Article:
Gowdy, J, Batchelor, M, Neelov, I et al. (1 more author) (2017) Non-Exponential Kinetics of Loop Formation in Proteins and Peptides: A Signature of Rugged Free Energy Landscapes? Journal of Physical Chemistry B, 121 (41). pp. 9518-9525. ISSN 1520-6106

https://doi.org/10.1021/acs.jpcb.7b07075

Non-Exponential Kinetics of Loop Formation in Proteins and Peptides: A Signature of Rugged Free Energy Landscapes?

James Gowdy[1], Matthew Batchelor[1], Igor Neelov[2], and Emanuele Paci[1,*]

[1] Astbury Centre for Structural Molecular Biology, University of Leeds, UK

[2] Institute of Macromolecular Compounds of Russian Academy of Sciences, St.Petersburg, Russia

[*]Correspondence to EP (e.paci@leeds.ac.uk)

Abstract

The kinetics of loop formation, *i.e*., the occurrence of contact between two atoms of a polypeptide, remains the focus of continuing interest. One of the reasons is that contact formation is the elementary event underlying processes such as folding and binding. More importantly, it is experimentally measurable and can be predicted theoretically for ideal polymers. Deviations from single exponential kinetics have sometimes been interpreted as a signature of rugged, protein-like, free energy landscapes. Here we present simulations, with different atomistic models, of short peptides with varied structural propensity, and of a structured protein. Results show exponential contact formation kinetics (or relaxation) at long times, and a power law relaxation at very short times. At intermediate times a deviation from either power law or simple exponential kinetics is observed that appears to be characteristic of polypeptides with either specific or non-specific attractive interactions, but disappears if attractive interactions are absent. Our results agree with recent experimental measurements on peptides and proteins and offer a comprehensive interpretation for them.

Introduction

The event of encounter for two atoms of a polypeptide chain is a necessary step for the formation of a persistent intramolecular interaction. It sets a lower limit for the time it takes for a protein to fold and more generally for molecular recognition to occur. The kinetics of formation of a contact between any two atoms of a polypeptide is also an important probe of its dynamics, which is relevant for function. This is true in general and particularly relevant for proteins whose function is not strictly related to their ability to populate a structurally well-defined state.[1-3]

Efforts to characterize the kinetics of formation of a contact between two atoms or group of atoms, and most frequently the two ends of a polypeptide chain, date back two decades to when time-resolved spectroscopy was used to monitor heme absorption following photodissociation of the carbon monoxide complex of denatured reduced cytochrome *c.*[4] More recently, generally applicable methods for probing contact formation have been used to directly measure loop formation in unfolded polypeptide chains.[5] One particularly promising experimental strategy exploits triplet-triplet energy transfer; with a suitable synthetic triplet donor and acceptor the transfer of energy is a diffusion-limited process that occurs at contact. Triplet-triplet energy transfer has also been used between tryptophan and cysteine but the reaction is slow and rates cannot be recovered without referring to a largely uncertain model of the quenching kinetics.[6]

The fraction of unreacted chains has been repeatedly reported to decay exponentially in time.[7-8] More recently, Kiefhaber and coworkers used diffusion-limited triplet-triplet energy transfer to probe contact formation over six orders of magnitude in time, ranging from picoseconds to microseconds.[9] Their results revealed processes occurring on different time scales that indicate a hierarchy of different peptide motions on the free energy surface. Non-exponential decay of the population of open (contact not yet formed) states on the sub nanosecond time scale was interpreted as a signature of motions within local wells on the energy landscape. Within these wells peptides can form loops by undergoing local motions without having to traverse significant barriers. Exponential kinetics observed on longer, nanosecond time scales were attributed to chain diffusion and the exploration of a greater proportion of conformation space through larger-scale motions. Their results indicate similar properties, at least locally, for the free energy landscapes of native proteins and unfolded polypeptide chains. Local energy minima reduce the size of accessible conformation space and accelerate the conformational search for energetically favorable local intrachain contacts.

While the methods mentioned above probe an equilibrium population of polypeptide chains, Volk and coworkers have exploited a technique where two residues of a polypeptide chain are chemically modified and joined together by a disulfide bond.[10-11] An aryl disulfide chromophore constrains the chain in a conformation away from equilibrium and can be rapidly photolysed (in less than 200 fs). The resulting thiyl radicals either undergo geminate recombination or diffuse apart as the polypeptide chain relaxes to equilibrium. The geminate pair is assumed to undergo diffusion-controlled recombination, and the radical population has been observed to decay

with non-exponential kinetics over a time range from picoseconds to microseconds. The same result has been recently shown to hold for a protein, which was constrained in a non-native conformation before a disulfide bond between two residues was photolysed.[12-13]

We were inspired by these experimental results that characterize the kinetics for loop formation, and intrigued to look at the relatedness of the two methods employed. Here we recapitulate existing theory and demonstrate the equivalence of these two experimental strategies, *i.e*., one where the time it takes for a loop to form is measured as an average for equilibrium conformations, and one where the loop formation time is measured selecting initial conformations where the loop is present. We also design, perform and analyze molecular dynamics simulations to interpret and reconcile the somewhat unexpected experimental results.

Atomistic simulation has been valuably used in the past to address the kinetics of loop formation in peptides.[14] One important use of computational models and simulation is to verify experimental assumptions and explore alternative ones.[6, 15] At the same time, experiments that probe the formation of contacts on short timescales (tens of nanoseconds for short unstructured peptides) are directly accessible by simulation and are a valuable tool for validating a range of assumptions.[6, 16]

Our simulations have been performed using a number of different models. A united-atom, transferable model was used to simulate short unstructured peptides; a variant of this, where attractive interactions are switched off has also been used. Another model used here is coarse-grained and native-centric, representing a protein with a given native structure on a minimally frustrated, funneled landscape. The polypeptide models have been chosen so that the distribution of end-to-end contact formation can, in all cases, be sampled rigorously. For longer or structured peptides and proteins, however, brute-force, sufficiently long equilibrium simulations may be practically unfeasible; the distributions of times of contact formation between the two ends of the chain converges particularly slowly, since such events are rare and become exponentially less frequent with chain length. Advanced sampling approaches may be used, however, these should be carried out in such a manner that the kinetics are preserved.[17]

Our results show, for all models considered, exponential kinetics of loop formation on long timescales. On short timescales our simulations replicate the remarkable deviation from simple exponential kinetics reported in advanced experimental studies. These “short” timescales can effectively extend for up to four orders of

magnitude in time, or longer depending on the system. Results are discussed in the context of the experimental finding that non-exponential kinetics persist to the millisecond timescale as recently reported by Volk and collaborators from non-equilibrium measurements[13] and the non-exponential kinetics observed by Kiefhaber and coworkers on timescales of 10 ns or less from equilibrium measurements.[9] We conclude that both sets of measurements probe effectively the free energy landscape of the polypeptide, and are likely not artifacts related to the finite speed of the chemical changes being probed or the fact that the quenching rate may be dependent on the distance between the two reacting groups.[6] We also conclude that the non-exponential kinetics observed can be related to the existence of a hierarchy of states due to the presence of attractive intramolecular interactions. The interactions are not necessarily native-like, and thus are not the signature of a funneled, protein-like free energy landscape. We also show that if long-range and attractive interactions are absent, the kinetics of loop formation expected for an ideal polymer is recovered; thus non-native interactions provide, at least in part, an alternative explanation for the non-exponential power law kinetics observed over a broad range of times.

