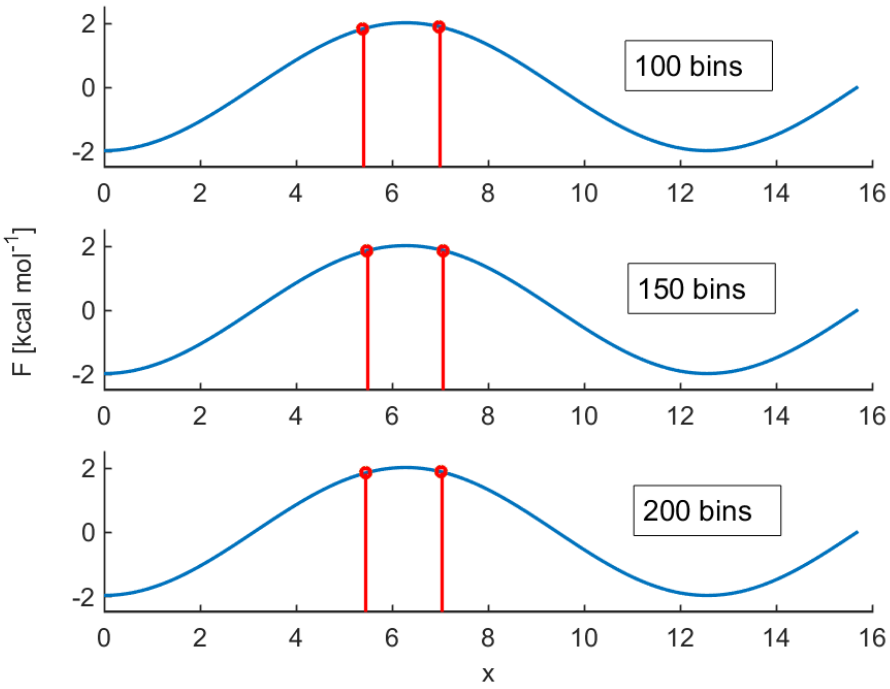


Figure S4. Optimal discretization of an asymmetrical two-state system into three states. The red lines and symbols indicate the locations of optimal boundaries using the HS coarse grained rates with 100, 150, and 200 bins from top to bottom, respectively.

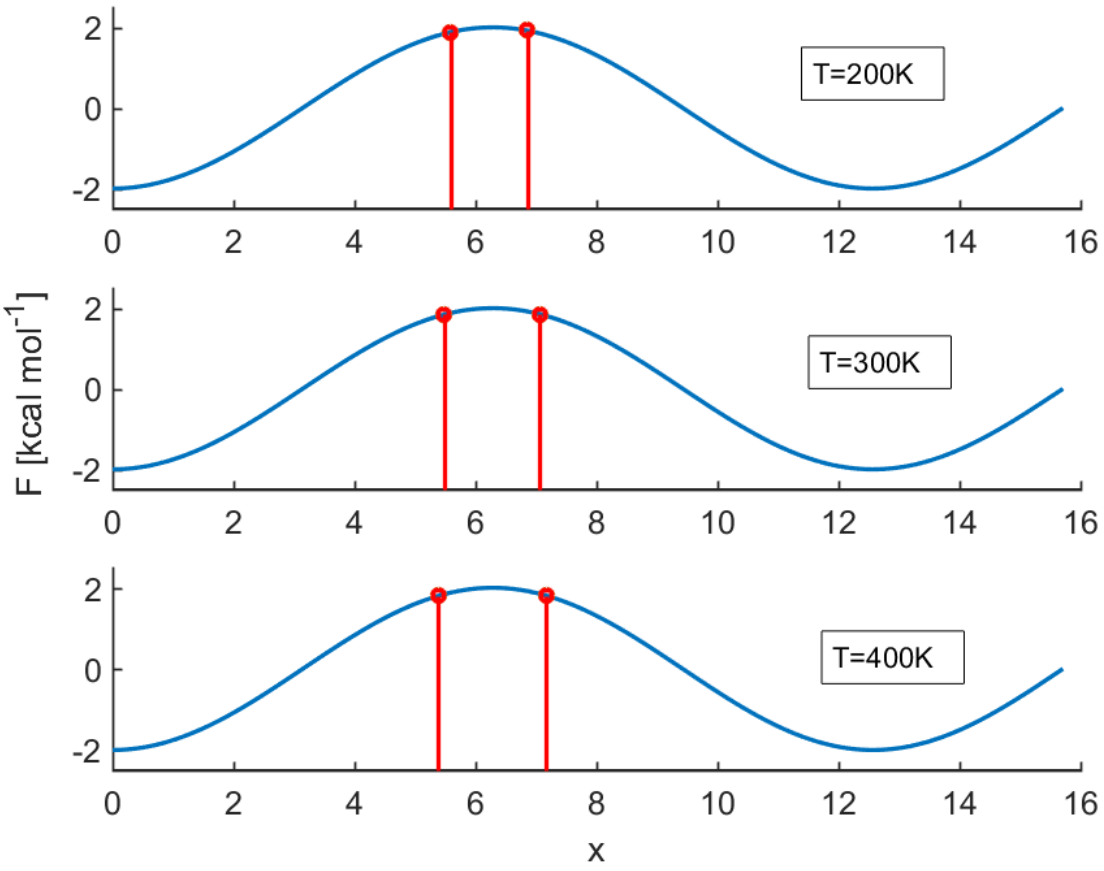


Figure S5. Optimal discretization of an asymmetrical two-state system into three states. The red lines and symbols indicate the locations of optimal boundaries using the HS coarse grained rates at 200, 300, and 400 K temperature, from top to bottom, respectively.

Optimal Three-State Coarse-Grained System

We calculated the optimal boundary positions for 1, 2, and 3 boundaries for a three-state system corresponding to the free energy profile of $\alpha = 0.5$ using the HS method (Fig. S6). Analogously, we obtain the same boundaries at $\tau \geq 100$. We can calculate the corresponding optimal Markov matrices at $\tau = 100$ and identify a TS after introducing the third boundary for $M = 4$:

$$M_2 = \begin{bmatrix} 0.9922 & 0.0078 \\ 0.0064 & 0.9936 \end{bmatrix}$$

$$M_3 = \begin{bmatrix} 0.9916 & 0.0072 & 0.0012 \\ 0.0553 & 0.5095 & 0.4352 \\ 0.0011 & 0.0526 & 0.9463 \end{bmatrix}$$

$$M_4 = \begin{bmatrix} 0.9912 & 0.0009 & 0.0068 & 0.0012 \\ 0.4923 & 0.0030 & 0.3111 & 0.1937 \\ 0.0522 & 0.0042 & 0.5074 & 0.4362 \\ 0.0011 & 0.0003 & 0.0523 & 0.9463 \end{bmatrix}.$$

The red numbers in M_4 indicate the probability to exit the TS (state 2) to states 1, 2, 3, and 4, with about 50%-50% probability to be on the left or right hand side after $\tau = 100$ time. We note that an appropriate choice for a lagtime varies, and it should be short enough not to reach full equilibrium, but long enough to result in the same consistent limit for optimal coarse states calculated with both the LE and HS kinetics.

