Supporting information to:

Hierarchical Biomolecular Dynamics:
Picosecond Hydrogen Bonding Regulates
Microsecond Conformational Transitions

Sebastian Buchenberg,† Norbert Schaudinnus,† and Gerhard Stock*,†,‡

Biomolecular Dynamics, Institute of Physics Albert Ludwigs University, 79104 Freiburg, Germany, and Freiburg Institute for Advanced Studies (FRIAS) Albert Ludwigs University, 79104 Freiburg, Germany

E-mail: stock@physik.uni-freiburg.de

*To whom correspondence should be addressed
†Albert Ludwigs University Freiburg
‡FRIAS Freiburg
Dihedral Angle Principal Component Analysis (dPCA)

To reduce the dimensionality of the trajectory, we employed the dihedral angle principal component analysis (dPCA) which uses the sine/cosine-transformed $\phi$ and $\psi$ dihedral angles of the protein backbone (for details see Ref. 1). This avoids possible artifacts due to the mixing of overall rotation and internal motion, which may occur when a PCA using Cartesian coordinates is employed to study large amplitude processes such as folding.\(^2\) As the terminal ends exhibit largely uncorrelated fluctuations, we excluded the dihedral angles of the first and last two residues of Aib\(_9\) from the analysis, yielding in total $5 \cdot 2 \cdot 2 = 20$ variables. Fig. S1a,b show the resulting free energy curves $\Delta G(V_j)$ ($j = 1, \ldots, 6$) along of the first six principal components, panel (d) displays the time evolution of the corresponding autocorrelation functions, and panel (c) shows the convergence of the cumulative fluctuations. Components 1-3 in (a) exhibit multi-peaked energy curves, while the energy curves of components 4-6 in (b) become more and more harmonic. In line with these findings, panel (d) shows a timescale separation between the component 1 and components 2-4. The latter show a further timescale separation with respect to the higher components. The normalized cumulative fluctuations in panel (c) reveal that the inclusion of two and five principal components yields 75 and 80 % of the overall fluctuations, respectively.

FIG. S1: Dihedral angle principal component analysis (dPCA) obtained from a MD simulation of the Aib\(_9\) peptide in chloroform at 300 K. Panels (a) and (b) show the free energy curves $\Delta G(V_j)$ ($j = 1, \ldots, 6$) along of the first six principal components and panel (d) displays the time evolution of the corresponding autocorrelation functions. Panel (c) shows the convergence of the cumulative fluctuations.
Global transition network and Markovianity

As explained in the main text, the intermediate states that constitute pathways from L to R can be described by a product state of the chiralities, right-handed (r) or left-handed (l), of the five inner residues of Aib$_9$. E.g., we obtain L = (lllll) and R = (rrrrr), as well as (rllll) if all but residue 3 show left-handed conformations. To define the chiral state “r” or “l” of an individual residue, we employed the definition of the cores 0$_l$ and 0$_r$ in Fig. S3. Using a lag time of $\tau = 10$ ps, we calculated the transition matrix $\{T(\tau)_{ij}\}$ for the resulting 2$^5$ states as described by Eq. (2) below. Figure 2c of the main text shows the resulting network representation of the global energy landscape.

![Graph showing the energy landscape](image)

FIG. S2: Test of the Chapman-Kolmogorov condition (1) at time $t = n\tau = 1$ ns using lag times $\tau = 10$, 20, 50 and 100 ps. Shown is the population probabilities of the 32 states of the system, obtained for the various lag times.

To show that the transition network represents a Markov state model, we test the Chapman-Kolmogorov condition

$$P(n\tau) = T(n\tau)P(0) = T^n(\tau)P(0),$$

(1)

where $P = (p_1, \ldots, p_{32})^T$ denotes the vector of the population probabilities of the 32 product states and $T = \{T(\tau)_{ij}\}$ represents the column-normalized transition matrix calculated for lag time $\tau$. Starting with a uniform initial distribution $p_i(t = 0) = 1/32$, Fig. S2 demonstrates that the dynamics is Markovian for lag times $\tau \geq 10$ ps.
Calculation of transition rates and mean first passage times