Theory and methods

If r is the distance between the first and the last atom of a polypeptide chain, the two ends are in contact if $r < r_c$, where r_c is the reaction radius. Experimentally r_c depends on the technique used, and it is generally assumed to correspond to van der Waals contact between the reactive groups engineered at the two ends of the polypeptide chain (the results below also hold if the reactive groups are anywhere in the polypeptide chain). We define $N_{eq}(t)$ as the number of molecules, initially at equilibrium, for which the two ends have not been in contact after a time t and were initially not in contact at $t = 0$. Likewise, $N_{ring}(t)$ is the number of molecules for which the two ends have not been in contact after a time t but were in contact at $t = 0$, *i.e.*, the initial conformation is a ring.

For both "equilibrium" and "ring" quantities the surviving fraction of unreacted molecules, $S(t) = N(t)/N(0)$ is their normalized counterpart. The following remarkable relation (see supplementary information for both a rigorous derivation and a schematic illustration based on simulation analyses) holds between $S_{eq}(t)$ and $S_{ring}(t)$:

$$\frac{S_{ring}(t)}{\int_0^\infty S_{ring}(t)\,dt} = -\frac{dS_{eq}(t)}{dt}$$

A time-dependent rate constant, $k_{ring}(t)$, can be defined for the decay in the survival of extended polymer populations.[18] The definition of $k_{ring}(t)$ is based on the rate equation

$$\frac{dS_{ring}(t)}{dt} = -k_{ring}(t)S_{ring}(t)$$

which can be re-arranged as:

$$k_{ring}(t) \equiv -\frac{\frac{dS_{ring}(t)}{dt}}{S_{ring}(t)} = -\frac{d\ln S_{ring}(t)}{dt} = -\frac{d}{dt}\left[\ln\left(-\frac{dS_{eq}(t)}{dt}\right)\right]$$

The average end-to-end contact formation time for an equilibrium ensemble of initial conditions is:

$$\tau_{eq} = \int_0^\infty dt\, S_{eq}(t)$$

If $S_{eq}(t) = \exp(-kt)$, then $k_{ring}(t) = k$ and $\tau = 1/k$.

In Table 1 are reported expressions for $k_{ring}(t)$ for different functional forms for $S_{eq}(t)$. In the supplementary information we report the same quantities for a broader set of functions, which have been previously used to fit the kinetics of loop formation.

	$\boldsymbol{S_{eq}(t)}$	$\boldsymbol{S_{ring}(t)}$	$\boldsymbol{k_{ring}(t)}$
Single exponential	e^{-bt}	e^{-bt}	b
Power law	$1 - at^b$	abt^{b-1}	$\frac{1-b}{t}$

Stretched exponential	$e^{-(bt)^c}$	$\frac{ce^{-(bt)^c}(bt)^c}{t}$	$bc(bt)^{c-1} - \frac{c-1}{t}$

Table 1. Instantaneous rate of ring formation $\boldsymbol{k_{ring}(t)}$ for different forms of the survival probability of the open conformation $\boldsymbol{S_{eq}(t)}$.

Molecular dynamics simulations of short polypeptides have been performed with two different molecular models. The most "chemically accurate" model considered here is based on the united atom CHARMM19 model[19] together with a continuum model for the solvent.[20] In this model, polar hydrogen atoms (*e.g.* in N–H groups) are treated explicitly whereas the non-polar hydrogen atoms (*e.g.* in methylene or methyl groups) are grouped together as a single entity with the attached heavy atom. Langevin dynamics has been used with a collision frequency of 3 ps^{-1}, corresponding to a friction coefficient about one order of magnitude lower than water but still in a high friction regime, *i.e.*, such that timescales depend linearly on friction and mechanisms are not affected.[21] Simulations were performed at a temperature of 300 K, with a timestep of 2 fs and were all at least 32 μs long. The second model, a soft-sphere (SS) model,[22] was also used. This model lacks electrostatics and solvation contributions, and the Lennard-Jones interactions are smoothly switched off continuously using a third order spline between 3 and 3.5 Å (resulting in an annihilation of all attractive interactions at atom-to-atom distances beyond this length). All the other parameters of the simulation were the same as for the united atom model.

In a third, structure-based model, each residue in the protein Sac7d is simplified to a single 'bead' located at the site of the alpha carbons. The pairwise attractive interactions are determined based on the type of amino acid present and their all-atom conformation in the native structure.[23] The reference native structure for the structure-based model of Sac7d was determined by solution-state NMR spectroscopy, and is deposited in the PDB with the code 1SAP.[24] Simulations for the structure-based model were also performed using Langevin dynamics. A timestep of 15 fs was used and simulations were 300 μs long. Simulations were performed over a broad range of temperatures between 230 K and 370 K.

Simulation trajectories were analyzed from an "equilibrium" perspective as this yields much larger datasets compared to a "ring" perspective analysis. End-to-end distances (distance between N-terminal N and C-terminal C) were calculated at *each* step. Contact events are enumerated and annotated as per the illustration in

Supplementary Fig. S1A–C. Each step in the trajectory where the ends are not in contact (*i.e.*, at a distance larger than the cut-off r_c) is a potential equilibrium start position. The contact formation time, or equilibrium-to-ring first passage time (FPT), was determined for each start position, and a frequency distribution of FPTs generated. Integration and normalization led to $S_{eq}(t)$ (see Supplementary Fig. S1E, G, I, K).

For the united-atom models r_c was set to 5 Å. For the coarse-grained model, where only distances between residues are defined, a r_c value of 8 Å was used. The average end-to-end contact formation time varies moderately (as expected longer times for smaller r_c values) but for values around those chosen, the onset of the different kinetic regimes remains unchanged (see Supplementary Figure S2).

Results

We first considered a number of short (up to 12 residues) homopolypeptides. In Figure 1A is shown the probability of survival of the open state (the two end groups not in contact) for three peptides: a glycine 6mer (G6), an alanine 12mer (A12) and a serine 6mer (S6). These have been plotted in two different ways: the log-linear plot (inset) of $S_{eq}(t)$ shows that the decay is well approximated by a single exponential at longer timescales. An exponential decay of $S_{eq}(t)$ has been widely reported.[5, 8]

The log-log plot of $1 - S_{eq}(t)$, however, shows a remarkable deviation from exponential decay at short times. A power law behavior at short times is a well-known consequence of non-Markovian dynamics due to chain connectivity[25], but in the case of the peptides above the kinetics of loop formation appears to be, perhaps unsurprisingly, more complex than for ideal polymers.

To investigate the origin of this complex kinetics a model identical to the implicit solvent model used above, but without electrostatic and attractive interactions (soft-sphere or SS model) can provide useful insight. The survival probability using the SS model is plotted, as $1 - S_{eq}(t)$, in Figure 1B for the same three homopolypeptides. At short times, *i.e.*, at times shorter than the relaxation time $\tau = 1/k$, $S_{eq}(t)$ decays as $1 - bt^a$, as expected for an ideal polymer such as the Rouse model.[26-28] The Rouse model consists essentially of non-interacting beads connected through springs. For the Rouse model $S_{eq}(t)$ can be obtained numerically[14] (and in some cases analytically[28]) from first principles. For the three homopolypeptides G6, S6, and A12 the power law behavior of $1 - S_{eq}(t)$ is observed over about three orders of

magnitude in time (least square fits shown in Figure 1A and 1B), with the exponent a estimated between 0.86 and 0.90.

The non-exponential kinetics, which extends to timescales considerably larger than those at which a power law is observed, appears to depend on the presence of attractive interactions, but only moderately on the sequence and on the length of the peptide. The timescale at which kinetics becomes exponential and the amplitude of the non-exponential component depends instead on the sequence and length of the peptide.

Before investigating further the complex kinetics observed at short times we consider the end-to-end loop formation in a structure-based coarse-grained protein model where only alpha carbon atoms are explicitly represented. We chose such a model because properties can be thoroughly sampled at equilibrium with available computational resources. We also chose a small, fast folding protein to obtain fully converged distributions of end-to-end contact times. Sac7d is a α-β DNA-binding protein and within the model we used folds reversibly, in a two-state manner, over a broad temperature range.

