Supplementary Information for

A method of incorporating rate constants as kinetic constraints in molecular dynamics simulations

Z. Faidon Brotzakis, Michele Vendruscolo, Peter G. Bolhuis

Corresponding Author name.
E-mail: p.g.bolhuis@uva.nl

This PDF file includes:
- Supplementary text
- Figs. S1 to S7
- Tables S1 to S2
- SI References
Supporting Information Text

Theory

In this section we give more detailed derivations supplementing the Main Text. We start with listing in Table S1, the symbols and functions used in the main text, together with a short explanation. We then explain how static and dynamic constraints can be applied in the MaxCal framework. We give a derivation of the independence of partial path distribution, and give an outline for solving the constraints for all interfaces simultaneously. Next, we derive that the committor is a natural solution for the configurational constraint using MaxEnt. We end this section with a numerical solution of the Volterra equation that gives the MaxCal biasing function.

Dynamical constraints. In Eqs. 9-12 of the Main Text we apply the MaxCal framework, applying a constraint using a function $s_i(x)$, which can relate to any measurement either giving rise to static/thermodynamic or dynamic/kinetic information. Starting

<table>
<thead>
<tr>
<th>CV</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(x)$</td>
<td>A configurational probability distribution, with $x$ denoting the configuration</td>
</tr>
<tr>
<td>$P^0_i(x)$</td>
<td>The prior configurational probability distribution</td>
</tr>
<tr>
<td>$P_{ME}^0(x)$</td>
<td>The maximum entropy posterior configurational probability distribution, after applying the constraints</td>
</tr>
<tr>
<td>$s_i(x)$</td>
<td>Collective variable/observable $i$ as a function of the microscopic configurations.</td>
</tr>
<tr>
<td>$s_i^{exp}$</td>
<td>The experimental observation/data $i$</td>
</tr>
<tr>
<td>$k_{AB}^{exp}$</td>
<td>Experimental observation/data of transition $k_{AB}$</td>
</tr>
<tr>
<td>$\sigma_i$</td>
<td>Level of confidence in the data $i$</td>
</tr>
<tr>
<td>$\sigma_{k,AB}$</td>
<td>Level of confidence in the rate constant $k_{AB}$</td>
</tr>
<tr>
<td>$\mu_i$</td>
<td>Lagrange multiplier associated with experimental data $i$</td>
</tr>
<tr>
<td>$\mu_{A,B}$</td>
<td>Lagrange multiplier associated with experimental rate constant data for $A$ ($k_{A,B}$), or $B$ ($k_{B,A}$)</td>
</tr>
<tr>
<td>$p(x_i \rightarrow x_{i+1})$</td>
<td>Short time Markovian probability representing the evolution of the probability density</td>
</tr>
<tr>
<td>$P^0(x)$</td>
<td>A full path distribution, with $x$ denoting the trajectory</td>
</tr>
<tr>
<td>$P^0_i(x)$</td>
<td>The prior full path distribution</td>
</tr>
<tr>
<td>$P_{ME}^0(x)$</td>
<td>The maximum caliber posterior path distribution, after applying the constraints.</td>
</tr>
<tr>
<td>$h_{A,B}(x)$</td>
<td>Indicator function identifying whether a configuration is in state $A$ or $B$</td>
</tr>
<tr>
<td>$P_{A,B}^0(x)$</td>
<td>A partial path distribution of trajectories starting in $A$, or $B$</td>
</tr>
<tr>
<td>$P_{A,B}^0(x)$</td>
<td>The prior partial path distribution of trajectories starting in $A$, $B$</td>
</tr>
<tr>
<td>$P_{ME}^0(x)$</td>
<td>The maximum caliber posterior partial path distribution of trajectories starting in $A$, after applying the constraints related to state $A$ i.e $k_{A,B}$</td>
</tr>
<tr>
<td>$P_{ME}^0(x)$</td>
<td>The maximum caliber posterior partial path distribution of trajectories starting in $A$, after applying the constraints related to state $B$ i.e $k_{B,A}$</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>Collective variable used to parameterize ordered set of interfaces between state $A$ and $B$</td>
</tr>
<tr>
<td>$\lambda(x)$</td>
<td>Value of collective variable $\lambda$ evaluated at configuration $x$</td>
</tr>
<tr>
<td>$\lambda_{max}$</td>
<td>Maximum value of $\lambda$ along an individual path $x$</td>
</tr>
<tr>
<td>$\lambda_{min}$</td>
<td>Minimum value of $\lambda$ along an individual path $x$</td>
</tr>
<tr>
<td>$P_\lambda(\lambda_{n}</td>
<td>Probability that a trajectory originating from $A$ and crossing interface $i$, reaches interface $n$ (i.e. $B$)</td>
</tr>
<tr>
<td>$P_\lambda(\lambda_{m}</td>
<td>Probability that a trajectory originating from $A$ and crossing the first interface $\lambda_{0}$, reaches interface $i$</td>
</tr>
<tr>
<td>$P_\lambda(\lambda_{0})$</td>
<td>Probability that a trajectory originating from $A$ and crossing the first interface, reaches a value $\lambda$</td>
</tr>
<tr>
<td>$\phi_0$</td>
<td>Flux through innermost interface $\lambda_{0}$ for trajectories starting in $A$</td>
</tr>
<tr>
<td>$f_{A}(\lambda(x)</td>
<td></td>