We wish to calculate the transition rates $k_{bf}$, $k_{01}$, $k_{lr}$, and $k_{LR}$ associated with the hierarchically coupled processes hydrogen (H) bond breaking $b \leftrightarrow f$, $(\phi, \psi)$ transitions $0 \leftrightarrow 1$, local $l \leftrightarrow r$ transitions, and global $L \leftrightarrow R$ transitions, respectively (see Fig. 5 in the main text). To this end, we applied two different measures. If the process is well characterized by a single rate (and given sufficiently converged statistics), we may simply calculate the corresponding transition matrix (see Sec. on Transition Matrix). More generally, we calculated the distribution of the first passage times $3^3, 4$ (see Sec. on the MFPT) and estimate the transition rate as the inverse of its mean. In both cases, we first need to carefully define the considered states (see following Sec.).

Definition of states

For a given partitioning into states $A, B, \ldots$, the MD trajectory in position space, $x(t)$, is converted into a trajectory in state space, $n(t) = \{A, B, \ldots\}$. Since the terminal ends exhibit largely uncorrelated fluctuations, we restricted the analysis to the inner residues of Aib$_9$. That is, we averaged the rates $k_{bf}$, $k_{01}$ and $k_{lr}$, over residues 4 to 6, and determined the rate $k_{LR}$ using residues 3 to 7. Moreover, we averaged all rates over left-handed and right-handed chiralities.

$b \leftrightarrow f$ transitions: To account for the $b \leftrightarrow f$ transitions describing the making and breaking of H bonds, we consider a H bond as formed, if the distance $d$ between the oxygen and nitrogen is less than 3.6 Å and if the angle $\Theta$ between the line connecting these atoms and the line connecting the O and H atom is less than 30°. The resulting trajectories for $d(t)$ and $\Theta(t)$ are found to exhibit rapid and large fluctuations (see the inset of Fig. 3c in the main text), which may lead to ultrafast crossings and recrossings of the state boundaries and therefore would naively be interpreted as ultrafast transitions. A closer analysis reveals, however, that the recrossings are mainly due to insufficient partitioning of the conformational space in the barrier regions. Since these spurious recrossings defy a simple Markov modeling of the dynamics, we applied a dynamical coring procedure to the data.$^5, 6$ That is, transitions $b \leftrightarrow f$ are counted only if the system stays in the new state at least for the time of 1.2 ps.

$0 \leftrightarrow 1$ transitions: The Ramachandran plot in Fig. S3 shows that the achiral peptide Aib$_9$ samples both left-handed ($l$) and right-handed ($r$) conformations with similar probability. In the discussion below, it turns out advantageous to label the states with respect to the diagonal indicated in Fig. S3. The conformational ground states $0_l$ and $0_l$ are centered at $\approx (-50°, -45°)$ and $(50°, 45°)$. To allow for a detailed description of the conformational sub-transitions discussed in Sec. on the Sequential Model, two excited conformational states $1_a$ and $1_b$ are introduced for each chirality. On the left-handed side, they are centered at $(68°, -45°)$ and $(45°, -100°)$; on the right-handed side at $(100°, -45°)$ and $(-68°, 45°)$. Choosing the radii $r_0 = 21°$ and $r_1 = 18°$ for the ground and
excited conformational states, respectively, we account only for the cores of these states. Moreover, transitions $0 \leftrightarrow 1$ are counted only if the system stays in the new state at least for the time of 1.2 ps. In this way, we avoid spurious recrossing at the boundary of these states.$^{5,6}$

**l$\leftrightarrow r$ and $L$ $\leftrightarrow R$ transitions:** A residue is considered to be in state $j \in \{l, r\}$ only if it is located in its corresponding ground state $0_j$. Similarly, Aib$_9$ is considered to be in state $j \in \{L, R\}$ only if all inner residues are simultaneously in their ground state $0_j$.

**FIG. S3:** Ramachandran plot of Aib peptide, which defines the conformational states of a residue as well as their core regions. The diagonal line separates left-handed conformations (which by convention are plotted at the right side of the Ramachandran plot) and right-handed conformations.