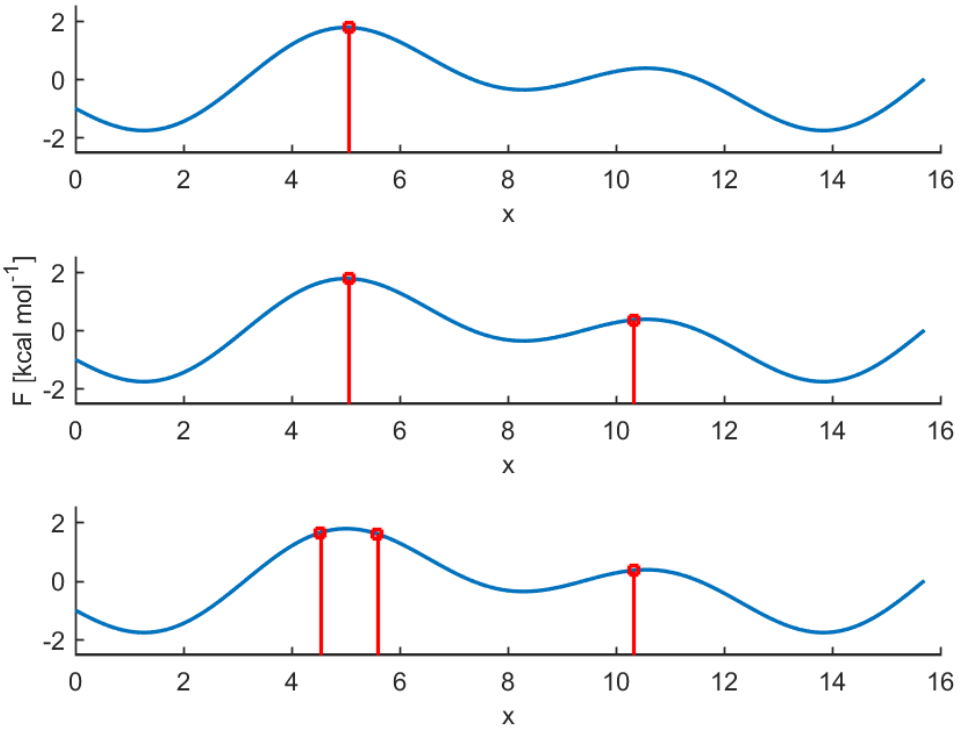


Figure S6. Optimal discretization of an asymmetrical three-state system into two, three, and four states (from top to bottom). The red lines and symbols indicate the locations of optimal boundaries using the HS coarse-grained rates. The free energy profile corresponds to $\alpha = 0.5$.

Optimal Coarse-Graining into Three States

We calculated the optimal boundary positions using 2 boundaries for 5 different free energy profiles to illustrate the variations using different coarse grained dynamics. We set the α parameter given in eq. (4) to five different values equally spaced in the interval $[0,1]$ (Figs. S7a to S7e), to monotonically tune the number of metastable states from 2 ($\alpha = 0$, Fig. S7a) to 3 ($\alpha = 1$, Fig. S7e). Accordingly, the optimal coarse graining results in either three MSs (here the fourth state that is not shown is a TS, see Fig. S6), or in two MSs and one TS state. Depending on the lagtime used for determining the coarse-grained Markov matrix, the optimal solution is slightly different. At short lagtimes (grey symbols and lines) the two-state system is more optimal, and the width of the TS approaches zero (limit not shown, see Fig. S3). At longer lagtimes, the optimal LE boundaries coincide with the HS boundaries.

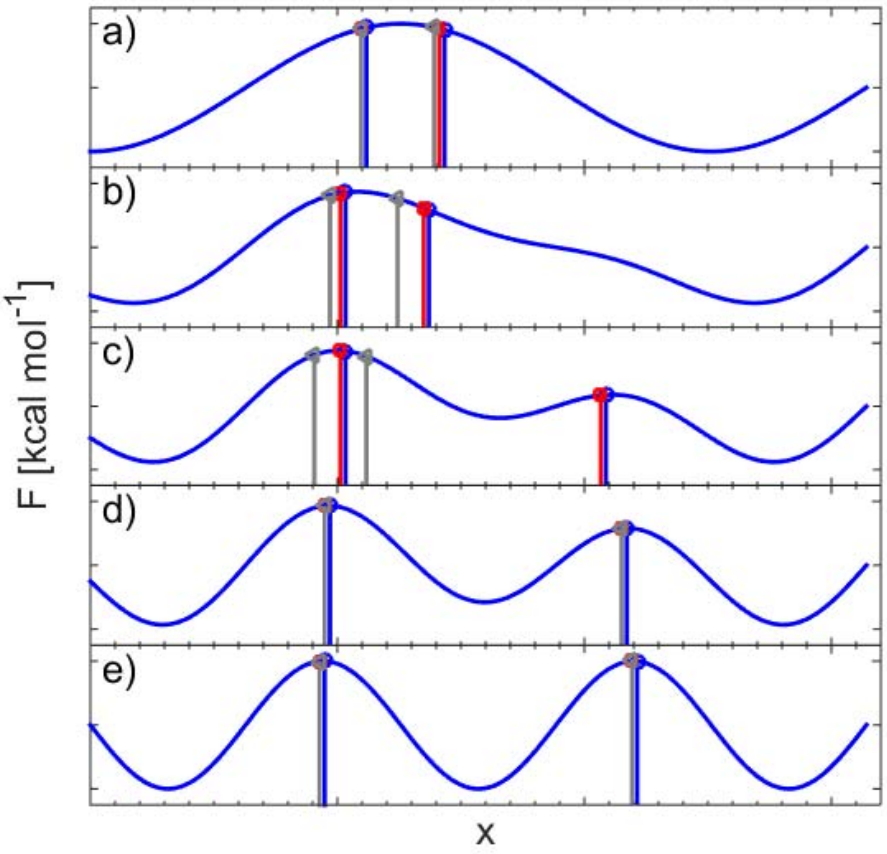


Figure S7. Optimal boundary positions to discretize the 1D free energy profiles using 2 boundaries with the LE method at lagtimes τ = 10 (grey) and

τ = 1000 (red), or using the HS reduced rate matrix (blue) [1]. Note that the optimal boundary positions for the HS method are defined such that they are shifted by 1 bin to allow these positions to be visually distinguished from the long lagtime boundaries and, at the same time, to illustrate the level of discretization for the profile.

