Dynamical fingerprints for probing individual relaxation processes in biomolecular dynamics with simulations and kinetic experiments

Frank Noé[a,1], Sören Doose[b], Isabella Daidone[c], Marc Löllmann[d], Markus Sauer[b], John D. Chodera[e], and Jeremy C. Smith[f]

[a]Research Center Matheon, FU Berlin, Arnimallee 6, 14159 Berlin, Germany; [b]Department of Biotechnology and Biophysics, Julius-Maximilians University, Am Hubland / Biozentrum, 97074 Würzburg, Germany; [c]Dipartimento di Chimica, University of L'Aquila, Via Vetoio (Coppito 1), 67010 Coppito (AQ), Italy; [d]Applied Laser Physics and Laser Spectroscopy, Bielefeld University, Universitätsstrasse 25, 33615 Bielefeld, Germany; [e]California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, CA 94720-3220; and [f]Center for Molecular Biophysics, University of Tennessee/Oak Ridge National Laboratory, One Bethel Valley Road, P.O. Box 2008, Oak Ridge, TN 37831-6255

Edited* by William A. Eaton, National Institutes of Health, Bethesda, MD, and approved December 27, 2010 (received for review April 10, 2010)

There is a gap between kinetic experiment and simulation in their views of the dynamics of complex biomolecular systems. Whereas experiments typically reveal only a few readily discernible exponential relaxations, simulations often indicate complex multistate behavior. Here, a theoretical framework is presented that reconciles these two approaches. The central concept is "dynamical fingerprints" which contain peaks at the time scales of the dynamical processes involved with amplitudes determined by the experimental observable. Fingerprints can be generated from both experimental and simulation data, and their comparison by matching peaks permits assignment of structural changes present in the simulation to experimentally observed relaxation processes. The approach is applied here to a test case interpreting single molecule fluorescence correlation spectroscopy experiments on a set of fluorescent peptides with molecular dynamics simulations. The peptides exhibit complex kinetics shown to be consistent with the apparent simplicity of the experimental data. Moreover, the fingerprint approach can be used to design new experiments with site-specific labels that optimally probe specific dynamical processes in the molecule under investigation.

Biological processes rely on the ability of macromolecules such as proteins and nucleic acids to dynamically change between different functional conformational and association states. Examples are the folded, unfolded, and long-lived intermediates in protein folding or apo, and complexed states in protein-ligand binding (1, 2). These processes may involve a few states that are distinct in terms of their signature under a given experimental observation, and experimental analyses often allow only one or two time scales to be distinguished (3, 4), suggesting simple two- or three-state models are sufficient to describe their behavior. In contrast, molecular dynamics (MD) simulations often reveal a considerably more complex picture with multiple metastable states and a multitude of relaxation times (5, 6). Theoretically, the macroscopically detectable changes have been proposed to arise from a stochastic walk on a rugged multidimensional energy landscape (7), possibly involving a hierarchy of barriers, resulting in a hierarchy of relaxation time scales (8), or, alternatively, a jump process on a transition network between conformational substates (5, 9, 10) for which a given structural change may involve multiple pathways (6). For simple systems, these different views—microscopic/theoretical and macroscopic/experimental—can be reconciled. For example, a system with ideal two-state kinetics has a single experimentally measurable relaxation time scale, corresponding structurally to the crossing of the predominant energy barrier. However, such an unambiguous mapping has not yet been established for more complex systems.

In this work, we present a general approach that evolves around the concept of *dynamical fingerprints* which visualize the essential kinetic features of the observed system in terms of peaks that have certain amplitudes and relaxation time scales. We derive a theoretical framework and computational methods that allow these fingerprint peaks to be calculated from simulation data and unambiguously linked to the structural processes present in the simulation. This link is the essential advantage over traditional analyses which do not enable the assignment of structural processes to relaxation time scales. Because fingerprints can also be calculated from experimental data directly, it is then possible to match peaks and thus assign structural processes to individual relaxation processes present in the measurement—provided that the simulation is a good model of the measured system (Fig. 1). Because matching a single pair of simulated and experimental fingerprints is often uncertain (e.g., due to force-field inaccuracies), we illustrate how the approach can be used in order to propose a new set of experiments with site-specific labeling that optimally probes specific features of the fingerprint.

This approach is applicable in conjunction with all experiments which measure relaxation profiles that decay with characteristic time scales of the molecular system. Such experiments include (*i*) experiments that monitor the relaxation of an ensemble towards equilibrium after an initial perturbation *via*, e.g., a jump in temperature (11, 12), pressure (13), a change in the chemical environment (14), or a photoflash (15–16); (*ii*) dynamical spectroscopic experiments such as inelastic neutron scattering (16), and (*iii*) low concentration or single molecule experiments accumulating auto- or cross-correlations of fluctuations, e.g., correlation spectroscopy of the fluorescence intensity (3, 17, 18–19) or Förster resonance energy transfer efficiency (20, 21).

Experimentally-measured relaxation profiles are typically analyzed by fitting a single- or multiexponential model. This approach is not objective as it requires the number of time scales to be fixed. For example, multiple exponentials with similar time scales, or a double-exponential where the larger time scale has a small amplitude will both yield visually excellent single-exponential fits with an effective time scale that may not exist in the underlying system [see (5) and *SI Appendix: Figs. 1 and 4*]. This effect arises from the fact that on the time axis exponentials cannot be visually separated as each exponential affects all times. To prepare the experimental data for a systematic analysis, we therefore propose a method that uniquely transforms the observed relaxation profile into an amplitude density of relaxation time scales (here called "dynamical fingerprints") present in the data. Several such methods have been

Author contributions: F.N., S.D., I.D., M.S., and J.D.C. designed research; F.N., I.D., and M.L. performed research; F.N. and M.L. contributed new reagents/analytic tools; F.N., S.D., and M.L. analyzed data; and F.N., S.D., I.D., J.D.C., and J.C.S. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: frank.noe@fu-berlin.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1004646108/-/DCSupplemental.