</tr>
<tr>
<td>$f_{B}(\lambda(x)</td>
<td></td>
</tr>
<tr>
<td>$P_\lambda(\lambda_{0})$</td>
<td>Probability that a trajectory of the prior path ensemble originating from $A$ and crossing the first interface, reaches a value $\lambda$</td>
</tr>
<tr>
<td>$P_{ME}^0(\lambda_{0})$</td>
<td>Probability that a trajectory of the maximum caliber posterior path ensemble originating from $A$ and crossing the first interface, reaches a value $\lambda$</td>
</tr>
<tr>
<td>$g(\lambda)$</td>
<td>MaxEnt biasing function of the configurational densities that reassures the experimentally determined equilibrium constant</td>
</tr>
<tr>
<td>$g^{exp}$</td>
<td>Experimentally determined value of the configurational density observable, i.e. the equilibrium constant</td>
</tr>
<tr>
<td>$\rho(\lambda)$</td>
<td>Configurational density projected onto $\lambda$</td>
</tr>
<tr>
<td>$\rho^0(\lambda)$</td>
<td>The prior configurational density projected onto $\lambda$</td>
</tr>
<tr>
<td>$\rho_{ME}(\lambda)$</td>
<td>The MaxCal configurational density projected onto $\lambda$</td>
</tr>
<tr>
<td>$\rho_{A,B}(\lambda)$</td>
<td>A configurational density projected onto $\lambda$ for trajectories originating in $A$, $B$</td>
</tr>
<tr>
<td>$\rho_{A,B}^0(\lambda)$</td>
<td>The prior configurational density projected onto $\lambda$ for trajectories originating in $A$, $B$</td>
</tr>
<tr>
<td>$\rho_{ME}^0(\lambda)$</td>
<td>The MaxEnt configurational density projected onto $\lambda$ for trajectories originating in $A$, $B$</td>
</tr>
<tr>
<td>$\rho_{A,B}^0(\lambda)$</td>
<td>The MaxCal configurational density projected onto $\lambda$ for pathways originating in $A$, $B$</td>
</tr>
<tr>
<td>$P_{A,B}^0(\lambda_{0})$</td>
<td>The reaching histogram for pathways originating from state $A$, crossing the innermost interface $\lambda_{0}$ and just reach the value of $\lambda$</td>
</tr>
<tr>
<td>$P_{A,B}^0(\lambda_{n})$</td>
<td>The reaching histogram for pathways originating from state $B$, crossing the innermost interface $\lambda_{m}$ and just reach the value of $\lambda$</td>
</tr>
<tr>
<td>$K_{eq}$</td>
<td>The equilibrium constant</td>
</tr>
<tr>
<td>$K$</td>
<td>The equilibrium fraction</td>
</tr>
<tr>
<td>$K_{eq}^{exp}$</td>
<td>The experimental equilibrium constant</td>
</tr>
<tr>
<td>$K_{eq}^0$</td>
<td>The prior equilibrium constant</td>
</tr>
<tr>
<td>$p_B(x)$</td>
<td>The probability that a trajectory initiated at configuration $x$ (with random momenta) ends up in $B$, a.k.a. the committor</td>
</tr>
<tr>
<td>$p_B(\lambda)$</td>
<td>The probability to end up in $B$, a.k.a. the committor, projected onto the collective variable $\lambda$</td>
</tr>
<tr>
<td>$S(P</td>
<td></td>
</tr>
</tbody>
</table>
with dynamical information, consider an arbitrary time correlation function
\[ c(t) = \langle s_i(0)s_j(t) \rangle = \int Dx P[x] s_i(x_0) s_j(x_\tau), \]  
[A1]
where \( s_{i,j}(x_\tau) \) denotes any observable as function of the configuration \( x_\tau \), with \( \tau = t/\Delta t \) corresponding to the frame index for time \( t \). As \( i \) and \( j \) can be identical, this definition includes autocorrelations. When one has access to experimental correlation data \( c^{Exp}(t) \) it is possible to impose a constraint on the path ensemble distribution, leading to the Lagrange function
\[ \mathcal{L} = -\int Dx P[x] \ln \frac{P[x]}{P^0[x]} - \nu \left( \int Dx P[x] - 1 \right) - \sum_\tau \mu_\tau \left( \int Dx P[x] s_i(0)s_j(t) - c^{Exp}(t) \right). \]  
[A2]
Following the same reasoning as for the MaxEnt approach we optimize the Lagrange function
\[ \frac{\delta \mathcal{L}}{\delta P[x]} = -\ln \frac{P[x]}{P^0[x]} - 1 - \sum_\tau \mu_\tau s_i(x_0) s_j(x_\tau) - \nu, \]  
[A3]
giving rise to the posterior
\[ P^{MC}[x] \propto e^{-\sum_\tau \mu_\tau s_i(x_0)s_j(x_\tau)} P^0[x]. \]  
[A4]
As an example, suppose that we are interested in a mobility function \( K_r[x] \), measuring, for example, the mean square displacement at a particular time \( \tau \) with respect to time \( \tau = 0 \). As this correlation only has to be constrained at \( \tau \), the posterior is simply
\[ P^{MC}[x] \propto e^{-\mu K_r[x]} P^0[x]. \]  
[A5]
Note that this is identical to the expression for the s-ensemble (with \( \mu = s \), not to be confused with the function \( s[x] \)), which biases path ensembles according to a time correlation function (1) and is usually presented in the context of large deviation theory. The s-ensemble biases all paths with a field \( s \) conjugate to the function \( K_r \). In the MaxCal approach the Lagrange multiplier \( \mu \) follows from the constraint imposed. Thus, the s-ensemble might also be interpreted as the field that imposes a certain constraint.