In Figure 2 is reported $1 - S_{eq}(t)$ for the protein Sac7d at three different temperatures, below and above the melting temperature of about 282 K (the temperature at which the protein is 50% folded). At all temperatures a single exponential fits $S_{eq}(t)$ for times longer than ~10–50 ns, and a unique power law, with an exponent independent of the temperature $(a = 0.81)$ fits at times between 5 and 1000 ps. The range where the decay of $S_{eq}(t)$ is neither power law nor exponential is larger at the lower temperature, 275 K, where the protein is in the folded state about 60% of the time, and smaller at 305 K, where the protein is 94% in an unfolded state.

We then analyzed more closely the kinetics, fitting the computed curves with specific, functional forms for the survival probability. In Figure 3, $S_{eq}(t)$ is compared for the three different models, focusing on a single case for each model. In Figures 3A and 3B $S_{eq}(t)$ of A12 is shown for the implicit solvent and the soft-sphere model, respectively. As previously noted these differ substantially. For the soft-sphere model $1 - S_{eq}(t)$ can be fitted with a power law at short times and a single exponential at long times (with time constant of ~0.93 ns).

For the united-atom model, for times longer than 1 ns, $S_{eq}(t)$ decays exponentially, with a time constant of 14.7 ns compared to an average end-to-end formation time of 14.5 ns. The sum of two exponentials, with a second time constant of ~24 ps, fits the

curve considerably better up to about 30 ps. Using a sum of one exponential and one stretched exponential one can fit $S_{eq}(t)$ accurately up to 1 ps, and between 10 fs and 1 ps $S_{eq}(t)$ is instead well approximated by a power law.

This result is consistent with what Fierz *et al.*[9] observed for a number of short, unstructured peptides. The short time behavior of the survival probability $S_{eq}(t)$ of the state where the triplet state is excited could be fitted either by a series of exponentials, or at least two exponentials and a stretched exponential. The stretch exponents β we obtained (0.69 for A12 and 0.58 for S6) are close to those reported by Fierz *at al.*[9] for *Xan*–Ser$_2$–*NAla* (0.72) and *Xan*–Ser$_6$–*NAla* (0.70), suggesting that experiment and simulation probe the same end-to-end dynamics of the polypeptide. The time constants of the two exponentials estimated from simulation for S6 are 50 ps and 700 ps while those determined experimentally for *Xan*–Ser$_6$–*NAla* (260 ps and 20 ns): the timescales of the simulations are expected to be roughly one order of magnitude faster given that they are performed in a medium with a friction about one order of magnitude lower than water. On the other hand, the amplitude of the non-exponential decay is considerably smaller than that reported by Fierz *at al.,*[9] which is about 1% for A12 and about 5% for S6.

In Figure 3C is shown $1 - S_{eq}(t)$ for the structure-based model of the protein Sac7d at 275 K. Here the single exponential approximation, with time constant ~132 ns, breaks down at times shorter than ~30 ns; at times below 1 ns a power law with exponent 0.82 fits $1 - S_{eq}(t)$.

At times between 1 and 30 ns neither a power law nor a single exponential describe the kinetics of contact formation. Whether physically meaningful or not, an excellent fit can be obtained if the single exponential is replaced by the sum of four exponentials (blue line in Figure 3C) or by the sum of an exponential (with time constant 5 ns) and a second exponential with an exponent β = 0.78.

Figure 3 also shows $k_{ring}(t)$ as defined by Hochstrasser, Volk and collaborators in Refs[10, 13, 18] (see Eq. 2). Experimentally $k_{ring}(t)$ was measured as the logarithmic derivative of the concentration of chains yet to recombine and form a ring at a time t after the disulfide bond keeping the two ends in contact was broken. The rate constant k_{ring} can be seen as a generalized rate of reformation of the end-to-end contact starting from a conformation where the two ends are in contact, and it is simply the inverse contact formation time in the case of single exponential relaxation (and in the case of single exponential kinetics it is irrelevant whether initial conformations are extracted from an equilibrium ensemble or from an ensemble

where the two ends are in contact, see Supplementary Information). In Figures 3D, 3E and 3F the time-dependence of $k_{ring}(t)$ is shown for the fits of $S_{eq}(t)$ in Figures 3A, 3B and 3C, respectively. As shown in Table 1, when $S_{eq}(t)$ decays exponentially (red curves), k_{ring} is simply a constant (*i.e.*, if $S_{eq}(t) \sim e^{-bt}$ then $k_{ring} = b$). When $S_{eq}(t)$ decays as a power law (green curves), k_{ring} is exactly proportional to the inverse of time (*i.e.*, if $S_{eq}(t) \sim 1 - at^b$ then $k_{ring} = (1-b)/t$).

Our result compares interestingly with that reported by Volk and coworkers for a helical peptide[18] and more recently for the 174-amino acid protein N-PGK (N-terminal domain of phosphoglycerate kinase from *Geobacillus steareothermophylus*).[13] From a direct measurement at different times they reported $k_{ring} \sim t^{-(0.94 \pm 0.03)}$ and this was valid over a broad time range from ps to ms. Our results show that $k_{ring} \sim t^a$, with $a = -1$ at very short times and $a = 0$ (*i.e.*, $k_{ring} \sim {}^1/_\tau = constant$) for long times. At intermediate timescales, for the protein model considered here, k_{ring} has a complex dependence on time that extends over four orders of magnitude in time. This is in line with the remarkable result recently reported by Milanesi *et al.,*[13] in which the time dependence of the rate extends up to milliseconds.

Analysis of the simulation of the structure-based model of the protein Sac7d allows direct analysis of the folding kinetics. The time required for k_{ring} to reach a constant value is comparable to, or larger than, the unfolding time. This is unsurprising since the two ends, or the two intrachain residues probed experimentally, cannot form a contact while the protein is in its native state. The time dependence encompasses the time it takes the protein to equilibrate from the initial unfolded conformation, and the time to possibly fold and necessarily unfold so that contact between the two ends can be reformed. The fact that N-PGK has a large folding and unfolding time[29] could then also explain the surprising time dependence of k_{ring} up to the millisecond timescale.

For completeness, plots of the potential of mean force (PMF) for the different systems considered above are shown in Figure 4. The PMF is the negative logarithm of the probability distribution of the end-to-end distance. For the more realistic model (atomistic with implicit solvent) we observe a number of local minima depending on the sequence (Figure 4A), and for all peptides a local minimum exists corresponding to when the two ends are in van der Waals contact (at about 4 Å). For A12 the PMF presents a minimum at around 17.5 Å corresponding to a populated helical state (~1.5 Å/residue). G6 and S6, instead appear to be disordered, in the sense that no

significantly populated states can be identified. For the short peptides, in the absence of attractive interactions (Figure 4B) the PMF is polymer-like, *i.e*., rather flat, except at short and large distances. In all cases $S_{eq}(t)$ reveals complex kinetics at short times regardless of the different thermodynamic properties. Comparison with what we observed for the more realistic model, suggests that complex kinetics can be attributed to non-specific attractive interactions, and not necessarily to the presence of significantly populated conformational states, *i.e*., local minima on the free energy surface.

In the case of the structure based coarse-grained model of the protein Sac7d, the PMF (Figure 4C) is polymer-like at temperatures well above the mid-point temperature. At lower temperatures it has minima corresponding to the native state (26 Å) and a native-like intermediate (30 Å). There is no minimum when the two ends are in contact because this is not a native contact and thus is set to be repulsive in the model.

Discussion

Simulations covering all relevant timescales using different models of polypeptides provide insight into the kinetics of loop formation. A number of recent experimental studies have shown intriguing deviations from exponential kinetics that these simulations reproduce and provide an explanation for.