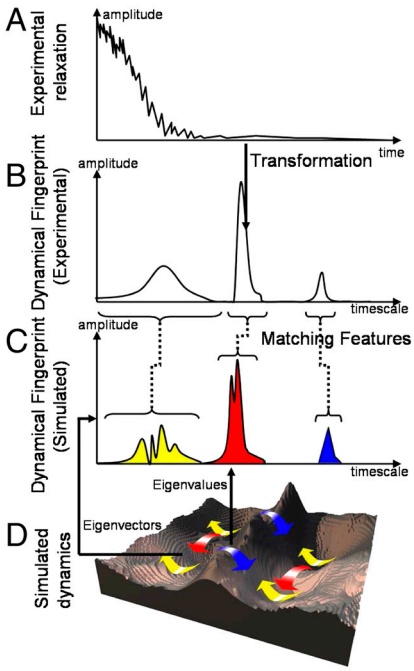

Fig. 1. Dynamical fingerprints: Experimental relaxation profiles (*A*) can be transformed into dynamical fingerprints (*B*), which represent the time scales and amplitudes of the relaxation processes in the data without having to predetermine a particular model or number of processes. On the theoretical side, the dynamics on an energy landscape (*D*) also generates dynamical fingerprints (*C*), but here each feature can be uniquely assigned to a particular transition or diffusion process on the landscape. If the simulation model is a sufficiently accurate model of the experimental system, structural processes can be assigned to the experimental data by matching features in the experimental and theoretical fingerprint.

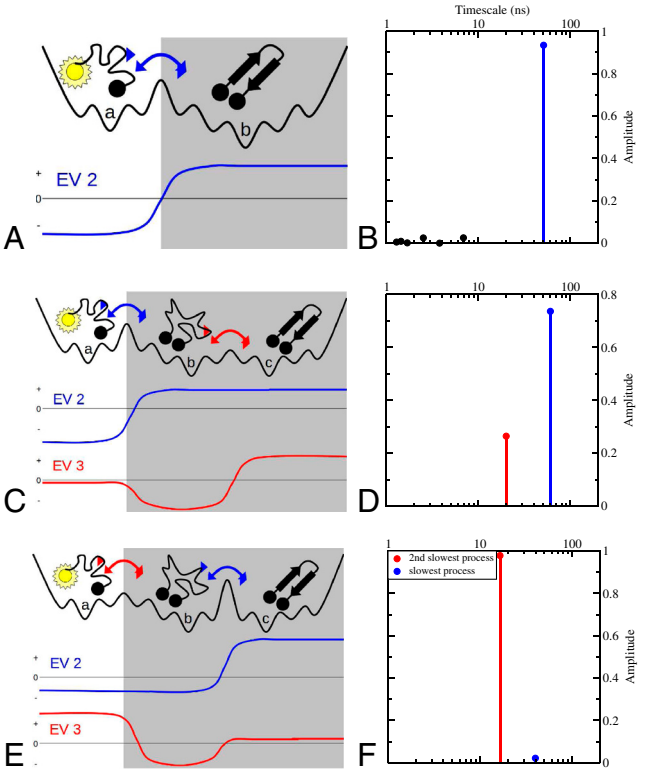

Fig. 2. Illustrative models of systems probed by fluorescence correlation. (*Left*) model energy landscapes with fluorescent states in white and dark states shaded in gray. The slowest transition process is governed by the second eigenvector (blue) of the dynamical operator, the next-slowest process by the third eigenvector (red). (*Right*) Dynamical fingerprints. The fingerprints are shown by blue and red lines and bullets, corresponding to the slowest and second-slowest transition processes on the left. (*B*) also shows the fingerprint peaks of the fast local diffusion processes as short black lines and bullets while for clarity this information is not shown in (*D*) and (*F*).

developed (22, 23) and in the present work a related method is given that calculates the maximum likelihood dynamical fingerprint from the measured relaxation profile under the experiment-specific measurement noise. Nonexponential models used in some analyses, such as stretched exponentials (24, 26) or power laws (25), represent effective models of sums of exponentials with particular spacings of time scales and amplitudes (26, 27) and thus also correspond to dynamical fingerprints with multiple peaks (see *SI Appendix: Figs. 1 and 2*).

The experimental relaxation profile and its dynamical fingerprint do not directly reveal the structural processes. However, detailed information on structural changes can be obtained from MD simulations. Here, we derive a theory and computational methods that allow dynamical fingerprints to be calculated from MD simulations. The critical advantage of the approach is that it associates each fingerprint peak uniquely to an eigenvector/eigenvalue pair of an operator describing the conformational dynamics (Fig. 1), thus permitting each peak to be uniquely associated with a specific process of structural change. If the simulated system is a realistic reproduction of the experimentally measured system, the dynamical fingerprints will be similar as well. Then, by matching peaks between the experimental and theoretical fingerprints, the experimental features can also be assigned to structural processes.