**Transition Matrix**

For all given pairs of states $i$ and $j$, we calculate the column-normalized transition matrix $T = \{T_{ij}\}$ by counting the number of transitions $C_{ij}$ that occur within a given lag time $\tau$. The resulting transition probability

$$T_{ij} = C_{ij} / \sum_j C_{ij}$$

yields the transition rate $k_{ij} = T_{ij} / \tau$. In the case of $b \leftrightarrow f$ and $0 \leftrightarrow 1$ transitions we choose a lag time $\tau = 1.2$ ps, since it is short enough to resolve the dynamics and long enough to sufficiently reduce the noise. The rates of these processes did not change significantly when we varied $\tau$ between 1.2 and 2.4 ps. As described above, we choose a lag time $\tau = 10$ ps to estimate the $l \leftrightarrow r$ transition rates.
Mean First Passage Time

To calculate mean first passage times (MFPT) of the considered processes, we computed from the state trajectory $n(t)$ the probability distribution of the corresponding first passage times (FPT). That is, for every MD time frame $t_A$ a trajectory is located in state $A$, the first passage time amounts for the time difference $t_B - t_A$, where $t_B$ is the time the trajectory hits state $B$. The MFPT is just the mean of the FPT distribution. Choosing temperatures of 280, 300 and 350 K, Fig. S4 shows the resulting FPT distributions of the four types of transitions under consideration. We find that the relatively "fast" transitions $b \leftrightarrow f$ and $0 \leftrightarrow 1$ are reasonably well described by their mean. At 300 K, for example, the $b \leftrightarrow f$ MFPT (48 ps) compares well to the corresponding transition time (45 ps) obtained from a transition matrix calculation. Similarly, the $0 \leftrightarrow 1$ MFPT (1 ns) compares well to the corresponding transition time of 1 ns. The situation is somewhat different for the "slow" transitions or rare events $l \leftrightarrow r$ and $L \leftrightarrow R$, which may exhibit a multi-peak distribution. At 300 K, the two maxima correspond to fast direct transitions (i.e., without any back-reaction, see Sec. on the Sequential Model) and more complicated transition pathways.

FIG. S4: First passage times distribution of the various types of transitions of Aib$_9$ at 280, 300 and 350 K. The distributions associated with the relatively well-sampled transitions $b \leftrightarrow f$ (a), $0 \leftrightarrow 1$ (b), and $l \leftrightarrow r$ (c) were discretized using 10 bins for each decade of the logarithmic time axis. The rare events of the $L \leftrightarrow R$ transitions (c) were binned using 30 bins in total.
Correlation between H bonds and the $0 \leftrightarrow 1$ transition

Table 1: Correlation between the existence of a H bond (including oxygen atoms O1...O5 or nitrogen atoms N4...N8) and the $0 \leftrightarrow 1$ conformational transition of the $n$th residue (monitored by the backbone dihedral angle $\psi_n$, $n = 3...7$). Shown is the probability difference $\Delta P = P_1 - P_0$ that a specific H bond exists in state 0 and state 1. Changes with $\Delta P \geq 25\%$ are shown in bold face.

<table>
<thead>
<tr>
<th>$\Delta P [%]$</th>
<th>O1</th>
<th>O2</th>
<th>O3</th>
<th>O4</th>
<th>O5</th>
<th>N4</th>
<th>N5</th>
<th>N6</th>
<th>N7</th>
<th>N8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\psi_3$</td>
<td>-16</td>
<td>-75</td>
<td>-2</td>
<td>7</td>
<td>-76</td>
<td>-12</td>
<td>-6</td>
<td>-7</td>
<td>-4</td>
<td></td>
</tr>
<tr>
<td>$\psi_4$</td>
<td>-28</td>
<td>-48</td>
<td>-62</td>
<td>-5</td>
<td>8</td>
<td>-2</td>
<td>-75</td>
<td>-26</td>
<td>-6</td>
<td>-6</td>
</tr>
<tr>
<td>$\psi_5$</td>
<td>-14</td>
<td>-19</td>
<td>-59</td>
<td>-65</td>
<td>5</td>
<td>-20</td>
<td>-7</td>
<td>-80</td>
<td>-42</td>
<td>-16</td>
</tr>
<tr>
<td>$\psi_6$</td>
<td>-12</td>
<td>-18</td>
<td>-12</td>
<td>-51</td>
<td>-57</td>
<td>-16</td>
<td>-11</td>
<td>-10</td>
<td>-82</td>
<td>-47</td>
</tr>
<tr>
<td>$\psi_7$</td>
<td>-5</td>
<td>-6</td>
<td>-7</td>
<td>-30</td>
<td>-53</td>
<td>-9</td>
<td>0</td>
<td>-4</td>
<td>-16</td>
<td>-80</td>
</tr>
</tbody>
</table>