Two recent experimental observations of anomalous kinetics of loop formation in structured and unstructured peptides and proteins have motivated this work.[9, 13] Both experimental setups can explore a broad range of timescales. Experimental measurements are substantially different, in that one probes loop formation for an ensemble of molecules at equilibrium while the other probes loop reformation in an ensemble of molecules where two residues are in contact at time zero (in a very rarely occupied conformation in the absence of a constraint). Also, interpretation of measurements obtained with either experimental technique relies on the assumption that the reaction between the two reacting groups is diffusion-limited. This is an important assumption since Makarov and co-workers have recently shown that if the quenching rate of two reactive groups depends even mildly on their distance it cannot be assumed that the reaction is diffusion-controlled,[30] *i.e*. there is no such thing as a “diffusion-controlled limit”.

In this study we have assumed an idealized situation, where collision between the reactive groups, in this case the polymer ends, takes place when their distance falls

below a certain radius as in previous simulation studies.[14, 31-32] We addressed both types of experiment by performing long equilibrium simulations, and calculating times of loop formation either from an ensemble of initial configurations that represent equilibrium, or one that satisfies the condition that the two ends are in contact at time zero. We showed that both measurements probe the same kinetics and that the probability that a contact has not occurred after time t, $S(t)$, is exactly related in the two cases.

Both sets of experimental measurements[9, 13] showed "anomalous kinetics" that have been explained using properties of the free-energy surface, which are key for both the folding and the function of proteins. In the case of the measurements of Kiefhaber and coworkers[9], anomalous kinetics consist of a relaxation of $S_{eq}(t)$ that could only be fitted with a stretched exponential or with multiple exponentials at short times. While the proposed functional forms were not specifically justified, the observed separation of time scales in the dynamics of loop formation was explained as originating from motions on different hierarchical levels of the free energy landscape. For structured proteins such a hierarchy of levels has been demonstrated using temperature jump experiments that have shown similarly complex kinetics on the microsecond time scale.[33] However, it is remarkable that largely unstructured peptides where no populated structured states are present also show analogous kinetics. Here we showed that multiple exponentials or a stretched exponential fit relatively well to $S_{eq}(t)$ calculated for either disordered peptides with a realistic force-field or a structured protein with a structure-based force field.

Volk and coworkers[13] showed instead that the time derivative of $S_{ring}(t)$, which they refer to as "instantaneous rate", was not a constant as expected from single exponential decay, but a power law with an unusual exponent, -0.94. While a power law with exponent -1 is expected for any polymer at very short times, and is observed in the simulations (between 1 ps to 1 ns depending on the model), Milanesi *et al.*[13] observed a power law behavior over nine orders of magnitude in time, from 10 ps to 1 ms, when systems as different as a fast folding short helix and a slow folding protein are considered simultaneously. The results presented here show that the instantaneous rate, or $k_{ring}(t)$, has typical polymer behavior at very short times ($k_{ring} \sim t^{-1}$), and is a constant at long times ($k_{ring} \sim t^{0}$). At intermediate times it may locally be fitted with a power law $k_{ring} \sim t^{a}$ with $0 \leq a \leq 1$ but no consistent power law behavior is observed; this is the same time regime where $S_{eq}(t)$ can be fitted with a stretched exponential as proposed by Fierz *et al.*[9] Interestingly, such anomalous

kinetics can extend to long timescales, depending on the features of the free energy landscape of the system considered; this was the main finding reported by Milanesi *et al.*[13] The unusual exponent -0.94 has been explained as arising from subdiffusion of the polypeptide chain[13, 34] due to "inhomogeneous trapping due to the multitude of local interactions which give rise to the rugged potential energy landscape with its hierarchy of well depths".[13] Milanesi *et al.*[13] showed that an approximate power law behavior for $k_{ring}(t)$ can be obtained by modeling a random walk on a landscape where well depths are distributed exponentially, even neglecting chain connectivity. Indeed, the dynamics of the end-to-end distance of a polypeptide chain is subdiffusive, although this may not be an attribute of protein dynamics but only of its projection on a reaction coordinate that is a single interatomic distance.[35] The consequence of projection-dependent subdiffusion have been shown to be observable over 13 decades in time;[36-37] self-similar, non-equilibrium behavior has been observed in the autocorrelation of a distance in atomistic simulations beyond the microsecond timescale.[38]

Here we have also provided some physical insight on the origin the anomalous kinetics, showing that is related to attractive non-specific interactions. If only steric interactions are present, the kinetics become polymer-like, with a power law behavior at times shorter than the Rouse time. Simulations provide information on such short timescales that have not, so far, been probed experimentally (<1 ns), where dynamics is non-Markovian and relaxation follows a power law as predicted for ideal polymers. We observe a universal power law for all models studied ($1 - S_{eq}(t) \sim t^{\alpha}$) with $\alpha \sim 0.8–0.9$. Yeung and Friedman[27] report that for very long Rouse chains with equilibrium initial conditions $\frac{dS_{eq}(t)}{dt} \sim t^{-1/4}$, thus $\alpha - 1 = -1/4$, *i.e.*, $\alpha = 0.75$, close to the value we found. Indeed, in this regime $k_{ring} \sim t^{-1}$ for any α close to unity, such as $k_{ring} \sim t^{-(0.94 \pm 0.03)}$ as has been reported by Milanesi *et al.*[13]

Conclusion

The results presented here reconcile a number of experimental measurements of end-to-end contact formation in apparent contradiction. For those experiments that predict purely exponential relaxation: we show that indeed the amplitude of the non-exponential phase is negligible if experiments probe a population of molecules at equilibrium. For those experiments that show non-exponential relaxation, either only for short times, or on all timescales explored (up to millisecond timescales): we show

that indeed the non-exponential kinetics can be observed on relatively long timescales, which depend on the system; for proteins this is related to the folding and unfolding timescales.

We also show that the different empirical models used to fit the non-exponential kinetics (stretched exponential for $S_{eq}(t)$ and unusual power law for the time-dependent "instantaneous rate constant" or $k_{ring}(t)$) give similar results, and overlap to a large extent with our simulation results.

Measurements using the approach of Milanesi *et al.*[13] on disordered peptides of various lengths and fast folding proteins should show that k_{ring} has a time dependence on a broad range of timescales but eventually converges to a constant at long timescales, *i.e*., at times on the order of the inverse of this constant.

More than 10 years ago Eaton *et al.*[39] stated that a motivation for developing fast kinetic methods was "to provide a much-needed reality check on computer simulations, which are flooding the protein-folding literature". In this paper we show that simulation can also provide a reality check on the way experimental data are interpreted, and a consistency check on experimental results that are either not obviously related, or even which appear contradictory.

Supporting Information Available

In supporting information a derivation of the relationship between equilibrium and ring survival probabilities is provided, as well as the analytical relation between equilibrium and ring survival probabilities and rates for special functional forms. The effect of the contact radius r_c on S_{eq} is shown in Figure S2.

Acknowledgments

This paper is dedicated to the memory of Alexei Likhtman. Enlightening discussions on polymer dynamics with Alexei inspired this work. JG was funded by a Wellcome Trust PhD studentship (093792/Z/10/Z).