Before applying this approach to a real biomolecular system, we illustrate it with three simple model systems, depicted in Fig. 2, probed by the correlation of a fluorescence signal that is quenched by end-to-end contact formation. Fig. 2*A* shows a potential dominated by one major energy barrier between two wells with some local roughness. The system exhibits two-state behavior, i.e., its slowest process is well separated in time from the fast local diffusion processes. Structurally, the slowest process corresponds to a transition across the barrier between a and b, as indicated by the different signs of the corresponding eigenvector in these states. State a is fluorescent while b is not, and therefore the experiment is sensitive to this transition. As expected, a single time scale with large amplitude (Fig. 2*B*, blue) would be seen in fluorescence correlation spectroscopy (FCS), while the local diffusion processes are fast and have small amplitudes. Fig. 2*C* shows a more complex case of a potential dominated by two major energy barriers. The slowest process (blue eigenvector) changes sign between states a and states b + c, and can thus be assigned to the transition across the a – b barrier. The next-slowest process (red eigenvector) has different signs in states b and c while being near-zero in a. This process thus further subdivides set b + c into b and c, and can be assigned to the transition across the b – c barrier. The fluorescence signal detects the transition across the higher barrier, but the faster transition occurs within the dark region. The system exhibits three-state behavior, and, although the faster transition occurs in the dark, it is nevertheless manifested in the FCS, albeit with a smaller amplitude than the slow process (Fig. 2*D*, blue and red). Fig. 2*E* is similar to *C*, but reversed in that the slow process is now between the dark states while the faster process is the one directly seen in the fluorescence signal. Although this system has three-state kinetics, it will appear as two-state in the FCS as the slow process has a very small amplitude (Fig. 2*F*, blue).

Theory

It is now briefly described how the dynamics on a particular energy landscape gives rise to dynamical fingerprints (Fig. 1 *D*–*C*) and allows their features to be uniquely linked to structural changes in these dynamics. Detailed derivations can be found

in the *SI Appendix*. The theory rests on a description of the conformational dynamics as time-dependent probability transport in state space (28). In practice, this description is approximated by a Markov model, now commonly used to analyze conformational dynamics (6, 10, 29–33). Consider the energy landscape to be partitioned into a number, m, of conformational states, all small enough to allow different metastable states to be distinguished and thus the slow conformation dynamics of interest can be described by interstate transitions and the dynamics is Markovian at time scale τ (6, 29, 30, 32). The transition probability, $T_{ij}(\tau)$, is the probability that the system being at state i will be found in state j a time τ later. The entire slow kinetics is then encoded in the transition matrix, $\mathbf{T}(\tau)$, which is related to a rate matrix $\mathbf{K}$ *via* $\mathbf{T}(\tau) = \exp(\tau\mathbf{K})$. Given the vector $\mathbf{p}(0)$, the ith element of which provides the population of conformation i at time 0, the population of the system at any of the discrete times $k\tau$ is given by (29, 30, 34):

$$\mathbf{p}^{\mathrm{T}}(k\tau) = \mathbf{p}^{\mathrm{T}}(0)[\mathbf{T}(\tau)]^k, \qquad [1]$$

where $[\mathbf{T}(\tau)]^k$ is the kth power of matrix $\mathbf{T}(\tau)$ and the superscript T indicates the transpose. Now consider the (left) eigenvectors, $\mathbf{l}_i$, and eigenvalues, λ_i, of $\mathbf{T}(\tau)$. For every pair of eigenvectors and eigenvalues the following relation holds:

$$\mathbf{l}_i^{\mathrm{T}}\mathbf{T}(\tau) = \lambda_i\mathbf{l}_i^{\mathrm{T}}. \qquad [2]$$

This eigenvalue equation has a straightforward interpretation: each eigenvector $\mathbf{l}_i$ corresponds to a collective transition process moving population between conformational states in which the eigenvector elements are of opposite sign (28). For ergodic systems, there is always one eigenvector, $\mathbf{l}_1 = \pi$, with corresponding eigenvalue $\lambda_1 = 1$, which provides the equilibrium distribution: $\pi^{\mathrm{T}}\mathbf{T}(\tau) = \pi^{\mathrm{T}}$, i.e., the Boltzmann distribution, for a molecular system at equilibrium. All other eigenvectors have eigenvalues less than 1 and represent perturbations from the equilibrium distribution which decay over time towards equilibrium, with a fraction per time unit τ specified by the corresponding eigenvalue λ_i. Thus, the dynamics described by $\mathbf{T}(\tau)$ has the characteristic relaxation times (29, 30, 34):

$$t_i = -\frac{\tau}{\ln(\lambda_i)}. \qquad [3]$$

The above illustrates how the complete kinetics of the molecule can be described as a superposition of collective transition processes represented by the eigenvectors of $\mathbf{T}(\tau)$. Note that the "time resolution" that can be practically described with this approach depends on the number of conformational states, m, used. Splitting into finer states resolves faster processes and the interesting slow processes are unaffected by this, as long as the state definition allows separation of the slowly interconverting conformations.

We can now derive general expressions for time-dependent relaxation profiles, $f(t)$, of kinetic experiments (see *SI Appendix* for details). These profiles have the simple general form:

$$f(t) = \gamma_1 + \sum_{i=2}^{m} \gamma_i \exp\left(-\frac{t}{t_i}\right), \qquad [4]$$

where t_i are the characteristic relaxation times of the system, thus being independent of the experiment conducted or the observables used. The amplitudes γ_i, however, do depend on the type of experiment and the observable: Let $\mathbf{a}$ define the value of a particular observable for every state. Then, experiments that monitor the relaxation of the ensemble average of a from some initially perturbed population $\mathbf{p}'(0)$ towards equilibrium have the amplitudes:

$$\gamma_i^{\mathrm{perturb}} = (\mathbf{a}^{\mathrm{T}}\mathbf{l}_i)([\mathbf{p}'(0)]^{\mathrm{T}}\mathbf{l}_i), \qquad [5]$$