Coexistence of $\alpha$- and $3_{10}$-helical structures

To study how the H bond dynamics in Aib$_9$ depends on temperature, Fig. S5(a) shows the probability distribution of the H bonding angle $\Theta_{n,n+3}$ at various temperatures ranging from 200 to 350 K. At low temperatures ($T < 230$ K), the system is almost exclusively in an $\alpha$-helix ($\Theta \approx 60^\circ$), although all trajectories were started in a $3_{10}$ conformation ($\Theta \approx 10^\circ$). At higher temperatures both $\alpha$ and $3_{10}$ helices co-exist at a ratio of $\approx 30$ to $70\%$. The latter finding is in line with previous experimental and computational work on Aib-rich peptides, which also reported a co-existence of $3_{10}$ and $\alpha$ structures, with the $3_{10}$ helix being the dominant one.\textsuperscript{7–9}

The localization of an Aib-rich peptide into an $\alpha$-helix at low temperatures, on the other hand, has to our knowledge not yet been described in the literature. As the backbone fluctuations of $3_{10}$ structures are found to be consistently higher than the fluctuations of $\alpha$ structures (Fig. S5c,d), the $3_{10}$-helix appears to be stabilized by entropic effects for $T \geq 230$ K,

$$\Delta G_{3_{10},\alpha} = \Delta U_{3_{10},\alpha} - T \Delta S_{3_{10},\alpha} < 0,$$

(3)

while at lower temperatures the enthalpically favored $\alpha$-helix predominates, $\Delta G_{3_{10},\alpha} > 0$.

The effect is further illustrated in Fig. S5b, which shows a sharp transition of the number of $3_{10}$ and $\alpha$ bonds at $\approx 230$ K. Representing the sum of H bonds (where each $3_{10}$-bond counts +1 and each $\alpha$-bond counts −1) as a function of time, Fig. S5e,f reveal that the change of H bonds at $T \geq 230$ K takes place on a timescale of 1 ns and in a predominantly cooperative manner. Since the opening of helix-stabilizing H bonds is a prerequisite for all other conformational transitions in Aib$_9$, we focus on temperatures $T \geq 240$ K in the main part of the paper.
FIG. S5: (a) Probability distribution of the H bonding angle $\Theta_{n,n+3}$ (averaged over all inner H bonds $n = 3, \ldots, 7$) shown for various temperatures. (b) The average number of $3_{10}$ (green) and $\alpha$ (red) H bonds reflect a dynamic transition of Aib$_9$ at 230 K, while the total number of H bonds (blue) stays about constant. (c,d) Average variance of the $\phi$ and $\psi$ dihedral angle, respectively, pertaining to residues 4 – 6 in the $\alpha$-helical state (green line) and $3_{10}$-helical state (red line). (e,f) Representative examples of the time evolution of the sum of H bonds (where each $3_{10}$ bond counts +1 and each $\alpha$ bond counts −1) at (e) 220 K and (f) 240 K.
Sequential model

To account for the details of the pathways of the conformational transitions in Aib$_9$, we describe the conformational state of an individual residue by the vector $(h,s,c)$, where $h = b,f$ (H bonding), $s = 0,1_a,1_b$ (ground/excited state) and $c = l,r$ (chirality). Since we average over the chirality in many cases, we furthermore introduce the short notation $s_h \equiv (h,s,c = \{l,r\})$, yielding the state space $\mathcal{V} = \{0_b, 0_f, 1_{ab}, 1_{bb}, 1_{af}, 1_{bf}\}$. Similarly we define $0 = \{0_b,0_f\}$, $1_a = \{1_{ab},1_{af}\}$, $1_b = \{1_{bb},1_{bf}\}$ as well as $1 = \{1_a,1_b\}$, to account for situations when we do not discriminate between the corresponding substates.