Comparison of the Optimal Coarse-Graining with the Committor Probability of 0.5

In general, the committor probability of 0.5 does not necessarily coincide with the location of the TS for complex networks. As a simple model potential, we used the free energy profile of $F(x) = 2\sin(x-\pi) - 0.1(x-5\pi/2)^2$ to demonstrate this in Fig. S8.

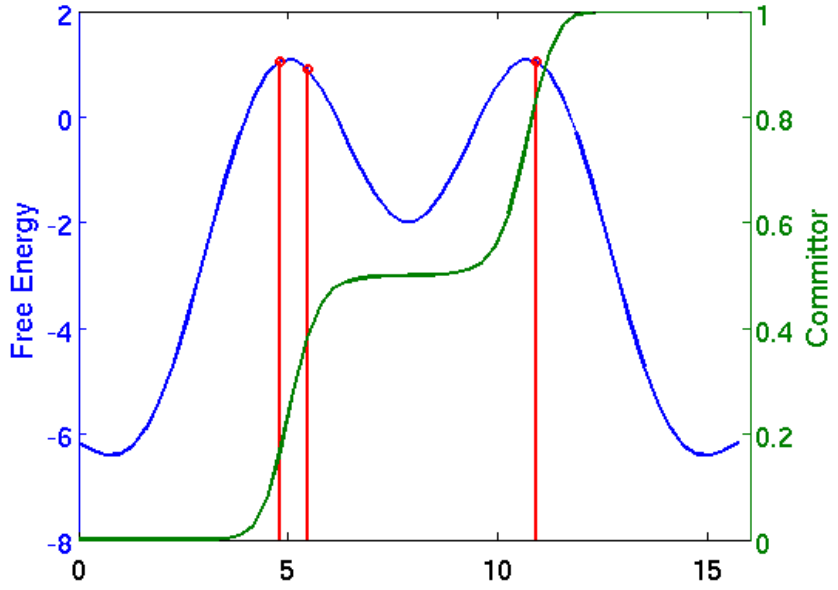


Figure S8. Optimal boundaries (red) for a 4-state model corresponding to the kinetic network built for the free energy profile (left axis, blue) with $F(x) = 2\sin(x-\pi) - 0.1(x-5\pi/2)^2$, using the HS reduced rates. The committor probability (right axis, green) of 0.5 corresponds to a metastable intermediate in this case, and not to a transition state, which is on the other hand, captured correctly between the first and second boundaries using our algorithm.

IV. ALA5 ANALYSIS

We analyzed MD trajectories of Ala_5 with TIP3P [2] water molecules initialized from 4 initial conditions, of 250 ns length each, and performed at two different temperatures T_1 = 300 K, T_2 = 350 K, for a total simulation time of 2 μs. The simulations were performed using GROMACS [3] with the Amber-GSS [4] force field, as described in detail in Ref. [5]. All five Ramachandran Ψ angles have been discretized, and MSMs were constructed for each angle (Fig. S9). For example, 3-state clustering of the backbone angle Ψ_1 at 350 K (Fig. 5A) with N_μ=100 microstates leads to a slowest relaxation time of t_2 = 212.45 ps, and the equilibrium probabilities are, π_0(H) = 0.662, π_1(TS) = 0.001, and π_2(C) = 0.337. The macrostate boundaries are listed in Table S1, and the transition probabilities are T_{22} = 0.0984, T_{20} = 0.4756, T_{21} = 0.4260.

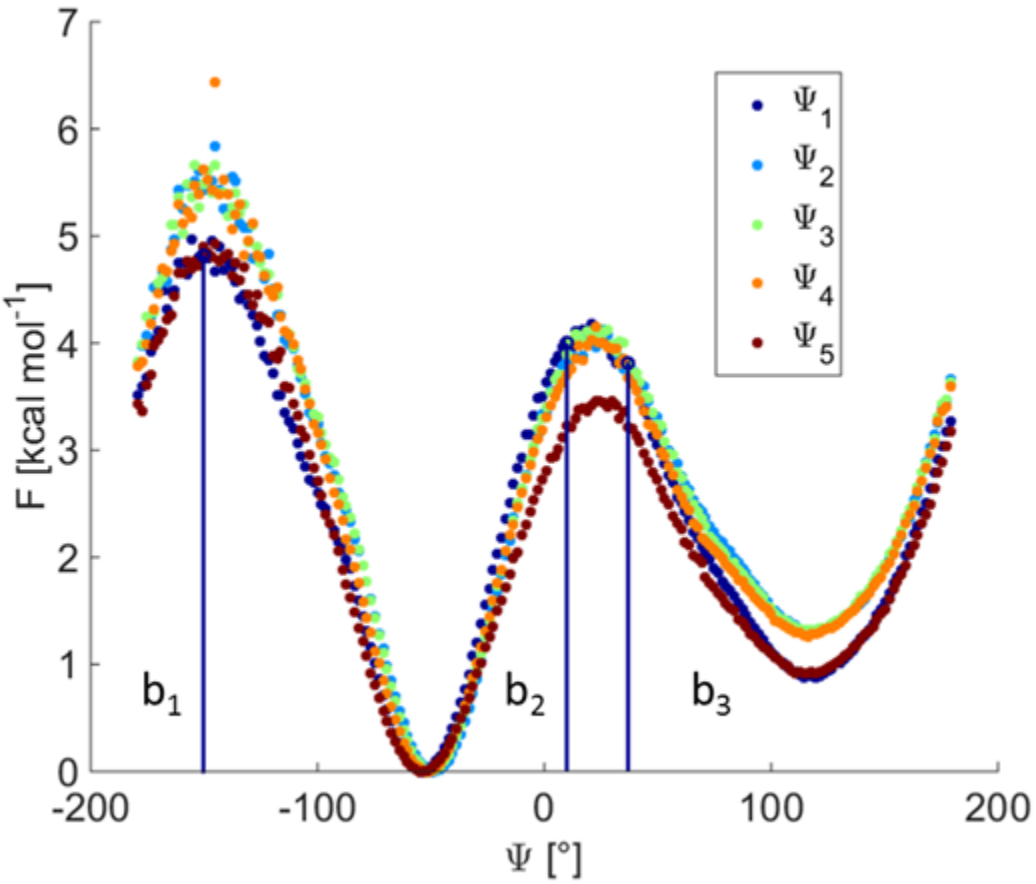


Figure S9. Free energies along the 5 backbone angles ψ of Ala_5 calculated for $N_\mu = 200$ μ-states at 350 K and lagtime τ=10 ps (see also Fig. 2A). Resulting M-states (vertical boundary lines: b1, b2, b3) are shown for the 3 M-states considering also periodic boundaries.