where $\mathbf{u}^T\mathbf{v}$ is equivalent to the scalar product that measures the overlap of two vectors $\mathbf{u}$, $\mathbf{v}$, and $p_i'(0) = p_i(0)/\pi_i$ denotes the relative excess of the initial population with respect to the equilibrium distribution. Correlation functions from equilibrium experiments (such as FCS) have amplitudes given by:

$$\gamma_i^{\mathrm{corr}} = (\mathbf{a}^{\mathrm{T}}\mathbf{l}_i)(\mathbf{b}^{\mathrm{T}}\mathbf{l}_i), \qquad [6]$$

where a and b are spectroscopic observables. For the particular example of FCS, we can simplify this equation. When the system states are defined such that there exists a set of fluorescent states, F, each giving rise to a constant fluorescence intensity, and the complementary set of dark states giving rise to no intensity, then the amplitudes due to conformational relaxations are proportional to:

$$\gamma_i^{\mathrm{fcs}} = \left[\sum_{j\in F} l_{ij}\right]^2. \qquad [7]$$

For FCS, an equivalent expression can be derived from Elson and Magde's theory of fluorescence correlation (35) when a reversible stoichiometry matrix is considered.

The above equations directly provide the structural interpretation of the processes responsible for the relaxation of the signal or the decay of the autocorrelation function: the signal is a weighted sum of exponentials, each term decaying with a characteristic time scale given by t_i. Each relaxation process has an amplitude that depends on the type of experiment and the observables employed. In relaxation experiments, the amplitude depends on how well the observable picks up the structural change involved in the process ($\mathbf{a}^T\mathbf{l}_i$) and on how much the initial perturbation populates relaxation mode i ($[\mathbf{p}'(0)]^{\mathrm{T}}\mathbf{l}_i$). In correlation experiments, the amplitude depends only on the sensitivity of the observable to the structural change involved. In FCS, in particular, the amplitude $[\sum_{j\in F} l_{ij}]^2$ quantifies the change of fluorescence intensity due to contact formation between dyes. The eigenvector corresponding to each term, $\mathbf{l}_i$, provides the structural interpretation for each term, as it corresponds to a collective transition between those states whose signs are opposite (see Fig. 2).

Based on the above theory, we can now give the mathematical expression for the sharp (i.e., noise-free) dynamical fingerprint, $S(t)$, given by:

$$S(t) = \sum_{i=2}^{m} \gamma_i\delta(t - t_i), \qquad [8]$$

with δ denoting the Dirac delta function. The slow part of $S(t)$ consists of discrete spectral lines at times t_i, indicating the relaxation times of all slow conformational processes, which are invariant properties of the system. The amplitudes of these spectral lines depend on the experimental observable; in the present example the fluorescence intensity change. Correspondingly, other types of experiment will have different $S(t)$ with equal positions but different intensities of the peaks associated with them. In principle, $S(t)$ has an additional continuous part close to times around 0, due to fast diffusion processes within the conformations, but this relatively uninteresting part of the fingerprint is only approximated here because only finitely many states m are considered. Note that in the limit of zero measurement noise, $S(t)$ is the inverse Laplace transform of the correlation or relaxation function (see *SI Appendix*).

Methods for extracting fingerprints from both experiment, $S_{\mathrm{exp}}(t)$, and simulation, $S_{\mathrm{sim}}(t)$, are described in the *SI Appendix*. $S_{\mathrm{exp}}(t)$ and $S_{\mathrm{sim}}(t)$ can now be compared. One cannot expect a perfect agreement between these fingerprints, because inaccuracies of the MD force-field will be reflected in the relaxation times and amplitudes that are computed. Moreover the fluorescence model used, i.e., the definition of fluorescent and nonfluorescent states, will affect the amplitudes (although not the time scales)

of the processes apparent in the fingerprint. However, when features between $S_{exp}(t)$ and $S_{sim}(t)$ can be qualitatively matched, the simulation model allows these features to be interpreted structurally through the association with the eigenvectors $\mathbf{l}_i$, i.e., each mode in the dynamical fingerprint can be associated with a particular structural process, thus yielding a detailed structural explanation of the experimentally observable features

Experimental fingerprint estimation has been implemented in scimex (https://simtk.org/home/scimex). Markov models were calculated with the EMMA package (https://simtk.org/home/emma).

Results

We now employ the concepts from above to investigate the conformation dynamics of MR121-(Gly-Ser)$_n$-Trp peptides with different lengths $n = 2$, 3, 5, and 9 where MR121 is a dye that, when specifically excited with a laser, will fluoresce unless it is specifically quenched by Trp which occurs efficiently upon contact formation. Experimentally, the conformational dynamics is probed *via* FCS, i.e., we compute the autocorrelation function of fluctuations of the fluorescence intensity in a very dilute sample. Computationally, all-atom explicit solvent MD simulations of the same peptides have been conducted. Dynamical fingerprints were computed from both experiment and simulation, their features compared, and the structural processes giving rise to these features investigated. The conformational dynamics of Gly-Ser and other flexible peptides were previously measured by various methods (3, 17, 19, 36). Usually, a fitting procedure suggested a single-exponential decay, consistent with a simple two-state model. Our present measurements have a time resolution of about 150 ps compared to at most several ns in previous experiments, and are at least tenfold longer in measurement time than previous FCS experiments, thus having reduced statistical noise and permitting the dynamical fingerprints to be resolved more reliably. The experimental setup and dynamical fingerprint extraction are described in the *SI Appendix* in detail.