In any case, in general, the posterior MaxCal distribution can be written as
\[ P^{MC}[x] \propto e^{-\sum_\tau \mu_i s_i[x]} P^0[x]. \]  
[A6]
Combining Eq. A6 with the path probability \( P[x] = \exp (-S[x]) \), in terms of the path action \( S[x] \) gives
\[ P^{MC}[x] \propto \exp \left( -\sum_\tau \mu_i s_i[x] \right) \exp \left( -S^0[x] \right), \]  
[A7]
which can be rewritten as
\[ S^{MC}[x] = \sum_\tau \mu_i s_i[x] + S^0[x], \]  
[A8]
Eq. A8 quantifies the correction of the prior action \( S^0 \) by the experimental constraints.

**Thermodynamic constraints.** Since equilibrium properties are not time dependent, they can be computed as time averages over path ensembles distributions:
\[ \langle s \rangle = \frac{1}{\langle L \rangle} \int Dx P[x] \sum_\tau s(x_\tau), \]  
[A9]
with \( \langle L \rangle \) being the average path length, and \( x_\tau \) the coordinates at each timestep of the path. Constraining an equilibrium property \( s^{Exp} \) then leads to a posterior distribution
\[ P^{MC}[x] \propto e^{-\mu \sum_\tau s(x_\tau)} P^0[x]. \]  
[A10]
An alternative way of constraining equilibrium properties is to first reduce the path space back to a configurational density \( \rho(x) \equiv P(x) \) by
\[ \rho(x) \propto \int Dx P[x] \sum_\tau \delta(x_\tau - x). \]  
[A11]
The average then becomes simply
\[ \langle s \rangle = \frac{\int dx \rho(x) s(x)}{\int dx \rho(x)}. \]  
[A12]
Indeed, substitution of Eq. A10 and Eq. A11 in Eq. A12 yields the same result as Eq. A9.
Independence of partial path distributions. In this section we show that kinetic constraints can be applied on partial path ensembles AB and BA separately. To show that, we apply two dynamical constraints onto the total distribution, one for each partial ensemble $P_A[x] \equiv P[x]h_A(x_0)$ and $P_B[x] \equiv P[x]h_B(x_0)$, consisting respectively, of all paths that start in A, and paths that start in B. Here $h_{A,B}(x)$ are the indicator functions, which are unity when the configuration $x$ is in state $A(B)$, and zero otherwise. The Lagrange function is

$$L = - \int Dxp[x] \ln \frac{P[x]}{P^0[x]} - \nu \left( \int Dxp[x] - 1 \right) - \mu_A \left( \int Dxp[x]h_A(x_0)s_A[x] - s_{A}^{exp} \right) - \mu_B \left( \int Dxp[x]h_B(x_0)s_B[x] - s_{B}^{exp} \right),$$

[A13]

where we used the definition of the partial ensembles. Maximisation of the caliber yields the posterior

$$P^{MC}[x] \propto e^{-\mu_A h_A(x_0)s_A[x] - \mu_B h_B(x_0)s_B[x]} P^0[x],$$

[A14]

or, expressing it in partial ensembles

$$P^{MC}_A[x] + P^{MC}_B[x] \propto e^{-\mu_A h_A(x_0)s_A[x] - \mu_B h_B(x_0)s_B[x]} (P^0_A[x] + P^0_B[x]).$$

[A15]

Clearly, for paths belonging to partial ensemble A $h_A(x_0) = 1$ and thus $h_B(x_0) = 0$

$$P^{MC}_A[x] \propto e^{-\mu_A h_A(x_0)s_A[x]} P^0_A[x],$$

[A16]

while for paths from partial ensemble B $h_A(x_0) = 0$ and $h_B(x_0) = 1$ it holds

$$P^{MC}_B[x] \propto e^{-\mu_B h_B(x_0)s_B[x]} P^0_B[x].$$

[A17]

Thus, both partial ensembles can be optimised and normalised independently. Indeed, when imposing kinetic constraints, this is what we aim to do.

Imposing the kinetic constraint for all $\lambda$. In the Main Text we observe that straightforward reweighting using Eq. 20 will introduce a discontinuity in the path distribution. We avoid this undesirable effect by realising that the condition that the reweighted rate should be equal to the experimental rate can be generalised to all values of $\lambda$. The constraint of the experimental rate should in fact apply to all values of $\lambda$. That is:

$$L = - \int Dxp_A[x] \ln \frac{P_A[x]}{P^{0}_A[x]} - \nu \left( \int Dxp_A[x] - 1 \right)
- \sum_{i=0}^{n} \mu_i \left( \int Dxp_A[x] \theta(\lambda_{max}[x] - \lambda_i) P_A(\lambda_n|\lambda_i) - k_{AB}^{exp} \right),$$

[A18]

where the sum runs over the $n+1$ interfaces. Following the usual minimisation of the Lagrange function gives

$$P^{MC}_A[x] \propto \exp \left[ - \sum_{i=0}^{n} \mu_i \theta(\lambda_{max}[x] - \lambda_i) P_A(\lambda_n|\lambda_i) \right] P^0_A[x].$$