References

1. Tran, H. T.; Mao, A.; Pappu, R. V., Role of backbone-solvent interactions in determining conformational equilibria of intrinsically disordered proteins. *J. Am. Chem. Soc.* **2008,** *130* (23), 7380-7392.
2. Vitalis, A.; Wang, X.; Pappu, R. V., Quantitative characterization of intrinsic disorder in polyglutamine: insights from analysis based on polymer theories. *Biophys. J.* **2007,** *93* (6), 1923-1937.
3. Mittal, J.; Yoo, T. H.; Georgiou, G.; Truskett, T. M., Structural ensemble of an intrinsically disordered polypeptide. *J. Phys. Chem. B* **2013,** *117* (1), 118-124.
4. Jones, C. M.; Henry, E. R.; Hu, Y.; Chan, C. K.; Luck, S. D.; Bhuyan, A.; Roder, H.; Hofrichter, J.; Eaton, W. A., Fast Events in Protein-Folding Initiated by Nanosecond Laser Photolysis. *Proc. Natl. Acad. Sci. U. S. A.* **1993,** *90* (24), 11860-11864.
5. Bieri, O.; Wirz, J.; Hellrung, B.; Schutkowski, M.; Drewello, M.; Kiefhaber, T., The speed limit for protein folding measured by triplet-triplet energy transfer. *Proc. Natl. Acad. Sci. U. S. A.* **1999,** *96* (17), 9597-9601.
6. Yeh, I. C.; Hummer, G., Peptide loop-closure kinetics from microsecond molecular dynamics simulations in explicit solvent. *J. Am. Chem. Soc.* **2002,** *124* (23), 6563-6568.
7. Krieger, F.; Fierz, B.; Bieri, O.; Drewello, M.; Kiefhaber, T., Dynamics of unfolded polypeptide chains as model for the earliest steps in protein folding. *J. Mol. Biol.* **2003,** *332* (1), 265-274.
8. Lapidus, L. J.; Eaton, W. A.; Hofrichter, J., Measuring the rate of intramolecular contact formation in polypeptides. *Proc. Natl. Acad. Sci. U. S. A.* **2000,** *97* (13), 7220-7225.
9. Fierz, B.; Satzger, H.; Root, C.; Gilch, P.; Zinth, W.; Kiefhaber, T., Loop formation in unfolded polypeptide chains on the picoseconds to microseconds time scale. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104* (7), 2163-2168.
10. Volk, M., Fast initiation of peptide and protein folding processes. *Eur. J. Org. Chem.* **2001,** (14), 2605-2621.
11. Volk, M.; Gnanakaran, S.; Gooding, E.; Kholodenko, Y.; Pugliano, N.; Hochstrasser, R. M., Anisotropy measurements of solvated HgI2 dissociation: Transition state and fragment rotational dynamics. *J. Phys. Chem. A* **1997,** *101* (4), 638-643.
12. Milanesi, L.; Jelinska, C.; Hunter, C. A.; Hounslow, A. M.; Staniforth, R. A.; Waltho, J. P., A method for the reversible trapping of proteins in non-native conformations. *Biochemistry* **2008,** *47* (51), 13620-13634.
13. Milanesi, L.; Waltho, J. P.; Hunter, C. A.; Shaw, D. J.; Beddard, G. S.; Reid, G. D.; Dev, S.; Volk, M., Measurement of energy landscape roughness of folded and unfolded proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109* (48), 19563-19568.
14. Pastor, R. W.; Zwanzig, R.; Szabo, A., Diffusion limited first contact of the ends of a polymer: Comparison of theory with simulation. *J. Chem. Phys.* **1996,** *105* (9), 3878-3882.
15. Voelz, V. A.; Singh, V. R.; Wedemeyer, W. J.; Lapidus, L. J.; Pande, V. S., Unfolded-state dynamics and structure of protein L characterized by simulation and experiment. *J. Am. Chem. Soc.* **2010,** *132* (13), 4702-4709.
16. Feige, M. J.; Paci, E., Rate of loop formation in peptides: a simulation study. *J. Mol. Biol.* **2008,** *382* (2), 556-565.
17. Shalashilin, D. V.; Beddard, G. S.; Paci, E.; Glowacki, D. R., Peptide kinetics from picoseconds to microseconds using boxed molecular dynamics: power law rate coefficients in cyclisation reactions. *J. Chem. Phys.* **2012,** *137* (16), 165102.
18. Volk, M.; Kholodenko, Y.; Lu, H. S. M.; Gooding, E. A.; DeGrado, W. F.; Hochstrasser, R. M., Peptide conformational dynamics and vibrational stark effects following photoinitiated disulfide cleavage. *J. Phys. Chem. B* **1997,** *101* (42), 8607-8616.
19. Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; *et al*., CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009,** *30* (10), 1545-1614.
20. Ferrara, P.; Apostolakis, J.; Caflisch, A., Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins* **2002,** *46* (1), 24-33.
21. Best, R. B.; Hummer, G., Diffusive model of protein folding dynamics with Kramers turnover in rate. *Phys. Rev. Lett.* **2006,** *96* (22), 228104.
22. Paci, E.; Karplus, M., Forced unfolding of fibronectin type 3 modules: an analysis by biased molecular dynamics simulations. *J. Mol. Biol.* **1999,** *288* (3), 441-459.

23. Karanicolas, J.; Brooks, C. L., 3rd, Improved Go-like models demonstrate the robustness of protein folding mechanisms towards non-native interactions. *J. Mol. Biol.* **2003,** *334* (2), 309-325.
24. Edmondson, S. P.; Qiu, L.; Shriver, J. W., Solution structure of the DNA-binding protein Sac7d from the hyperthermophile Sulfolobus acidocaldarius. *Biochemistry* **1995,** *34* (41), 13289-13304.
25. Doi, M.; Edwards, S. F., *The Theory of Polymer Dynamics*. Clarendon: Oxford, 1988; p xiii, 391 p.
26. Yeung, C.; Friedman, B., Cyclization of Rouse chains at long- and short-time scales. *J. Chem. Phys.* **2005,** *122* (21), 214909.
27. Yeung, C.; Friedman, B. A., Relation between cyclization of polymers with different initial conditions. *Europhys. Lett.* **2006,** *73* (4), 621-627.
28. Likthman, A. E.; Marques, C. M., First-passage problem for the Rouse polymer chain: An exact solution. *Europhys. Lett.* **2006,** *75* (6), 971-977.
29. Parker, M. J.; Spencer, J.; Clarke, A. R., An integrated kinetic analysis of intermediates and transition states in protein folding reactions. *J. Mol. Biol.* **1995,** *253* (5), 771-786.
30. Cheng, R. R.; Uzawa, T.; Plaxco, K. W.; Makarov, D. E., The rate of intramolecular loop formation in DNA and polypeptides: the absence of the diffusion-controlled limit and fractional power-law viscosity dependence. *J. Phys. Chem. B* **2009,** *113* (42), 14026-14034.
31. Feige, M. J.; Groscurth, S.; Marcinowski, M.; Yew, Z. T.; Truffault, V.; Paci, E.; Kessler, H.; Buchner, J., The structure of a folding intermediate provides insight into differences in immunoglobulin amyloidogenicity. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105* (36), 13373-13378.
32. Toan, N. M.; Morrison, G.; Hyeon, C.; Thirumalai, D., Kinetics of loop formation in polymer chains. *J. Phys. Chem. B* **2008,** *112* (19), 6094-6106.
33. Yang, W. Y.; Gruebele, M., Folding at the speed limit. *Nature* **2003,** *423* (6936), 193-197.
34. Volk, M.; Milanesi, L.; Waltho, J. P.; Hunter, C. A.; Beddard, G. S., The roughness of the protein energy landscape results in anomalous diffusion of the polypeptide backbone. *Phys. Chem. Chem. Phys.* **2014,** *17* (2), 762-782.
35. Krivov, S. V., Is protein folding sub-diffusive? *PLoS Comput. Biol.* **2010,** *6* (9), e1000921.
36. Lim, M.; Jackson, T. A.; Anfinrud, P. A., Nonexponential protein relaxation: dynamics of conformational change in myoglobin. *Proc. Natl. Acad. Sci. U. S. A.* **1993,** *90* (12), 5801-5804.
37. Min, W.; Luo, G. B.; Cherayil, B. J.; Kou, S. C.; Xie, X. S., Observation of a power-law memory kernel for fluctuations within a single protein molecule. *Phys. Rev. Lett.* **2005,** *94* (19), 198302.
38. Hu, X. H.; Hong, L.; Smith, M. D.; Neusius, T.; Cheng, X. L.; Smith, J. C., The dynamics of single protein molecules is non-equilibrium and self-similar over thirteen decades in time. *Nat. Phys.* **2016,** *12* (2), 171-174.
39. Eaton, W. A.; Munoz, V.; Hagen, S. J.; Jas, G. S.; Lapidus, L. J.; Henry, E. R.; Hofrichter, J., Fast kinetics and mechanisms in protein folding. *Annu. Rev. Biophys. Biomol. Struct.* **2000,** *29*, 327-359.