The fingerprints computed from the experimental fluorescence correlation functions are shown in Fig. 3 *C–F* (top). Each fingerprint has a negative amplitude at 1.5 ns–2 ns, which is the signature of the limited excited state lifetime (37), i.e., an anticorrelation in the fluorescence signal resulting from the failure to excite an already-excited dye. For this reason, contributions close to this peak are unreliable from experiment and are grayed out in the figure. Furthermore, relaxation times in the range of >500 ns are found which may be due to decay of long-lived triplet states or arise from terms necessary to correct for the imperfect model of the diffusional decay—these do not yield reliable information on conformational transitions and thus the corresponding region is grayed out as well. The remaining ranges of 4 ns–500 ns contain useful information on the relaxation times of conformational transitions and their associated fluorescence amplitudes, and this is validated by the fact that the fingerprint of the pure dye contains no features in these ranges.

For all four peptides, $n = 2$, 3, 5, and 9, most of the amplitude of the experimental fingerprint is in peaks at time scales 4–8, 5–10, 7-30, and 15–50 ns, respectively. Due to these dominant amplitudes, the slow relaxations appear *almost* single exponential and could be well fitted with a single time scale, as in previous studies that used a corresponding two-state model to compute contact-formation rates depending on chain length (3, 36). However, both broadening and multimodality of these features imply that the existence of multiple distinct conformational processes with similar relaxation times is also consistent with the data. Of particular interest is the presence of peaks indicating additional slow processes: the fingerprints of all peptides resolve separate low-amplitude peaks at >30 ns to >200 ns. We have tested the statistical significance of these long-time scale features (*SI Appendix: Table S2*), and found them to have nearly 100% confidence for $n = 2$ and $n = 3$, 93.9% confidence for $n = 9$, but only 77.3% confidence for $n = 5$. These findings indicate that the processes which contribute the strongest to fluorescence change and have been identified in previous experimental studies are probably not the slowest processes in the systems, and the systems may be considerably more complex than consisting only of two metastable states that are open and closed.

In order to obtain detailed insight into the structural events associated with the dynamical processes involved in the experimental signals, extensive all-atom MD simulations for the above systems in explicit solvent were conducted. The state spaces were discretized into ≈200 substates, and each substate classified as either dark (with dye and Trp in van der Waals contact) or

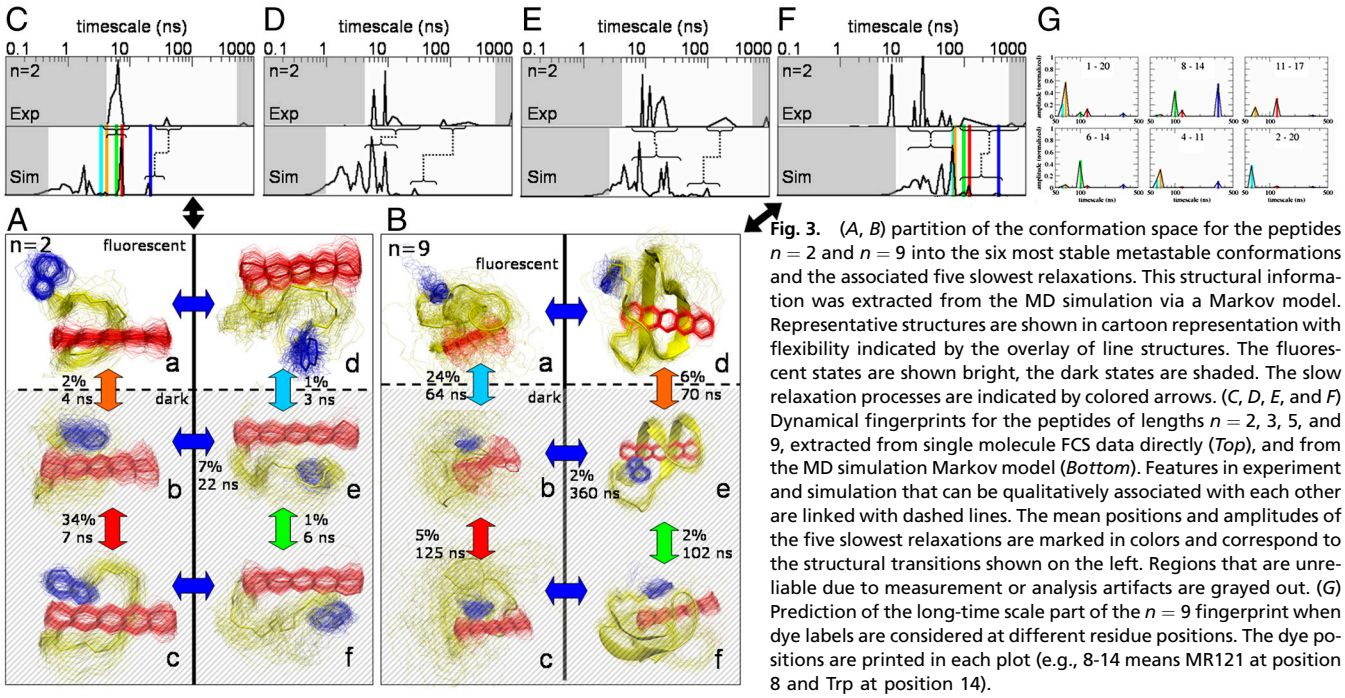

Fig. 3. (*A*, *B*) partition of the conformation space for the peptides $n = 2$ and $n = 9$ into the six most stable metastable conformations and the associated five slowest relaxations. This structural information was extracted from the MD simulation via a Markov model. Representative structures are shown in cartoon representation with flexibility indicated by the overlay of line structures. The fluorescent states are shown bright, the dark states are shaded. The slow relaxation processes are indicated by colored arrows. (*C*, *D*, *E*, and *F*) Dynamical fingerprints for the peptides of lengths $n = 2$, 3, 5, and 9, extracted from single molecule FCS data directly (*Top*), and from the MD simulation Markov model (*Bottom*). Features in experiment and simulation that can be qualitatively associated with each other are linked with dashed lines. The mean positions and amplitudes of the five slowest relaxations are marked in colors and correspond to the structural transitions shown on the left. Regions that are unreliable due to measurement or analysis artifacts are grayed out. (*G*) Prediction of the long-time scale part of the $n = 9$ fingerprint when dye labels are considered at different residue positions. The dye positions are printed in each plot (e.g., 8-14 means MR121 at position 8 and Trp at position 14).