[A19]

This needs to obey $n+1$ constraints, for $k = 0 \ldots n$

$$\left( \theta(\lambda_{max}[x] - \lambda_k) P_A(\lambda_n|\lambda_k) \right) = \frac{\int Dxp_A[x] \theta(\lambda_{max}[x] - \lambda_k) P_A(\lambda_n|\lambda_k)}{\int Dxp_A[x]} = k_{AB}^{exp},$$

[A20]

or by substitution of $P^{MC}_A[x]$

$$\int Dxp_A[x] \exp \left[ - \sum_{i=0}^{n} \mu_i \theta(\lambda_{max}[x] - \lambda_i) P_A(\lambda_n|\lambda_i) \right] \theta(\lambda_{max}[x] - \lambda_k) P_A(\lambda_n|\lambda_k) = k_{AB}^{exp},$$

[A21]

where we for this moment assumed that $\int Dxp_A[x] = 1$.

We realise that the sum in the exponent is in fact only dependent on $\lambda_{max}[x]$ (and of course on $P_A(\lambda_n|\lambda)$), but for a given system $P_A(\lambda_n|\lambda)$ is a function of $\lambda$, so the sum can be written as

$$- \sum_{i=0}^{n} \mu_i \theta(\lambda_{max}[x] - \lambda_i) P_A(\lambda_n|\lambda_i) \equiv f_A(\lambda_{max}[x]),$$

[A22]

where the $P_A(\lambda_n|\lambda)$ dependence is implicit in the function $f_A$. The interpretation is that the weight of each path in the posterior path ensemble is entirely dependent on the $\lambda_{max}[x]$. We show that this is indeed correct in the next section.
Solving the constraint function. To solve the constraint equation Eq. A21 we make use of the fact that the integral over the paths can be split into the intervals between the interfaces. For each path with $\lambda_{max}$ between $i$ and $i+1$, the situation is then identical. For each interval the situation is different. For instance, for $k = n$ it holds that only paths beyond $\lambda_n$ will contribute.

$$D_n e^{F_n} p_A(\lambda_n | \lambda_0) = k_{exp}^{A_{AB}},$$

where $D_n$ is the path fraction for paths beyond $\lambda_n$, so $D_n = P^0_A(\lambda_n | \lambda_0)$, and we have defined the sum $F_n \equiv -\sum_{i=0}^{n-1} \mu_i p_A(\lambda_n | \lambda_i)$. Since $P_A(\lambda_n | \lambda_n) = 1$ by definition this changes into

$$e^{F_n} = k_{AB}^{exp} / D_n = k_{AB}^{exp} / P^0_A(\lambda_n | \lambda_0) \equiv e^{f_A(\lambda_n)}. $$

Indeed, this is the correct reweighting that is needed. For the case of $k = n - 1$

$$(D_{n-1} e^{F_{n-1}} + D_ne^{F_n}) P_A(\lambda_n | \lambda_{n-1}) = k_{AB}^{exp}. $$

Now $D_{n-1} = P^0_A(\lambda_{n-1} | \lambda_0) - P^0_A(\lambda_n | \lambda_0)$, and

$$P_A(\lambda_n | \lambda_{n-1}) = P_A(\lambda_n | \lambda_0) / P_A(\lambda_{n-1} | \lambda_0).$$

This will give

$$D_{n-1} e^{F_{n-1}} + D_ne^{F_n} = k_{AB}^{exp} P_A(\lambda_{n-1} | \lambda_0) / P_A(\lambda_n | \lambda_0),$$

or

$$D_{n-1} e^{F_{n-1}} + k_{AB}^{exp} = k_{AB}^{exp} P_A(\lambda_{n-1} | \lambda_0) / P_A(\lambda_n | \lambda_0),$$

which is

$$D_{n-1} e^{F_{n-1}} = k_{AB}^{exp} (P_A(\lambda_{n-1} | \lambda_0) - P_A(\lambda_n | \lambda_0)),$$

or, since $k_{AB}^{exp} = P_A(\lambda_n | \lambda_0)$ by definition, the weight for paths in the $k = n - 1$ slot becomes

$$e^{F_{n-1}} = P_A(\lambda_{n-1} | \lambda_0) - P_A(\lambda_n | \lambda_0) / P_A(\lambda_{n-1} | \lambda_0) - P_A(\lambda_n | \lambda_0).$$

This gives a regular pattern, and for $k = i$ it holds that for paths in this slot the weight is as follows

$$e^{F_i} = P_A(\lambda_i | \lambda_0) - P_A(\lambda_{i+1} | \lambda_0) / P_A(\lambda_i | \lambda_0) - P_A(\lambda_{i+1} | \lambda_0).$$

The next question is what the function $F_i$ is. This function can be expressed as a function of $\lambda$ by noticing that all the paths fall in a slot $i$ and $i + 1$, which are determined by $\lambda_{max} [x]$. This means that for paths in this slot

$$F_i = f_A(\lambda_{max} [x]).$$

Taking the large $n$ limit, and interfaces arbitrary close, this would mean that each path should be reweighted using

$$P^{MC}_A [x] \propto \exp( f_A(\lambda_{max} [x]) ) P^0_A [x],$$

where the function $f_A(\lambda)$ should be the biasing function.