Figures

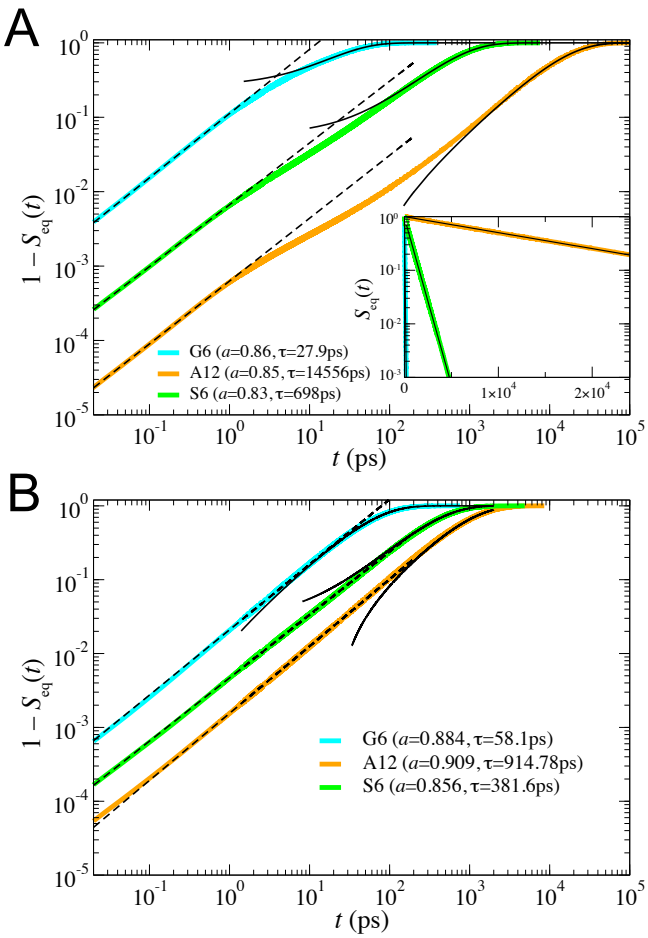

Figure 1. (A) Surviving fraction starting from the open conformation $\boldsymbol{S_{eq}(t)}$ for three homopolypeptides simulated using a united-atom model with implicit solvent. At short times the $\boldsymbol{S_{eq}(t)}$ is not exponential and complex kinetics are observed; at times on the order of tens of ps $\boldsymbol{S_{eq}(t)}$ decays as $\boldsymbol{1 - t^a}$. Dashed lines indicate power law fits at these times, the resulting value of $\boldsymbol{a}$ is indicated in the legend. At intermediate times neither the power law nor the single exponential description fit $\boldsymbol{S_{eq}(t)}$; this is particularly evident for polyalanine. The log-linear plot of $\boldsymbol{S_{eq}(t)}$ in the inset shows that at long times the distribution of survival times of the open conformation decays exponentially. Solid lines in the figure and inset are exponential fits over these longer times; the $\boldsymbol{\tau}$ value in the legend is the fitted decay time. (**B**) Same as A but for the soft-sphere variant of the atomistic model. The $\boldsymbol{S_{eq}(t)}$ behavior for all peptides can be well approximated at short times by $\boldsymbol{1 - S_{eq}(t) \sim t^a}$ and on long timescales by a single exponential fit, without an intermediate phase.

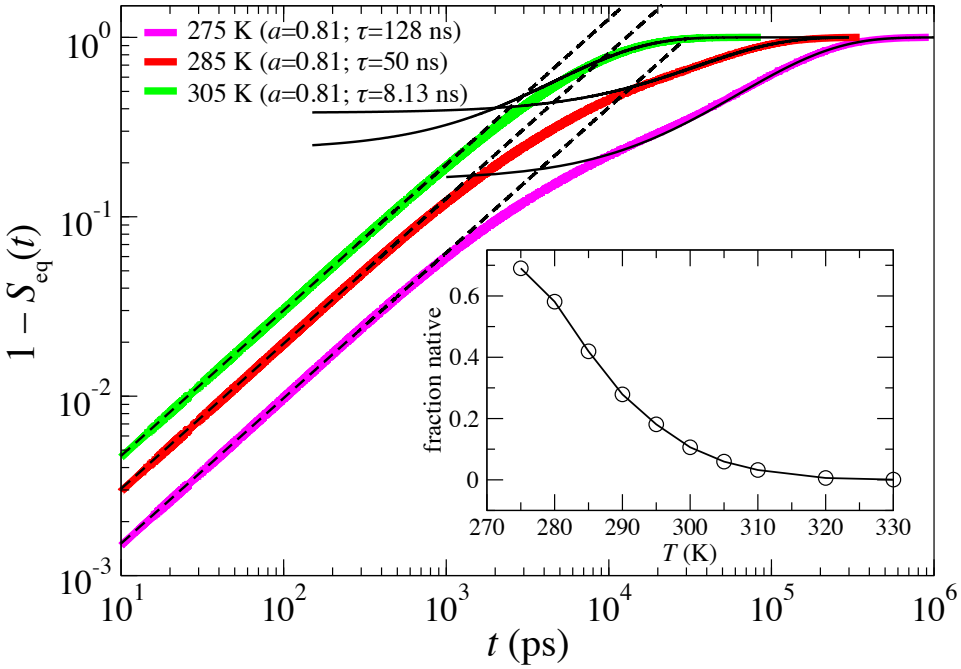

Figure 2. Surviving fraction starting from the open conformation $S_{eq}(t)$ for the structure-based model of the Sac7d protein. Values are shown at three different temperatures. Inset: fraction folded as a function of the temperature. The a values in the legend indicate the exponents of the power law ($1 - S_{eq}(t) \sim t^{a}$) fitted at short times (dashed blue lines). The τ value in the legend is the fitted decay time for exponential fits over longer times (solid blue lines).