otherwise fluorescent. Markov models of the transition dynamics between discrete states were built with the EMMA package and extensively validated (*SI Appendix*). Both experimental and simulated fingerprints yield very similar correlation functions (*SI Appendix: Fig. S8*). A comparison of fingerprints is more discriminatory and shows that some peaks and features are similar and form a putative match, indicated in Fig. 3. Although not every single peak can be matched with confidence, the similar degree of complexity justifies detailed examination of which structural processes the simulated fingerprint peaks correspond to. This examination is done here for $n = 2$ and $n = 9$ in detail.

The slowest processes can be matched between simulation and experimental fingerprints and have simulated time scales of 22 ± 2, 29 ± 3, 100 ± 5, and 360 ± 14 ns for $n = 2$, 3, 5, and 9, respectively. The slowest process amplitudes are small (between 1.5 and 7 %) but statistically significant. Although it is often assumed in peptide and protein folding that the slowest process is the folding process itself, this is not the case here. None of these slow processes separate open and closed states (see Fig. 3), and thus the slowest time scales cannot be assigned to folding or chain closure. For the smallest peptide, $n = 2$, the slowest process (22 ns) exchanges between sets (a, b, c) and (d, e, f) which differ in the stacking order of the MR121 and Trp ring systems (see Fig. 3*A*). While this process does not separate open and closed states, it involves a significant contribution to the fluorescence intensity change because it changes the probability of occupying an unfolded, or fluorescent, structure, this being 25% for one stacking order and 10% for the other. The accompanying change in effective fluorescence upon swapping the stacking order results in a 7% decay of the autocorrelation function. For $n = 2$, the two next-slowest processes correspond to changes of the Trp orientation relative to the MR121 dye [one splitting (a, b, and c) into (a, b) and c; the other one splitting (d, e, and f) into (d, e) and f]. Similar processes are observed for the $n = 3$ and $n = 5$ peptides, in which the two or three slowest processes again correspond mainly to exchange of stacking order and change in orientation of the labels. This observation confirms a previous finding that for the small peptides the dynamics is dominated by the interaction of the dyes (38, 39). In $n = 9$, the peptide is long enough to form secondary structure and the slowest process involves mainly formation of β-sheet structures (see Fig. 3*B*).

For all peptides, the largest amplitude processes, i.e., those involving the strongest fluorescence changes, arise from multiple processes with relaxation times similar to that of the main peak in the experimental fingerprint. One can thus interpret an experimentally-fitted single exponential as an "effective time scale" that combines a group of different chain closure processes with similar relaxation times. For $n = 2$, the highest-contributing individual process contributes 34% of the decay, has an estimated time scale of 7 ± 0.5 ns, and structurally corresponds to an exchange of end-group orientations (see Fig. 3*A*), and this process involves a strong fluorescence change as it switches between entirely closed states and a rapid equilibrium of closed and open states. For $n = 9$, the process with the maximum intensity of 24% has an estimated time scale of 64 ± 4 ns and directly involves chain closure.

All simulation-derived fingerprints exhibit fast processes in the time ranges <2, 3, 5, and 30 ns for $n = 2$, 3, 5, and 9, respectively, that could not be reliably estimated from the measurement data. All fast processes exchange probability between very few (often two) conformational states and are local because they correspond to fluctuations and diffusion between states that are embedded within the larger basins of the energy surface (see also Fig. 2*B*). The fact that these processes make a significant contribution to the fluorescence change indicates that the energy landscape is not shaped in such a way that the major barriers exactly separate open and closed species, but rather allows for rapid contact formation of the end groups while the main chain stays within a metastable state.

A key concern of the analysis is how to reliably match features of fingerprints derived from simulation and experiment, and, more constructively, how experiments can be used to test the existence of interesting processes proposed by the fingerprint analysis. Systematic errors such as force-field inaccuracies in the simulation and sample impurities in the measurement will affect the positions, amplitudes, and even the existence of fingerprint peaks, making a one-to-one comparison sometimes unreliable. To resolve ambiguities, we propose an approach that uses the present theoretical analysis to design specific new experiments permitting a one-to-one investigation of fingerprint features. This approach involves proposing to place site-specific labels at different positions and predicting the corresponding changes in the fingerprints which can be tested experimentally.

The approach uses the fact that the relaxation time scales t_i are basic physical properties of the molecular system, and thus the positions of the fingerprint peaks are independent of the observable used to probe the conformational dynamics. However, the peak amplitudes directly depend on the observable used, *via* Eq **5** or **6**, and it is thus possible to propose label positions based on the simulation model that maximally enhance the amplitude of selected individual observed relaxation processes while suppressing the other amplitudes. When the predicted enhancement matches experiment, experimental and simulation peaks can be matched with relative confidence even in the presence of systematic simulation or measurement error.