Now we should still check whether $\int Dx P_A [x] = 1$. To do so we expand the integral in intervals, as before, to get

$$\int Dx P_A [x] = \sum_{i=0}^{n} D_i \exp \left[ \sum_{j=0}^{n} \mu_j \theta(\lambda_{max} [x] - \lambda_j) P_A(\lambda_n | \lambda_j) \right]$$

$$= \sum_{i=0}^{n} P_A(\lambda_i | \lambda_0) - P_A(\lambda_{i+1} | \lambda_0) = 1,$$

where we used the fact that $P_A(\lambda_0 | \lambda_0) = 1$. So indeed the normalisation is guaranteed. At first sight it seems that, when the path histograms, e.g. the crossing probabilities $P_A(\lambda | \lambda_0)$, are reweighted with a function $f_A(\lambda)$, we can simply replace

$$P^{MC}_A (\lambda | \lambda_0) = C \int Dx P^0_A [x] \theta(\lambda_{max} [x] - \lambda) e^{f_A(\lambda_{max} [x])},$$
by $P_A^0(\lambda|\lambda_0)e^{f_A^{\lambda}(\lambda_{\text{max}}[x])}$. But this would be wrong. The $\theta$-function in the integral means we have to sum over ensembles of paths for a certain $\lambda_{\text{max}}[x]$. To see this we should look at the ‘reaching’ histogram:

$$R_A^0(\lambda|\lambda_0) = C \int \mathcal{D}x P_A^0[x] \delta(\lambda_{\text{max}}[x] - \lambda),$$  \[A36\]

where $C$ is a normalisation constant. This histogram counts the paths that ‘just reached’ $\lambda$. The crossing probability can be simply obtained from this by integration:

$$P_A^0(\lambda|\lambda_0) = C \int \lambda R_A^0(\lambda|\lambda_0) d\lambda.$$  \[A37\]

This can be easily seen by realising $\theta(\lambda_{\text{max}}[x] - \lambda) = \int_{\lambda_n}^{\lambda} \delta(\lambda_{\text{max}}[x] - \lambda) d\lambda$. For $\lambda = \lambda_n$, all paths that cross $\lambda_n$ should be included. Effectively this could be achieved by taking $\lambda_n \to \infty$.

When the paths are reweighed with $e^{f_A^{\lambda}(\lambda_{\text{max}}[x])}$ this should be done on $R_A^0(\lambda|\lambda_0)$, and not directly on $P_A^0(\lambda|\lambda_0)$. Thus

$$R_A^{MC}(\lambda|\lambda_0) = C \int \mathcal{D}x P_A[|x] \frac{\delta(\lambda_{\text{max}}[x] - \lambda)}{P_A^{\lambda}(\lambda|\lambda_0) e^{f_A^{\lambda}(\lambda_{\text{max}}[x])}}.$$  

Substitution in the crossing probability yield

$$P_A^{MC}(\lambda|\lambda_0) = C \int \lambda R_A^0(\lambda|\lambda_0) e^{f_A^{\lambda}(\lambda)} d\lambda.$$  \[A39\]

This is the proper reweighting of the crossing probabilities.

We are now in the position to check whether the reweighting assumption Eq. A33 is correct, by using Eq. A31 and substitute Eq. A39.

$$e^{F_i} = \frac{P(\lambda|\lambda_0) - P(\lambda_{i+1}|\lambda_0)}{P^0(\lambda|\lambda_0) - P^0(\lambda_{i+1}|\lambda_0)} = \frac{\int_{\lambda_n}^{\lambda} R_A^0(\lambda|\lambda_0) e^{f_A(\lambda)} d\lambda - \int_{\lambda_n}^{\lambda_{i+1}} R_A^0(\lambda|\lambda_0) e^{f_A(\lambda)} d\lambda}{\int_{\lambda_n}^{\lambda} R_A^0(\lambda|\lambda_0) d\lambda - \int_{\lambda_n}^{\lambda_{i+1}} R_A^0(\lambda|\lambda_0) d\lambda} = \frac{\int_{\lambda_n}^{\lambda} R_A^{\lambda}(\lambda|\lambda_0) e^{f_A^{\lambda}(\lambda)} d\lambda}{\int_{\lambda_n}^{\lambda} R_A^{\lambda}(\lambda|\lambda_0) d\lambda} = e^{f_A(\lambda)},$$  \[A40\]

where the latter equality follows from the fact that all paths between $\lambda_i$ and $\lambda_{i+1}$ get the same weight.

**The committor is a solution for $g(\lambda)$**. The natural solution of the equation

$$\frac{\int d\lambda g(\lambda) \rho(\lambda)}{\int d\lambda \rho(\lambda)} = K,$$  \[A41\]

with $K$ the equilibrium fraction, is that $g(\lambda)$ is equal to the (averaged or projected) committor $p_B(\lambda)$. This can be seen as follows. The definition of the committor is the fraction of all paths that lead to $B$ (2)

$$p_B(\lambda) = \frac{\rho_{BB}(\lambda) + \rho_{AB}(\lambda)}{\rho(\lambda)} = \frac{\rho_B(\lambda)}{\rho(\lambda)},$$  \[A42\]

where $\rho_B(\lambda) = \rho_{BB}(\lambda) + \rho_{AB}(\lambda)$ is the density of points that commit to $B$, and the total density is $\rho(\lambda) = \rho_A(\lambda) + \rho_B(\lambda)$ (2). Substituting $g(\lambda)$ by $p_B(\lambda)$ gives

$$\frac{\int d\lambda p_B(\lambda) \rho(\lambda)}{\int d\lambda \rho(\lambda)} = \frac{\int d\lambda \rho_B(\lambda)}{\int d\lambda \rho(\lambda)},$$  \[A43\]

which is indeed the fraction of points committed to $B$, thus by definition equal to the equilibrium fraction $K^{\text{eq}}$.