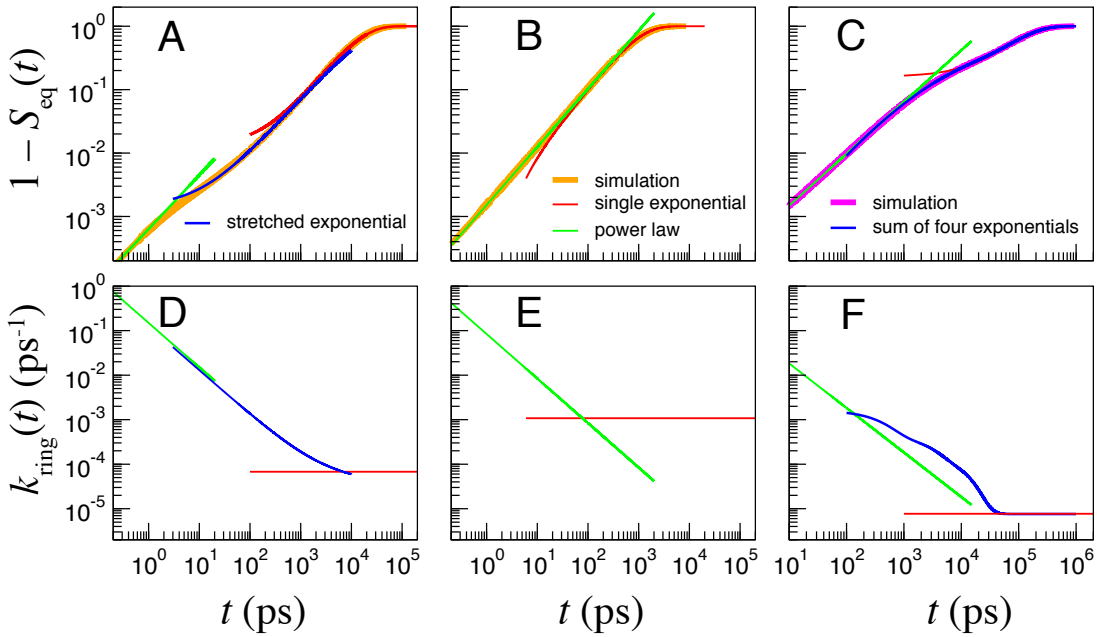

Figure 3. Comparison of the short time kinetics of loop closure for A12 with the united-atom model with implicit solvent (A and D), its soft-sphere variant (B and E) and for the structure-based model of the Sac7d protein at 275 K (C and F). In the first row (A, B and C) is shown the survival times of the open conformation $\mathbf{1} - \boldsymbol{S_{eq}(t)}$. The thick orange (A, B) or black (C) curve is the result from the simulation; the red curve is a single exponential fit, the green curve is a power law fit; the blue lines are empirical fit of the region that is neither power law nor single exponential, a stretched exponential fit for the peptide A12 and a sum of four exponential for the Sac7d protein. In the second row (D, E and F) is shown $\boldsymbol{k_{ring}}$ (the various fits are show with the same color code).

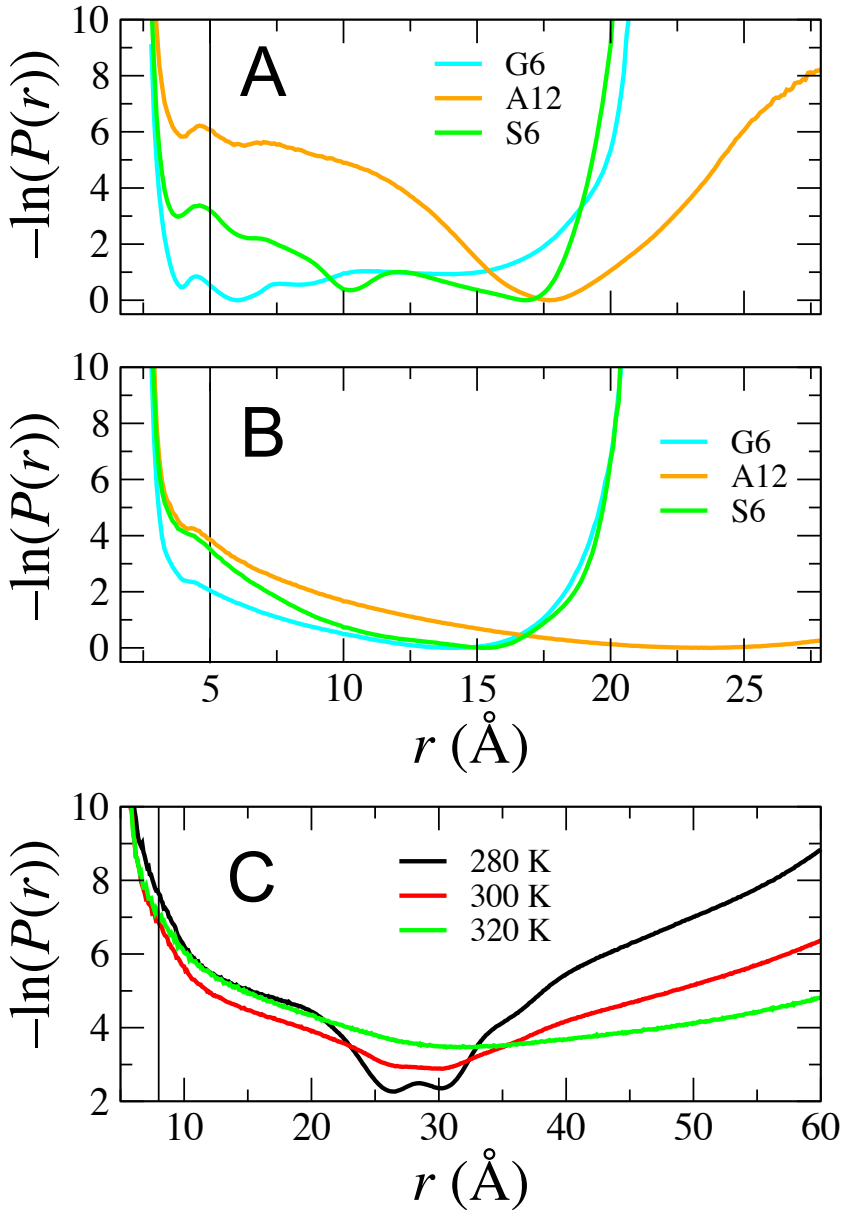

Figure 4. Potential of mean force associated with the end-to-end distance. (A) United-atom model (red curves are A12, green S6, black, G6). (B) Soft-sphere model (red curves are A12, green S6, black, G6). (C) Structure-based model of the protein Sac7d. The vertical lines at 5 Å and 8 Å show the contact radii used when analyzing the atomistic and coarse-grained model, respectively.

Supporting Information for: "Nonexponential Kinetics Of Loop Formation In Proteins And Peptides: A Signature Of Rugged Free Energy Landscapes?"

James Gowdy,† Matthew Batchelor,† Igor Neelov‡ and Emanuele Paci†

† Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, UK

‡ Institute of Macromolecular Compounds of Russian Academy of Sciences, St. Petersburg, Russia

Relationship between equilibrium and ring survival probabilities

Here we prove the relation

$$\frac{S_{ring}\ (t)}{\int_0^\infty S_{ring}(t)dt} = -\frac{dS_{eq}(t)}{dt} \tag{1}$$

A notable related equation has been has been proved by Yeung and Friedman[1]

$$S_{ring}(t)\ \propto -\frac{dS_{eq}(t)}{dt} \tag{2}$$

We start by providing an intuitive demonstration of relation (2). To this end we use a toy dataset obtained from a short stretch of simulation of the A12 peptide. The end-to-end distance r is shown over a 4000 fs window. The contact radius r_c has been set in this example at 5 Å (Figure S1-A). When $r < r_c$ we consider the two ends to be in contact forming a "ring" conformation (Figure S1-B). Time is a discreet variable (*i.e.*, the integration timestep, 1 fs in this example).

We define a ring-to-ring first passage time (FPT) as the time the peptides spend in an open conformation without the end-to-end distance falling below r_c. The first passage is therefore bordered by two ring conformations one upon entering the open state (time entered) and one upon exiting the open state (time exited). An example ring-to-ring first passage is illustrated in Figure S1-D. In the example we enumerate N = 12 contact events (to each of which is assigned a different color).

The frequency distribution for the ring-to-ring FPTs is

$$n_{ring}(t) = \sum_{i=1}^{N} \delta_{t_i,t} \tag{3}$$

i = 1 where $\delta_{i,j}$ is the Kronecker delta function, and plotted in Figure S1-F. For convenience we take t to be a continuum variable in the following,

$$n_{ring}(t) = \sum_{i=1}^{N} \delta(t_i - t)\, dt \tag{4}$$

where $\delta(x)$ is the Dirac delta function with dimension of reciprocal time (thus $n_{ring}(t)$ is dimensionless). The frequency distribution of equilibrium open-to-ring FPTs (Figure S1-E)

$$n_{eq}(t) = \sum_{i=1}^{N} \theta(t_i - t) \tag{5}$$

where $\theta(t)$ is the Heaviside step function, as shown in Figure S1-G.