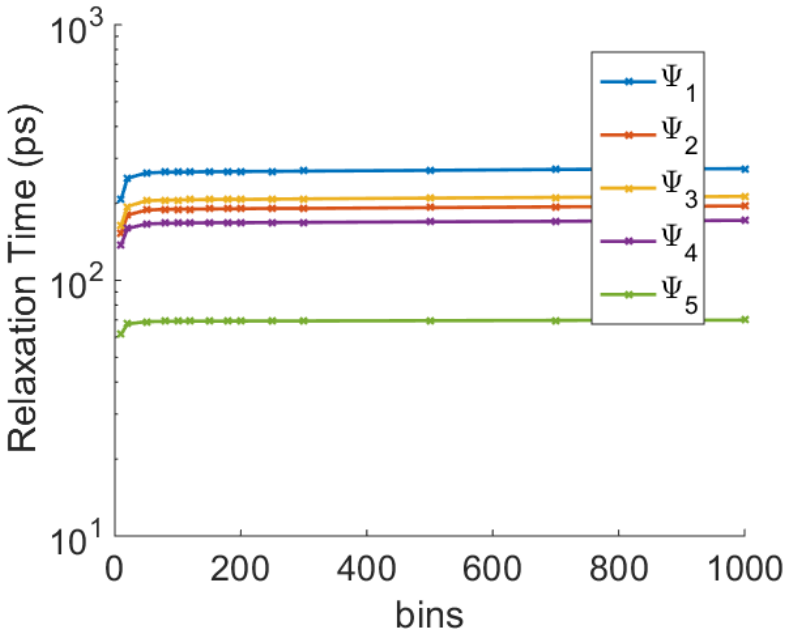


Figure S10. Relaxation times calculated using different number of bins to construct the initial full dimensional MSM along the 5 backbone angles ψ of Ala_5 at 350 K.

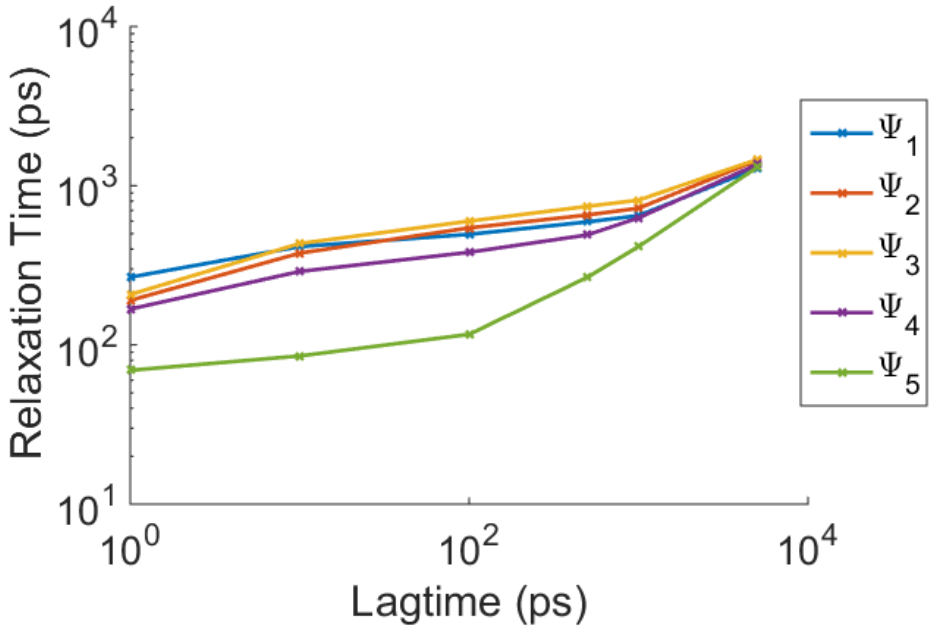


Figure S11. Relaxation times of the full dimensional MSM calculated at different lagtimes along the 5 backbone angles ψ of Ala_5 at 350 K using 200 bins.

To ensure that our full dimensional Markov models are in the correct Markovian limit, we tested how the slowest relaxation time varies as a function of the number of bins (Fig. S10) and lagtimes (Fig. S11) used to construct the MSMs. The relaxation time levels off with increasing bin numbers up to ranges where enough statistics can be obtained in each bin, and our clustering with 100 or 200 bins is in this Markovian limit already where the relaxation time has leveled off.

Similarly the relaxation time levels off before increasing for the largest lagtime (5000 ps > t_2), where numerical errors are already observed. This indicates that this lagtime is too large relative to the intrinsic relaxation time of the system, and for the data statistics to be meaningful. For Ψ_5, which has the fastest relaxation time of the 5 angles, already the last three lagtimes show numerical errors in calculating t_2. Our Markovian relaxation times considering each angle

individually are in good agreement with the overall relaxation time (using all 5 backbone angles) of ~1100 ps determined previously [5]. An even more faithful rendering of the dynamics could be achieved by using a transition-based assignment of states [5] that uses core sets to filter out fast recrossings between the states.

To demonstrate that our lagtime used for the coarse-graining (τ = 1 ps) is already in the Markovian limit with respect to the optimal boundary positions, we analyzed the optimal boundaries applying the HS coarse-graining method on the full dimensional MSMs (Fig. S12). Calculating the optimal clustering as shown in Fig. S12, it is observed that the boundary positions remain effectively constant when changing the lagtimes in the range of 1, 10, and 100 ps using 200 bins. The same results are also obtained using 100 bins with the LE method (Table S1).

Tables S2 and S3 provide details for the final coarse graining into Ω-states, considering all 5 angles at the same time. Ω_1 consists predominantly of only one all-helical M-state $\Sigma_0 = [00000]$ that has an equilibrium population of ~40.84% at 350 K (Table S2). The 2nd most populated state, Ω_2, consists of several macrostates that add up to 6.91% of the total equilibrium population, of which 00001 is predominant (~6.66%). Thus, unfolding at the last C-terminal residue of Ala_5 is not sufficient to perturb the overall stability of helical conformations, and it maintains Ala_5 in its folded ensemble. On the other hand, the unfolded ensemble appears to consist of 5 Ω states (Ω_4 to Ω_8) connected with noticeably slower rates than those within the folded states, and accounting for more than 52% of the total equilibrium population. The "unfolded" ensemble is separated from the helix-rich "folded" ensemble by a TS, Ω_3, that consists of 10 *M*-states with a total population of only ~0.19%, of which the most populated *M*-states are [2000*], [0200*] and [0002*], where the "*" symbol means that at that position either 0 (H) or 1 (C) conformations are

observed. Interestingly, at 300 K the 4th residue is the only residue that significantly modulates the TS dynamics (Table S3).