So the task is now to find the function $p_B(\lambda)$ that solves

$$\frac{\int d\lambda \rho_B(\lambda)}{\int d\lambda \rho(\lambda)} = K^{\text{eq}},$$  \[A44\]

given that original distributions obey

$$\frac{\int d\lambda \rho_B^0(\lambda)}{\int d\lambda \rho^0(\lambda)} = K^0,$$  \[A45\]

Indeed, using the ME reweighted densities this leads to the self consistent relation for the committor, Eq. 37 in the Main Text.
Numerical solution of the Volterra equation. In practice, the Volterra equation of the first kind Eq. 41 in the Main Text should be solved numerically. For instance, when the configurational density histograms $\rho_0^A(\lambda, \lambda_m)$ are computed for each interface value $\lambda_m$, the maximum value of each path in that ensemble, this equation becomes

$$\int_{\lambda_m=\lambda_n}^{\lambda} \rho_0^A(\lambda, \lambda_m)e^{f_A(\lambda_m)} = e^{-\mu g(\lambda)} \int_{\lambda_m=\lambda_n}^{\lambda} \rho_0^A(\lambda, \lambda_m).$$

[A46]

or in a discrete version

$$\sum_{\lambda_m=\lambda_n}^{\lambda} \rho_0^A(\lambda, \lambda_m)e^{f_A(\lambda_m)} = e^{-\mu g(\lambda)} \sum_{\lambda_m=\lambda_n}^{\lambda} \rho_0^A(\lambda, \lambda_m).$$

[A47]

This function can be solved numerically starting from the final value $\lambda_m = \lambda_n$ for which holds

$$\rho_0^A(\lambda, \lambda_n)e^{f_A(\lambda_n)} = e^{-\mu g(\lambda_n)} \rho_0^A(\lambda, \lambda_n).$$

[A48]

Iteration by back-substitution leads to the desired weighting function $f_A(\lambda)$.

Equation scheme of the method. The entire procedure is summarized as an equation scheme in Fig. S1. This scheme can be read as an algorithm, in which the blue boxes on the first row are the input, and the red boxes in the one but last row are the output. Note that the recalculation of the rates in last row of the blue column is just a check whether the constraints are imposed correctly.
Results

Restraining only the final interface $\lambda_B$. Here, we demonstrate how reweighting only pathways that cross the interface of state B leads to a discontinuity and sudden increase in the otherwise monotonically decreasing crossing probability histogram, (see Fig. S2a) as well as a sudden discontinuity in the configurational density (Fig. S2b).

![Graphs showing crossing probability and configurational density](image)

**Fig. S2.** a) Crossing probability out of state A and b) the respective configurational density.

Influence of the choice of CV. In this section we explore the influence of the choice of CV on the optimisation. We examine two different 2D potentials which are shown in Fig. S3 employing standard replica exchange transition interface sampling (RETIS) \(^{(3, 4)}\). These potentials are of the form

$$
\begin{align*}
v(x, y) &= 10e^{-a((x+b)^2 + (y-b)^2)} \\
&\quad - 3e^{-0.3(x-y)^2 - 0.3(x+y-8)^2 - 3e^{-0.3(x+y+y)^2}} \\
&\quad + \frac{32}{1800} (0.00625(x+y)^4 + 10(x-y)^2).
\end{align*}
$$

[B1]

The left potential is obtained by setting $a = 0.1$, and $b = 2$, while the right potential is defined by $a = 0.5$, and $b = 0$.

For the first potential (Fig. S3left) we run a path sampling simulation with RETIS using the $x$ coordinate to define the interfaces. Here, we have set 29 interfaces, at respectively, $x$ values of \{-3.7, -3.55, -3.4, -3.2, -2.9, -2.6, -2.2, -1.8, -1.5, -1.2, -1.0, -0.8, -0.5, 0, 0.5, 0.8, 1.0, 1.2, 1.5, 1.8, 2.2, 2.6, 2.9, 3.2, 3.4, 3.55, 3.7\}. The reciprocal temperature was set $\beta = 1.5$. We performed $10^5$ trials shots per interface, and included replica exchanges and path reversals. Just as before we obtained the densities, which reproduced the potentials.

Next, we apply the MaxCal approach with a strong kinetic constraint of $\mu_A = -3$ and $\mu_B = 3$; the reconstructed posterior RPE free energy is shown in Fig. S4. Then we run a path sampling simulation with RETIS using the $y$ coordinate to define the interfaces; the corresponding RPE free energy landscape is also shown in Fig. S4bottom. The posterior free energy landscapes are clearly shifted, but behave rather independent from the choice of collective variable as order parameter for the RETIS.