Measurement techniques using site-specific labels that do not significantly affect the conformational dynamics are ideal for this approach, e.g., isotope labeling in NMR, IR spectroscopy, or neutron scattering. In fluorescence-based techniques the dyes should not perturb the relaxation times concerned. Unfortunately, in the system studied here, the dyes are relatively large compared to the GS peptides themselves and changing label positions would effectively change the sequence studied. Nevertheless, we use the MR121-GS_9-W simulation model to illustrate the usefulness of the approach. Assuming that fluorescence quenching can be approximately predicted based on C_α–C_α distances of the labeled residue pairs, fingerprints were predicted for all 190 possible positions of placing the MR121 and Trp dyes in the 20-mer (see *SI Appendix*). It was found that, depending on the labeling position, individual peak amplitudes may indeed be tuned to very small or very large values. Moreover, for most peaks, one or more label positions exist at which the target peak has a large amplitude and the other peaks small amplitudes, indicating that it is possible to design experimental constructs that are optimal probes of individual relaxation processes. For each of the five slowest processes, those label positions were chosen that maximize the fraction of amplitude in the selected process relative to the other four slowest processes, yielding MR121-Trp dye positions 8–14, 11–17, 6–14, 4–11, and 2–20 for the slowest to fifth-slowest process, respectively (see Fig. 3).

Conclusions

The present analysis of fluorescently labeled peptides reveals highly complex processes even for these relatively simple systems. These processes include very fast ones in the ps and few ns range, the overlap of several large-amplitude conformational processes at the time scales experimentally most apparent, and the existence of slow, low-amplitude processes. The slowest processes have signatures of kinetic partitioning (40), i.e., they are not due to direct end-to-end contact formation, but rather to transitions between metastable states that contain rapidly equilibrating dark and fluorescent states in different proportions. This scenario is depicted in the schematic Fig. 1*D*. For the smaller peptides, the slowest processes involve changing the stacking order and

orientations of the fluorophore and the Trp quencher. In the larger peptides, the slowest processes are due to secondary structure rearrangements, including β-sheet formation and dissociation, during which the Trp and MR121 remain close. These results were unexpected, because in all cases observed, the slowest processes involve transitions from dark to dark states or from bright to bright states, which may intuitively seem to be invisible. However, these processes are indeed visible because they switch between regions of conformation space which contain different ratios of dark and fluorescent species. The two-state kinetic picture that was suggested in previous experimental analyses of these systems can now be seen as an effective model combining into a single transition those processes that involve most of the fluorescence intensity change. However, this picture is insufficient to quantitatively describe the complex energy landscape typical of biological systems.

Dynamical fingerprints are a natural approach to reconcile kinetic experiments and molecular simulation, and can be expected to find broad application in the characterization of biomolecular processes. Comparing the features of fingerprints is cleaner and more insightful than comparing relaxation profiles directly or fits to them, because in relaxation profiles all relaxation processes mix at all times. Critically, each simulation-derived peak can be unambiguously associated to a structural process present in the simulation data, yielding a direct and general approach for structurally interpreting kinetic experiments *via* peak-matching.

In a broader sense, the approach presented here illustrates how the intrinsically complex conformational dynamics observed in MD simulations is compatible with the comparatively simple models suggested by standard experimental analyses. The fingerprint prediction approach has shown that in many experimental observables only a few of the slow relaxation time scales are probed with significant amplitudes. This observation suggests that there may be many cases in which systems may appear to be only two- or three-state in a single experiment, while they are in fact more complex. To determine how many states a system possesses it is useful to conduct multiple experiments with labels at different sites. The theory laid out here suggests how simulations can be used in order to propose a small set of experimental constructs to be measured in order to probe individual relaxation processes, and to thus construct a detailed picture of the conformational dynamics of biomolecules step by step. Markov model simulations have already shown considerable complexity in larger and slower systems than the peptides reported here, including metastable denatured states and multiple folding pathways (6, 10, 41). The present approach will help in designing experiments to unravel this complexity.

ACKNOWLEDGMENTS We acknowledge the Deutsche Forschungsgemeinschaft (DFG) for financial support under Grants HN 1201/2-1, SM 63/12-1, NO 825/2-1. Also, Research Center Matheon: the Department of Energy (DOE) BER and ASCR "Multiscale Mathematics" grant number ERKPE84. J.D.C. acknowledges a QB3-Berkeley distinguished postdoctoral fellowship.