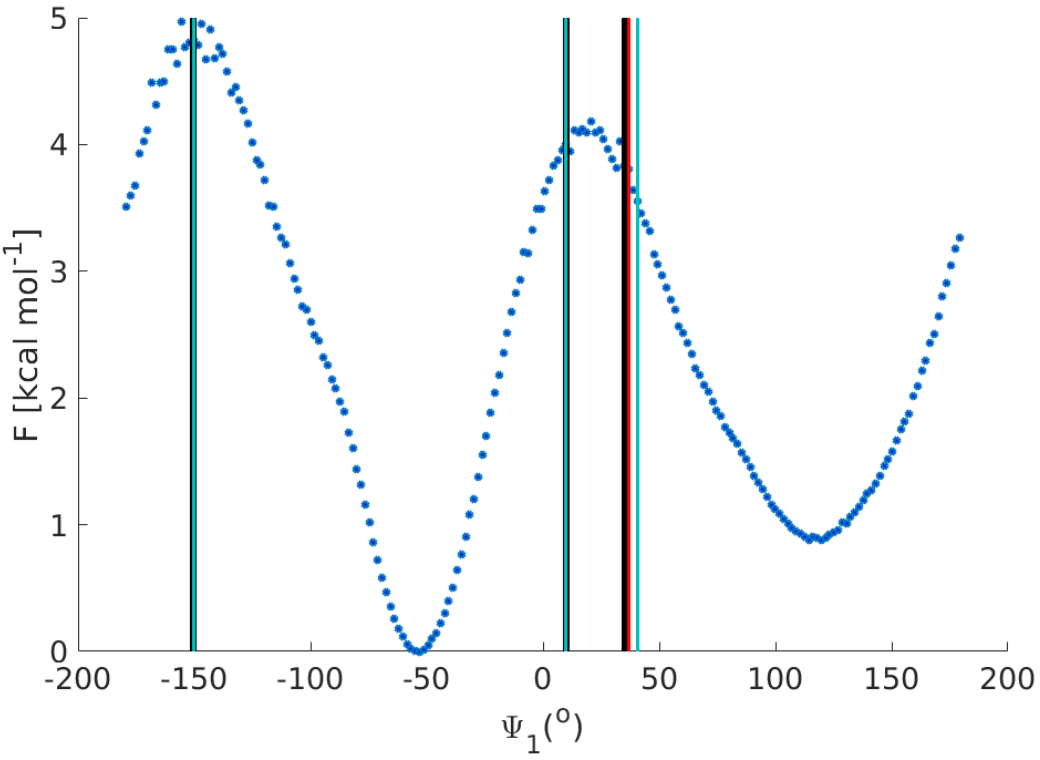


Figure S12. Free energy (blue symbols) along the first backbone angle ψ_1 of Ala_5 calculated for $N_\mu = 200$ μ-states at 350 K and lagtime τ=10 ps (see also Fig. 2A). Resulting optimal boundaries are shown for the 3 M-states at 3 different lagtimes (black: 1 ps, red: 10 ps, green: 100 ps) considering periodic boundary conditions.

Table S1. Optimal boundaries of all 5 Ψ RCs for Ala5 (in °) corresponding to Fig. 3B in the main text, determined at T=350 K, lagtime τ = 1 ps, N=100.

Ψ_1	-150°	10°	31°
Ψ_2	-150°	10°	41°
Ψ_3	-150°	20°	41°
Ψ_4	-150°	20°	41°
Ψ_5	-150°	20°	51°

Table S2. Optimal Ω coarse-grained states for Ala5 at 350 K (corresponding to Fig. 3B in the main text). The corresponding total population of each Ω cluster and the corresponding number of macrostates included are shown, together with the largest population M-state and its value.

Ω-States (at 350 K)	Prob.	No. of M-States	M-State w. largest Prob.	Largest Prob. M-State
Ω1	40.84%	2	"00000"	40.84%
Ω2	6.91%	2	"00001"	6.66%
Ω3 (TS)	0.19%	10	"20000"	0.05%
Ω4	12.59%	7	"10000"	9.26%
Ω5	5.16%	19	"10001"	2.28%
Ω6	12.58%	36	"11001"	1.91%
Ω7	14.94%	49	"10111"	2.10%
Ω8	6.79%	24	"11111"	2.44%

Table S3. Optimal Ω coarse-grained states for Ala5 at 300 K (see main text). The corresponding total population of each Ω cluster and the corresponding number of macrostates included are shown, together with the largest population M-state and its value.

Ω-States (at 300 K)	Prob.	No. of M-States	M-State w. largest Prob.	Largest Prob. M-State
Ω1	56.51%	1	"00000"	56.51%
Ω2	5.48%	3	"00002"	4.88%
Ω3 (TS)	0.13%	9	"0001*"	0.08%
Ω4	11.52%	8	"20000"	9.60%
Ω5	2.30%	8	"20002"	1.34%
Ω6	0.39%	13	"02002"	0.30%
Ω7	3.83%	14	"22000"	1.83%
Ω8	2.63%	18	"22002"	0.95%
Ω9	9.13%	39	"22022"	1.91%
Ω10	8.07%	36	"22222"	2.33%