For the other potential with two channels we do exactly the same, albeit with a milder constraint $\mu_A = -1$ and $\mu_B = 0$, and show the results in Fig. S5b. Note that here the unbiased RPEs are the same; it does not matter whether $x$ or $y$ is used as an order parameter to define interfaces. However, in the posterior FE landscapes, there are differences, depending on which CV is used as an order parameter. The upper channel is biased more when using $y$ as an order parameter while the lower one is biased more when using $x$ as an order parameter. This leads to a higher free energy for the upper or lower channels respectively. This discrepancy can most likely be resolved by using the generalised approach. We leave this for a future study.

![Graphs of 2D toy potentials](image)

**Fig. S3.** 2D toy potentials to study the influence of the order parameter.
As described in the Main Text, it is possible to compute the MaxCal entropy or equivalently the KL divergence for the MaxCal optimised distributions, and identify which CV perturbs the distribution the least. Here, care needs to be taken how to add the different AB and BA sub-distributions.

Since the potential is symmetric along the diagonal, there is no difference between the \(x\) and \(y\) CV, and both give an identical entropy or KL divergence.

Using equation Eq. 43 of the Main Text we can compute the entropy or equivalently the KL divergence for the 2D potentials using different bias settings for two different CVs: \(\lambda = x\) along the x axis (or equivalently the y axis), and along the diagonal \(\lambda = x + y\). In Table S2 we report the values for the individual path distributions \(S_A, S_B\), as well as the total \(S_A + S_B\), using Eq. 42 of the Main Text. In addition, we compute the correction based on the \(\alpha\) parameter due to the normalisation, from Eq. 43 of the Main Text. The last three columns report the \(\alpha\) weighted entropies (KL divergence), as well as the corrected value of \(S\). Note that we report the negative entropy, or KL divergence, in order to keep all numbers positive. Also, here we can set \(\alpha_0 = 0.5\) due to the symmetry of the potential.

Since the diagonal is much more aligned with the reaction coordinate, we expect the diagonal CV to give better KL divergences. And indeed, for instance for the bias \(\mu_A = -1\) and \(\mu_B = 0\), the KL divergence for the AB path distribution

| CV \(\mu_A\) \(\mu_B\) \(S_A\) \(S_B\) \(S_A + S_B\) \(\alpha\) correction \(\alpha S_A\) \(\alpha S_B\) \(\alpha S_A + (1-\alpha)S_B\) \(S\) corrected |
|---|---|---|---|---|---|---|---|---|---|
| x -1 -1 | 0.000250 | 0.000187 | 0.000437 | 0.500 | \(4.23 \times 10^{-10}\) | 0.000125 | 0.000125 | \(9.35 \times 10^{-3}\) | 0.000219 | 0.000219 |
| x+1 -1 -1 | 6.35 \times 10^{-5} | 4.73 \times 10^{-5} | 0.000011 | 0.500 | \(5.24 \times 10^{-10}\) | 0.000011 | 0.000011 | \(3.18 \times 10^{-5}\) | 0.000025 | 0.000025 |
| x -1 0 | 0.000012 | 0.731 | 0.111 | \(8.96 \times 10^{-5}\) | 0 | 0.000012 | 0.000012 | \(8.96 \times 10^{-5}\) | 0.111 |
| x+y -1 0 | 5.55 \times 10^{-5} | 0.731 | 0.111 | \(4.06 \times 10^{-5}\) | 0 | 0.000012 | 0.000012 | \(4.06 \times 10^{-5}\) | 0.111 |
| x -1 1 | 7.97 \times 10^{-5} | 0.000049 | 0.881 | 0.328 | \(7.02 \times 10^{-5}\) | 0.000049 | 0.000049 | \(5.65 \times 10^{-5}\) | 0.328 |
| x+y -1 1 | 6.38 \times 10^{-5} | 0.000164 | 0.000028 | 0.881 | 0.328 | \(5.65 \times 10^{-5}\) | 0.000164 | 0.000164 | \(5.65 \times 10^{-5}\) | 0.328 |
| x -2 2 | 0.000038 | 0.00888 | 0.00902 | 0.982 | 0.603 | 0.000135 | 0.000135 | 0.000135 | 0.603 |
| x+y -2 2 | 0.000153 | 0.00186 | 0.00182 | 0.982 | 0.603 | 0.000150 | 0.000150 | 0.000150 | 0.603 |
| x -3 3 | 0.000146 | 0.0518 | 0.0519 | 0.998 | 0.676 | 0.000146 | 0.000146 | 0.000146 | 0.676 |
| x+y -3 3 | 0.000255 | 0.0136 | 0.0139 | 0.998 | 0.676 | 0.000254 | 0.000254 | 0.000254 | 0.676 |
Fig. S5. Analysis for the influence of choice of order parameter on the RPEs for a potential with two channels Fig. S3 right. Free energy landscapes from the RPEs for the curved potential. Left is the original sampling, with the x coordinate (top) and y coordinate (bottom) to place interfaces. The right column shows the free energy landscapes from the MaxCal posterior RPEs.

is a factor of two lower for the diagonal. When the BA ensemble is also perturbed, care needs to be taken to weight the contributions of the perturbations in the right way, as given in Eq. 43 of the Main Text. For instance, for $\mu_A = -1$ and $\mu_B = 1$, the BA path distribution KL divergence is at least four times smaller for the diagonal CV, as for the x-axis CV. However, since the path distributions contribute in different proportions, the total entropy for the diagonal CV is only twice as improved.