As explained in the main text, the interpretation of the $0 \rightarrow 1$ transition as a one-step process, $0_b \xrightarrow{k_{01}} 1_f$ is misleading. To account for the fact that first the breaking of H bonds is required, we consider the sequential process

\[
0_b \xrightarrow{k_{bf}} 0_f \xrightarrow{k_{01}'} 1_f,
\]

which leads to the rate equations

\[
\frac{d}{dt} 0_b = -k_{bf} 0_b + k_{fb} 0_f, \tag{5}
\]
\[
\frac{d}{dt} 0_f = -(k_{fb} + k_{01}') 0_f + k_{bf} 0_b, \tag{6}
\]
\[
\frac{d}{dt} 1_f = k_{01}' 0_f. \tag{7}
\]

Since $k_{fb}, k_{bf} \gg k_{01}'$ (see below), the states $0_b$ and $0_f$ can be assumed stationary on the timescale on which we model the $0 \rightarrow 1$ transition, that is, $d/dt 0_b \approx 0$ and $d/dt 0_f \approx 0$. Hence Eq. (6) reduces to

\[
\frac{d}{dt} 1_f = \frac{k_{01}'}{k_{fb}} k_{bf} 0_b, \tag{8}
\]

which gives rise to the interpretation of $\frac{k_{01}'}{k_{fb}} k_{bf}$ as the effective rate for $0 \rightarrow 1$ transition. As described above, the transition matrix associated with state space $\mathcal{V}$ was evaluated to give the desired rates $k_{fb}$, $k_{bf}$ and $k_{01}'$. Fig. S6 displays the temperature dependence of these rates and confirms that the back-rate $k_{fb}$ is indeed much higher than the direct forward rate $k_{01}'$. The interscale coupling between tiers C (H bonding) and tier B (0→1) is thus found to be a consequence of a sequential reaction with a high back-rate rather than caused by different barrier heights. To illustrate the coupling between tier B (0→1) and tier A (l → r) in a similar way, we tried to model
FIG. S6: Transition rates and exponential fits. The left figure shows the transitions $(f,0) \rightarrow (b,0)$ (black) and $l \rightarrow 0$ (blue) The right figure shows the transitions $(f,0) \rightarrow (f,1)$ (black) and $(f,1,l) \leftrightarrow (f,1,r)$ (blue). We report the mean over residues 4, 5, 6 and standard error.

the transition $0_l \leftrightarrow 0_r$ via the multi-step process

$$0_l \overset{k_{01}}{\underset{k_{10}}{\rightleftharpoons}} l_{a,r} \overset{k_{1l}}{\underset{k_{0l}}{\rightleftharpoons}} 0_l,$$

(9)

corresponding to the the “upper” path of the Ramachandran space in Fig. S3 (and analogously for the lower path). Again, we found that the corresponding back-rate $k_{10}$ is considerably higher than the direct forward rate $k'_{lr}$ (Fig. S6). By calculating the transition matrix for the corresponding state space, however, we found that it is not possible to construct a Markov state model for this multiscale process [in the sense of Eq. (1)]. This is because the $l \leftrightarrow r$ dynamics is not Markovian, when we choose the lag time of the transition matrix short enough to resolve the $l \leftrightarrow r$ dynamics. Moreover, the tacitly made approximation that the conformational dynamics of an individual residue of the peptide can be considered as independent of the dynamics of the remaining residues, is found to brake down in the case of the $l \leftrightarrow r$ transition. Nonetheless we can model the overall rate $k_{lr}$ in terms of a product of the rate of the underlying rate $k_{01}$ and an exponential factor, see Fig. 5c in the main text.
FIG. S7: Calculated self-diffusion constant $D = \langle \Delta x(t)^2 \rangle / (6 \delta t)$ (red lines) for (a) chloroform and (b) water along with experimental values (green crosses) taken from Prielmeier and Luedemann\textsuperscript{10} and Price, Ide and Arata\textsuperscript{11}.

**Effect of solvent**

To study if the dynamical transition of Aib\textsubscript{9} in chloroform is of intramolecular origin or driven by the solvent, we calculated the self-diffusion constant $D(T)$ of the solvent. Fig. S7 shows that calculated and experimental results are in good agreement for both chloroform and water. In the case of water, we find a biphasic temperature dependence of the diffusion constant, which reflects the well-established dynamical transition of proteins in water that is associated with the glass transition of solvent water at $\approx 220$ K.\textsuperscript{12–15} However, no such effect is seen for chloroform in the entire considered temperature range of 200 - 350 K. This indicates that chloroform does not originally cause the peculiar temperature dependence of the conformational dynamics of Aib\textsubscript{9}.
References