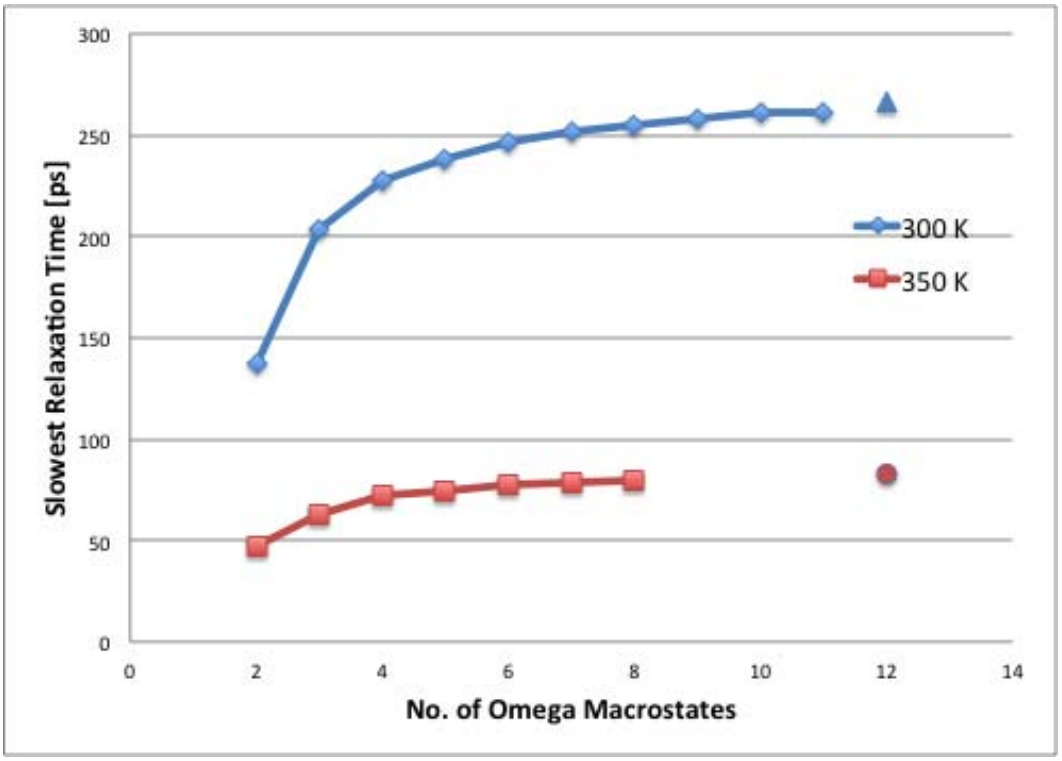


Figure S13. Convergence of the corresponding slowest relaxation time, t_2, at 300 K (blue) and 350 K (red), calculated for increasing number of Ω coarse-grained states (with the exception of the most r.h.s. points that show the values corresponding to the entire set of 149 macrostates).

V. EGFR ANALYSIS

The EGFR simulations by the Shaw group include the extracellular module, the transmembrane segment, the juxtamembrane segment, and the intracellular kinase domain separately [6]. We analyzed the positions of all protein atoms for the N-terminal transmembrane dimer helices, saved with a time step of $\Delta t = 1$ ns at 310 K. Clustering of the one-dimensional tICA trajectories is given in Figs. 7A and S15 using a symmetrized transition count matrix (an iteratively determined reversible transition matrix was also used with equivalent results, data not shown). The small equilibrium probabilities for each component indicate that the low-populated state is indeed TS-like for each tICA coordinate, which was also confirmed by the corresponding transition probabilities.

We used the first four tICA components to obtain the corresponding global network of transitions between the optimal coarse-grained Ω-states (Figs. S14-15). The kinetically most important coordinate was the second component (Fig.7A), which had a slower relaxation time than the first one. The 4^{th} component did not significantly alter the obtained global network as compared with using only the first three components (data not shown). The 5^{th} component had an even faster relaxation time than the 4^{th} (Table S4), and thus was not included in subsequent analysis. To test if the ordering along the second eigenvector is a good approximation, we performed an iterative algorithm where we considered moves of every individual M-state to any of the Ω states. While this procedure led to a minor improvement for the Ala5 data, it did not identify new optimal arrangements in this case. A representation of the Ω_5 TS conformation is illustrated in Fig. 7B, colored according to the rms correlation between the CA atom coordinate and the second tICA component (color scheme blue to red corresponds to low to high). The overall network is illustrated in Fig. 7C and presented in more detail in Table S5. Four coarse grained states (Ω_1 through Ω_4) form one conformational basin separated by the TS-like state Ω_5 (orange) from a second basin formed by Ω_6.

Interestingly, as shown in Table S5, the 2^{nd} tICA component captures best the conformational dynamics between the TS state and neighboring conformations corresponding to conformational changes at the N terminus (Fig. 7B). We found significant differences when analyzing the non-symmetrized transition count matrix (data not shown). This together with the linear structure of the global network indicates that longer trajectories are likely needed for more converged results. Our analysis can thus identify key starting points for additional simulations from the TS configurations that could be aimed for additional sampling of the dynamics.

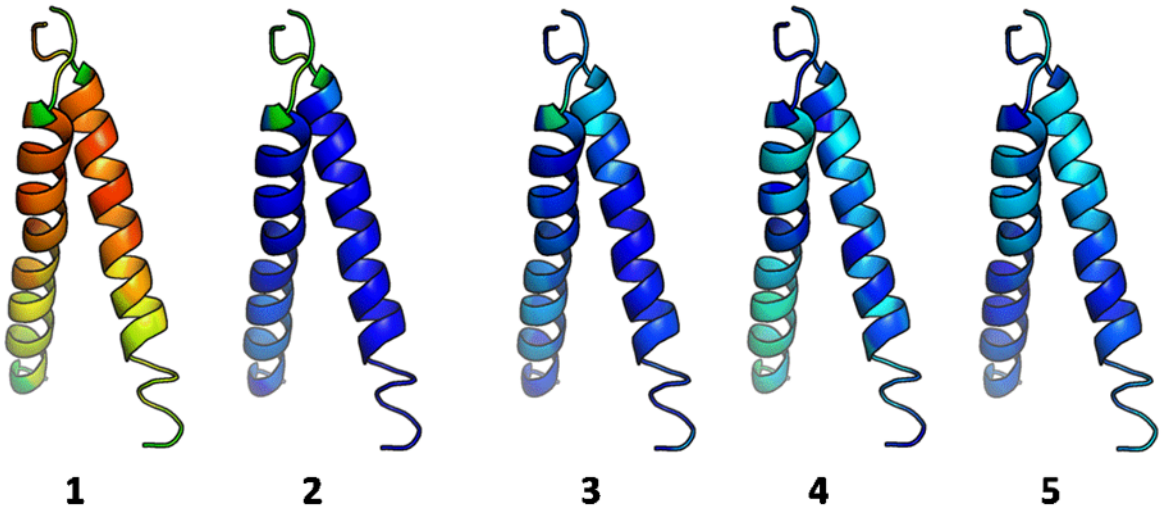
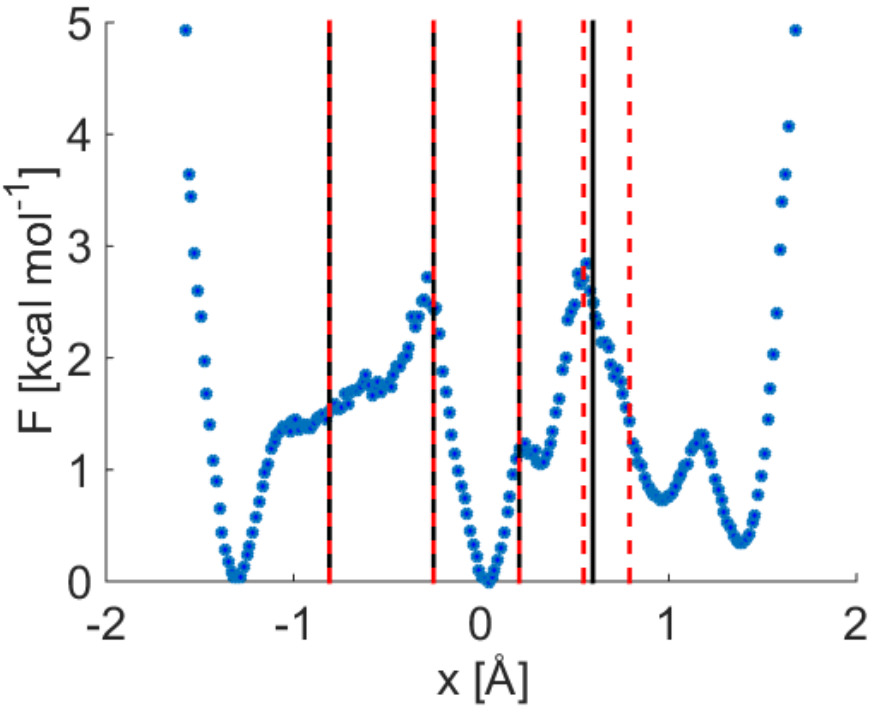
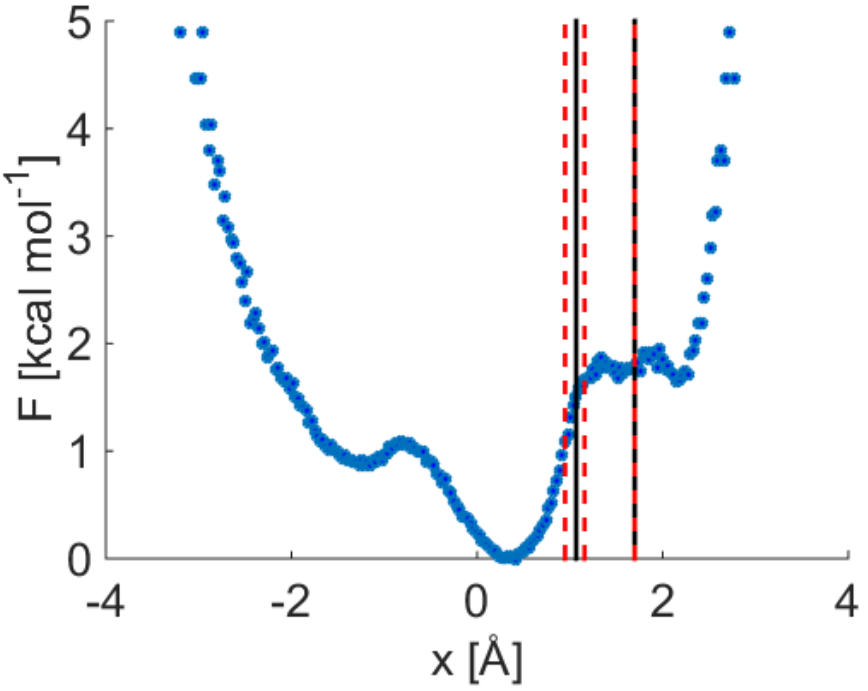