Note that the diagonal CV is also always performing better for the total path distributions ($S_A + S_B$). This is to be expected, as the diagonal is more aligned to the reaction coordinate, and thus the distribution are expected to be less perturbed. However, for the individual AB and BA path distribution a strong bias (e.g. $\mu_A = -2$) will make the diagonal CV for the AB distribution seem worse (4th column), indicating that diagonal AB distribution is perturbed more than the AB distribution for $\lambda = x$. This is possibly caused by difficulties in the numerical solution of the Volterra equation. We stress that the MaxCal approach should focus on the total distribution, to make a proper comparison, and indeed, both the sum $S_A + S_B$ and the $\alpha$ weighted, and total corrected entropy all show improvements for the diagonal.

Note also that the final KL divergence reported in the last column, including the correction, is steadily increasing when the bias becomes more asymmetric. In fact it is largely dominated by the correction term, due to this asymmetry. This reflects the fact that due to the asymmetry the BA and AB distributions are biased in different directions. However, when just the individual distribution are considered, the larger difference in weighting (the value of $\alpha$) is causing a much larger KL divergence. It is thus the asymmetry, reflecting a change in the thermodynamic equilibrium constant that dominates the MaxCal entropy. Note that for symmetric bias the correction vanishes.

This analysis can show how the choice of the CV influences the optimisation, and how one can use this to optimize the order parameter progress variable along which the MaxCal is performed.

**Association and dissociation of two diphenylalanine dipeptides**

Next, we illustrate the method to characterise the association and dissociation transition of two diphenylalanine dipeptides (FFs), as studied in Refs. (5, 6). In this system A and B refer to the bound and unbound states respectively. Here, we focus on trajectories coming from state A. We obtain the kinetic ensemble, by using minimum distance between the peptides, $d_{\text{min}}$, as a forward model order parameter $\lambda$. As shown in Fig. S6 we first reweight with $\mu_A = -1$ and $\mu_B = 0$. Notably, as shown in Fig. S6a, the self-consistent solution to the committor is shown to exhibit a shift with respect to the prior $p_B$, and the position of the isocommittor point $p_B(\lambda) = 0$ shifts to larger values of the minimum distance between the peptides. The fact that we constrain $k_{AB}$ to a value of $k_{AB}^0 \exp(-1)$ is reflected by the steeper posterior $p_B(\lambda)$, signifying a steeper barrier and thereby slower kinetics, as expected.

Fig. S6b illustrates the numerical solution of the Volterra equation Eq. 41 using back substitution. The original weight
Fig. S6. Dissociation of two diphenylalanine dipeptides. (a) Committor \( p^B_0(\lambda) \) function (black) and solution of the self-consistent Eq. 37 (green) for the explicit simulation using \( \mu_A = -1 \). (b) Original weight function \( e^{-\mu g(\lambda)} \) (green) and back-iterated function \( e^{-f(\lambda)} \) (red). (c) Logarithm of the configurational densities, original (black), reweighted with \( g \) function (green) (not visible, behind red), corrected with \( e^{-f(\lambda)} \) (red). (d) Logarithm of the crossing probabilities, original (black), and corrected with \( e^{-f(\lambda)} \) (red).

\[ e^{-\mu g(\lambda)} = e^{-\mu P^B(\lambda)} \] is shown in green, the back iterated solution \( e^{-\mu f(\lambda)} \) in red. As for the toy models described above, the weighting function is nonlinear (red curve) and oscillates due to numerical errors. Fig. S6c illustrates the original and reweighted densities \( \rho_A \). The reweighted densities with the \( e^{-\mu g(\lambda)} \) are equal to the RPE weighted ones, as expected by construction. Note the smaller probability density at values of minimum distance larger than 0.4 nm. This is exactly the effect of a smaller transition rate constant which steepens the free energy barriers. Finally, Fig. S6d shows the effect of the kinetic constraint to the crossing probabilities. The original crossing probability is shown in red, and the RPE-corrected in black. Indeed, the crossing probability value at interface B is lowered by a factor of \( 1/e \), as imposed. In Fig. S7 we restrain both the forward and the backward rate by \( \mu_A = -1 \) and \( \mu_B = -1 \) and the log crossing probabilities now change on both sides by a factor \( -1 \). The reweighted densities with the \( e^{-\mu g(\lambda)} \) are equal to the RPE-weighted ones, as expected by construction. Interestingly, making the rate exp(-1) times slower increases both the dissociation barrier \( (d \approx 0.4 \text{ nm}) \) as well as the association barrier \( (d \approx 0.5 \text{ nm}) \). The latter is done by disfavoring the stability of a secondary state \( A' \) at \( d \approx 0.5 \text{ nm} \), where one water-hydration layer is mediating peptide contacts. This water-mediated peptide-contact state \( A' \) now becomes a transition state region configuration, also found in a previous study in the context of protein-protein association/dissociation (7).

Fig. S7. Association and dissociation of two diphenylalanine dipeptides. (a) Free energy for a rate correction of \( \mu_A = -1, \mu_B = -1 \), original (black), reweighted with \( g \) function (green) (not visible, behind red), corrected with \( e^{-f(\lambda)} \) (red). (b) Crossing probability histogram of original (black/green), corrected with \( e^{-f(\lambda)} \) (red/blue). c) Representative configurations of the bound (A), secondary (A') and unbound (B) states.
References