1. Frauenfelder H, Sligar SG, Wolynes PG (1991) The energy landscapes and motions of proteins. *Science* 254:1598–1603.
2. Ostermann A, Waschipky R, Parak FG, Nienhaus UG (2000) Ligand binding and conformational motions in myoglobin. *Nature* 404:205–208.
3. Neuweiler H, Löllmann M, Doose S, Sauer M (2007) Dynamics of unfolded polypeptide chains in crowded environment studied by fluorescence correlation spectroscopy. *J Mol Biol* 365:856–869.
4. Jäger M, Nguyen H, Crane JC, Kelly JW, Gruebele M (2001) The folding mechanism of a beta-sheet: the ww domain. *J Mol Biol* 311:373–393.
5. Muff S, Caflisch A (2007) Kinetic analysis of molecular dynamics simulations reveals changes in the denatured state and switch of folding pathways upon single-point mutation of a -sheet miniprotein. *Proteins* 70:1185–1195.
6. Noé F, Schütte C, Vanden-Eijnden E, Reich L, Weikl TR (2009) Constructing the full ensemble of folding pathways from short off-equilibrium simulations. *Proc Natl Acad Sci USA* 106:19011–19016.
7. Onuchic JN, Wolynes PG (2004) Theory of protein folding. *Curr Opin Struc Biol* 14:70–75.
8. Frauenfelder H, et al. (2009) A unified model of protein dynamics. *Proc Natl Acad Sci USA* 106:5129–5134.
9. Noé F, Fischer S (2008) Transition networks for modeling the kinetics of conformational transitions in macromolecules. *Curr Opin Struc Biol* 18:154–162.
10. Bowman GR, Beauchamp KA, Boxer G, Pande VS (2009) Progress and challenges in the automated construction of Markov state models for full protein systems. *J Chem Phys* 131:124101.
11. Jäger M, et al. (2006) Structure-function-folding relationship in a ww domain. *Proc Natl Acad Sci USA* 103:10648–10653.
12. Sadqi M, Lapidus LJ, Munoz V (2003) How fast is protein hydrophobic collapse? *Proc Natl Acad Sci USA* 100:12117–12122.
13. Dumont C, Emilsson T, Gruebele M (2009) Reaching the protein folding speed limit with large, sub-microsecond pressure jumps. *Nat Meth* 6:515–519.
14. Chan CK, et al. (1997) Submillisecond protein folding kinetics studied by ultrarapid mixing. *Proc Natl Acad Sci USA* 94:1779–1784.
15. Volkmer A (2000) One- and two-photon excited fluorescence lifetimes and anisotropy decays of green fluorescent proteins. *Biophys J* 78:1589–1598.
16. Doster W, Cusack S, Petry W (1989) Dynamical transition of myoglobin revealed by inelastic neutron scattering. *Nature* 337:754–756.
17. Lapidus LJ, Eaton WA, Hofrichter J (2000) Measuring the rate of intramolecular contact formation in polypeptides. *Proc Natl Acad Sci USA* 97:7220–7225.
18. Michalet X, Weiss S, Jäger M (2006) Single-molecule fluorescence studies of protein folding and conformational dynamics. *Chem Rev* 106:1785–1813.
19. Hudgins RR, Huang F, Gramlich G, Nau WM (2002) A fluorescence-based method for direct measurement of submicrosecond intramolecular contact formation in biopolymers: an exploratory study with polypeptides. *J Am Chem Soc* 124:556–564.
20. Kim HD, et al. (2002) Mg^{2+}-dependent conformational change of RNA studied by fluorescence correlation and FRET on immobilized single molecules. *Proc Natl Acad Sci USA* 99:4284–4289.
21. Nettels D, Hoffmann A, Schuler B (2008) Unfolded protein and peptide dynamics investigated with single-molecule FRET and correlation spectroscopy from picoseconds to secondst. *J Phys Chem B* 112:6137–6146.
22. Provencher SW (1982) Contin—a general purpose constrained regularization program for inverting noise linear algebraic and integral-equations. *Comput Phys Commun* 27:229–242.
23. Steinbach P (2002) Analysis of kinetics using a hybrid maximum-entropy/nonlinear-least-squares method: application to protein folding. *Biophys J* 82:2244–2255.
24. Klafter J, Shlesinger MF (1986) On the relationship among three theories of relaxation in disordered systems. *Proc Natl Acad Sci USA* 83:848–851.
25. Min W, Luo G, Cherayil BJ, Kou SC, Xie XS (2005) Observation of a power-law memory kernel for fluctuations within a single protein molecule. *Phys Rev Lett* 94:198302.
26. Hagen SJ, Eaton WA (1996) Nonexponential structural relaxations in proteins. *J Chem Phys* 104:3395–3398.
27. Witkoskie JB, Cao J (2004) Single molecule kinetics. i. theoretical analysis of indicators. *J Chem Phys* 121:6361–6372.
28. Schütte C, Fischer A, Huisinga W, Deuflhard P (1999) A direct approach to conformational dynamics based on hybrid monte carlo. *J Comput Phys* 151:146–168.
29. Swope WC, Pitera JW, Suits F (2004) Describing protein folding kinetics by molecular dynamics simulations: 1. theory. *J Phys Chem B* 108:6571–6581.
30. Chodera JD, et al. (2007) Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J Chem Phys* 126:155101.
31. Buchete NV, Hummer G (2008) Coarse master equations for peptide folding dynamics. *J Phys Chem B* 112:6057–6069.
32. Sarich M, Noé F, Schütte C (2010) On the approximation error of Markov state models. *SIAM Multiscale Modeling Simulation* 8:1154–1177.
33. Voelz VA, Bowman GR, Beauchamp K, Pande VS (2010) Molecular simulation of ab initio protein folding for a millisecond folder ntl9. *J Am Chem Soc* 132:1526–1528.
34. Noé F, Horenko I, Schütte C, Smith JC (2007) Hierarchical analysis of conformational dynamics in biomolecules: transition networks of metastable states. *J Chem Phys* 126:155102.
35. Elson EL, Magde D (1974) Fluorescence correlation spectroscopy. i. conceptual basis and theory. *Biopolymers* 13:1–27.
36. Krieger F, Fierz B, Bieri O, Drewello M, Kiefhaber T (2003) Dynamics of unfolded polypeptide chains as model for the earliest steps in protein folding. *J Mol Biol* 332:265–274.
37. Felekyan S, et al. (2005) Full correlation from picoseconds to seconds by time-resolved and time-correlated single photon detection. *Rev Sci Instrum* 76:083104.
38. Daidone I, Neuweiler H, Doose S, Sauer M, Smith JC (2010) Hydrogen-bond driven loop-closure kinetics in unfolded polypeptide chains. *Public Library of Science* 6: e1000645.
39. Doose S, Neuweiler H, Sauer M (2009) Fluorescence quenching by photoinduced electron transfer: a reporter for conformational dynamics of macromolecules. *ChemPhysChem* 10:1389–1398.
40. Thirumalai D, Woodson SA (1996) Kinetics of folding of proteins and RNA. *Acc Chem Res* 29:433–439.
41. Voelz VA, Dill KA (2007) Exploring zipping and assembly as a protein folding principle. *Proteins* 66:877–888.