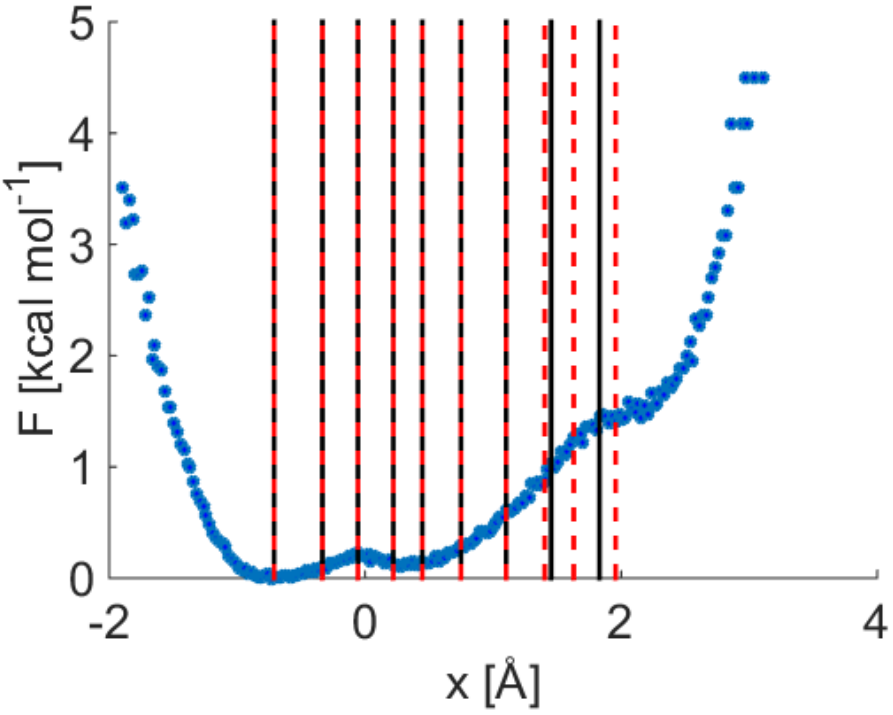
Figure S14. Illustration of the first five tICA components used in our analysis. The colors correspond to the rms values of the tICA correlation matrix for the CA atoms (with blue to red for increasing magnitude).



A



B



C

Figure S15. Free energy profiles for tICA coordinates 1, 3, and 4 (A-C, respectively). See Fig. 4A for the 2nd tICA coordinate. $N_\mu = 200$ μ-states were used for each tICA component. Metastable state boundaries (solid black) and the optimal states including the first TS (dashed red) are illustrated as vertical lines. The TS was defined as having larger overall outgoing probabilities to two metastable states compared with the probability to remain in the state.

Table S4. Clustering of the first five tICA components of the EGFR simulations into 2 to 11 M-states, determined by the appearance of the first transition state. The slowest relaxation times (t_2, in μs) of the coarse grained M-states are given corresponding to the increasing number of clustered states. The relaxation times of the full $N_\mu = 200$ μ-state model (first row) are well approximated by the final macrostates. Note that the slowest relaxation time corresponds to the second tICA component, while the 3rd-5th tICA components have significantly faster relaxation times. The relaxations times of the optimal coarse-grained Ω-states are also given (μs). In this case, the initial product space was formed using the first four tICA components, and a total of 245 states were sampled out of the 6*3*4*11=792 possible states.

	TICA 1	**TICA 2**	**TICA 3**	**TICA 4**	**TICA 5**	**Ω STATES**
ALL STATES	**77.54**	**86.27**	**9.32**	**6.40**	**6.07**	**200.49**
M=2	27.26	45.65	5.00	2.73	3.17	58.96
M=3	40.01	78.70	6.17	3.74	4.17	93.07
M=4	51.48		6.77	4.41		113.25
M=5	57.41			4.93		128.08
M=6	60.52			5.28		148.88
M=7				5.51		
M=8				5.72		
M=9				5.86		
M=10				5.95		
M=11				6.01		

Table S5. Optimal Ω coarse-grained states for EGFR at 310 K (see main text, Fig. 4C) using the first 4 tICA components. The TS was defined as having larger outgoing probabilities to at least two other metastable states, compared with the probability to remain in the state. The number of M-states along each tICA component were 6, 3, 4, 11, respectively (Table S4). The corresponding total population of each Ω cluster and the corresponding number of macrostates included are shown, together with the largest population M-state and its equilibrium probability. The M-states are labelled according to their location along each tICA component (e.g., the state "2-2-1-4" had projections in states numbered S2, S2, S1 and S4, along each tICA component, respectively). The 4th tICA component was not necessary to further distinguish the Ω clusters, accordingly, "*" indicates several different states of this corresponding tICA component.

Ω-States	Equilibrium Probability	No. of M-States	M-State w. Largest Prob.	Largest Prob. M-State
Ω1	34.41%	104	"6 3 1 *"	12.79%
Ω2	3.26%	34	"4 3 3 2"	0.43%
Ω3	3.20%	9	"3 3 1 4"	0.75%
Ω4	33.60%	32	"1 3 1 *"	8.19%
Ω5 (TS)	0.78%	29	"2 2 1 *"	0.28%
Ω6	24.75%	37	"3 1 1 *"	5.03%

VI. MGA2 TRANSMEMBRANE HELIX ANALYSIS

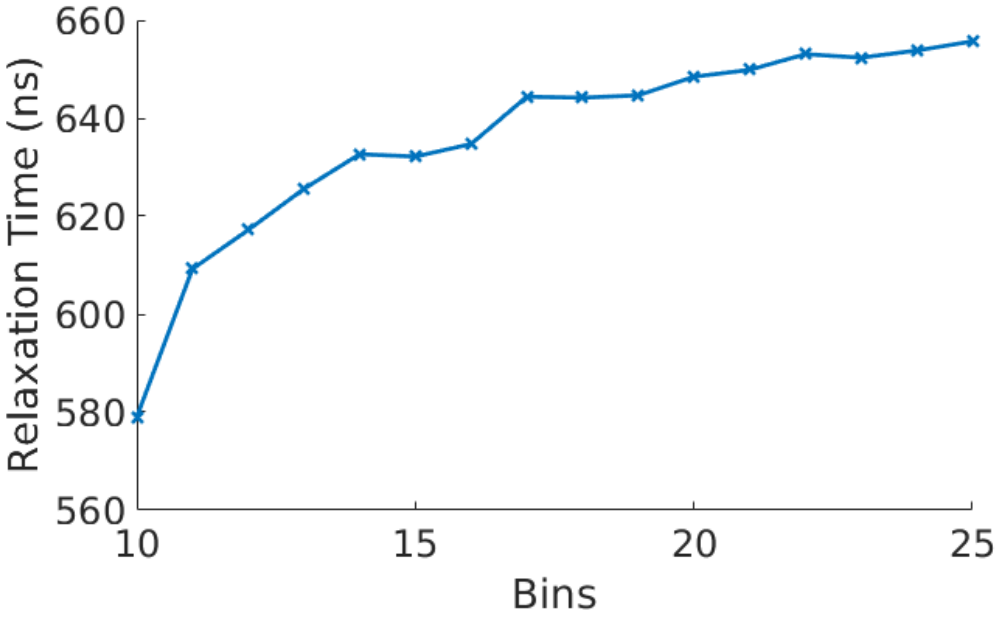


Figure S16. Mga2 dimer rotation relaxation times of the full dimensional MSM calculated for different discretizations of the two reaction coordinates using a lagtime of τ = 100 ns.

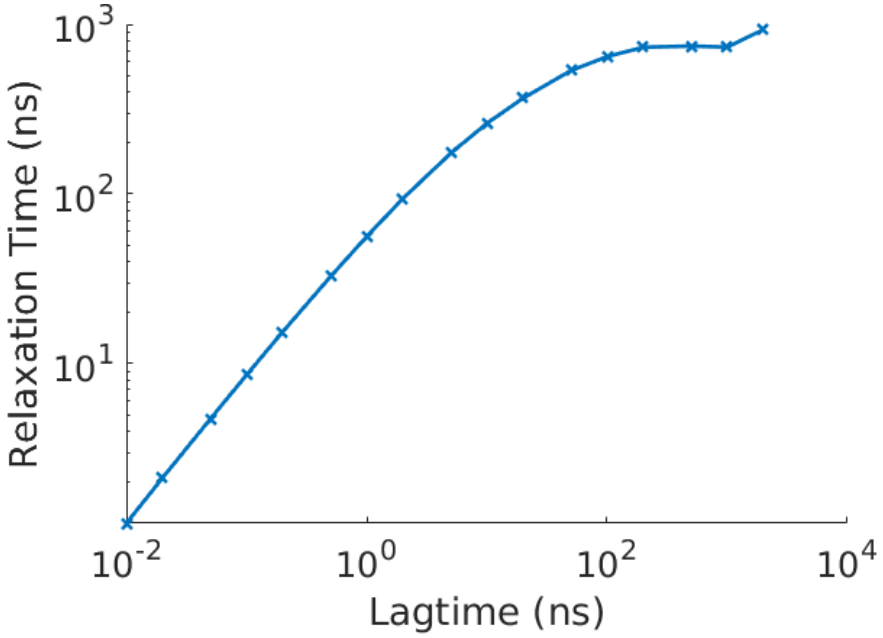


Figure S17. Mga2 dimer rotation relaxation times calculated using different lagtimes to construct the full dimensional MSM with 17 bins for the 2D grid along each reaction coordinate, θ and χ.

REFERENCES

1. Hummer, G. and A. Szabo, *Optimal Dimensionality Reduction of Multistate Kinetic and Markov-State Models.* The Journal of Physical Chemistry B, 2015. **119**(29): p. 9029-9037.
2. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulating liquid water.* The Journal of Chemical Physics, 1983. **79**(2): p. 926-935.
3. Lindahl, E., B. Hess, and D. van der Spoel, *GROMACS 3.0: A package for molecular simulation and trajectory analysis.* Journal of Molecular Modeling, 2001. **7**(8): p. 306-317.
4. Nymeyer, H. and A.E. García, *Simulation of the folding equilibrium of α-helical peptides: A comparison of the generalized Born approximation with explicit solvent.* Proceedings of the National Academy of Sciences, 2003. **100**(24): p. 13934-13939.
5. Buchete, N.-V. and G. Hummer, *Coarse master equations for peptide folding dynamics.* The Journal of Physical Chemistry B, 2008. **112**(19): p. 6057-6069.
6. Shan, Y., et al., *Oncogenic mutations counteract intrinsic disorder in the EGFR kinase and promote receptor dimerization.* Cell, 2012. **149**(4): p. 860-870.