Thus, using the relationship between Dirac delta and Heaviside function $\int_{-\infty}^{x} \delta(x')dx' = \theta(x)$, the following equation holds between equilibrium and ring frequency distributions:

$$n_{eq}(t) \propto \int_{-\infty}^{t} n_{ring}(t')\, dt'$$

or equivalently,

$$\frac{dn_{eq}(t)}{dt} \propto -n_{ring}(t) \tag{6}$$

Note that proportionality in equation (6) differs from the equality in the equation relating the Dirac delta function to the Heaviside function. This is to account for the definition of n_{ring} in equation (4). Thus the proportionality constant Δ has units of time

$$n_{eq}(t) = \int_{-\infty}^{t} n_{ring}(t')\, dt' \cdot \Delta^{-1}$$

$$\frac{dn_{eq}(t)}{dt} = -n_{ring}(t) \cdot \Delta^{-1}$$

The unnormalized distribution of survival times of the open state, *i.e.* $N_{eq}(t)$ and $N_{ring}(t)$, are related to the respective distributions of FPTs, $n_{eq}(t)$ and $n_{ring}(t)$, *i.e.* the distribution of loop formation times required by the polypeptides to sample conformations up to the point where $r(t) = r_c$. This relationship can be expressed as

$$N(t) = \int_{-\infty}^{t} n(t')dt' \tag{7}$$

for both the ring and equilibrium cases (shown in Figure S1-H and S1-I, respectively). Thus equation (2) can be obtained by integrating both sides of equation (6)

$$N_{ring}(t) \propto -\frac{dN_{eq}(t)}{dt}$$

or with the dimensionally correct equality:

$$N_{ring}(t) = -\frac{dN_{eq}(t)}{dt} \cdot \Delta \tag{8}$$

The normalized counterparts of N_{ring} and N_{eq} are the surviving fraction functions. Assuming the system is ergodic, the surviving fraction is equal to the probability that a polypeptide has not yet undergone loop formation

$$S(t) = \frac{N(t)}{N(0)} = \frac{\int_{-\infty}^{t} n(t')\,dt'}{\int_{-\infty}^{t} n(t')\,dt'} = \Pr(t_c > t) \tag{9}$$

Specifically,

$$S_{ring}(t) = \frac{N_{ring}(t)}{N_{ring}(0)}$$

and

$$S_{eq}(t) = \frac{N_{eq}(t)}{N_{eq}(0)} \tag{10}$$

Substituting equations (9) and (10) into equation (8) results in

$$S_{ring}(t) = -\frac{N_{eq}(0)}{N_{ring}(0)}\frac{dS_{eq}(t)}{dt}\Delta \tag{11}$$

Thus

$$\frac{dS_{eq}(t)}{dt} \propto -S_{ring}(t) \tag{12}$$

The proportionality constant can be shown to be equal to $\int_0^\infty dt S_{ring}(t)$, by considering the equation for the total number of ring-to-ring FPTs

$$N_{ring}(0) = \int_{-\infty}^{0} n_{ring}(t)dt \tag{13}$$

And the total number of equilibrium-to-ring FPTs that occur, which from equation (7) and (8) can be calculated as:

$$N_{eq}(0) = \int_{-\infty}^{t'} \frac{dt'}{\Delta} \int_{-\infty}^{t=0} n_{ring}(t)dt = \frac{1}{\Delta} \int_{-\infty}^{0} t\, n_{ring}(t)dt \tag{14}$$

Therefore

$$\frac{N_{eq}(0)}{N_{ring}(0)} = \frac{1}{\Delta} \frac{\int_{-\infty}^{0} t\, n_{ring}(t)dt}{\int_{-\infty}^{0} n_{ring}(t)dt} = \frac{\tau_{ring}}{\Delta} = \frac{\int_{-\infty}^{0} S_{ring}(t)dt}{\Delta} \tag{15}$$

Where $\tau_{ring}$is the ring-to-ring mean first passage time and is equal to the integral of the respective survival function, *i.e.* $\tau = \int_{-\infty}^{0} S(t)dt$. Thus when equation (15) is substituted into equation (11) we obtain the sought relation given in equation (1).

$$S_{ring}(t) = -\int_{-\infty}^{\infty} S_{ring}(t)dt \frac{dS_{eq}(t)}{dt}$$

or

$$\frac{S_{ring}(t)}{\int_{0}^{\infty} S_{ring}(t)dt} = -\frac{dS_{eq}(t)}{dt} \tag{16}$$

We also define

$$k_{ring}(t) = -\frac{d \ln S_{ring}(t)}{dt} = -\frac{d}{dt} \ln \left[-\frac{dS_{eq}}{dt} \right], \tag{17}$$

which is identical to the instantaneous first-order rate constant k_{rec} or k_{inst} measured by Volk and collaborators.[2-3] And

$$k_{eq}(t) = -\frac{d \ln S_{eq}(t)}{dt} \tag{18}$$

Relations between equilibrium and ring survival probabilities and rates for special functional forms

Single exponential

If

$$S_{eq}(t) = e^{-\alpha t}$$

then

$$S_{ring}(t) = e^{-\alpha t}$$

and

$$k_{ring} = k_{eq} = \alpha$$

Two exponentials

If

$$S_{eq}(t) = ((1-\gamma)e^{-\alpha t} + \gamma e^{-\delta t})$$

then

$$S_{ring}(t) = \frac{\alpha(1-\gamma)(e^{-\alpha t}) + \gamma\delta e^{-\delta t}}{\alpha(1-\gamma) + \gamma\delta}$$

and

$$k_{ring} = -\frac{\alpha^2(1-\gamma)(-e^{-\alpha t}) - \gamma\delta^2 e^{-\delta t}}{\alpha(1-\gamma)e^{-\alpha t} + \gamma\delta e^{-\delta t}}$$

$$k_{eq} = -\frac{\alpha(1-\gamma)(-e^{-\alpha t}) - \gamma\delta e^{-\delta t}}{(1-\gamma)e^{-\alpha t} + \gamma e^{-\delta t}}$$

One exponential and one stretched exponential

If

$$S_{eq}(t) = (1-\gamma)e^{-(\alpha t)^\beta} + \gamma e^{-\delta t}$$

then

$$S_{ring}(t; t \neq 0) = \frac{\alpha\beta(1-\gamma)(e^{-(\alpha t)^\beta})(\alpha t)^{\beta-1} + \gamma\delta e^{-\delta t}}{\alpha\beta(1-\gamma) + \gamma\delta}$$

and

$$k_{ring} = -\frac{-\alpha^2\beta^2(1-\gamma)e^{-(\alpha t)^\beta}(\alpha t)^{2\beta-2} + \alpha^2(\beta-1)\beta(1-\gamma)e^{-(\alpha t)^\beta}(\alpha t)^{\beta-2} - \gamma\delta^2 e^{-\delta t}}{\alpha\beta(1-\gamma)e^{-(\alpha t)^\beta}(\alpha t)^{\beta-1} + \gamma\delta^{-\delta t}}$$

$$k_{eq} = -\frac{\alpha\beta(1-\gamma)(-e^{-(\alpha t)^\beta})(\alpha t)^{\beta-1} - \gamma\delta e^{-\delta t}}{(1-\gamma)e^{-(\alpha t)^\beta} + \gamma e^{-\delta t}}$$

Fit used by Fierz et al.[4] (two exponential and a stretched exponential).

If

$$S_{eq}(t) = (1-\gamma-\epsilon)e^{-(\alpha t)^{\beta}} + \gamma e^{-\delta t} + \epsilon e^{-\zeta t}$$

then

$$S_{ring}(t; t \neq 0) = \frac{\alpha\beta(1-\gamma-\epsilon)(e^{-(\alpha t)^{\beta}})(\alpha t)^{\beta-1} + \gamma\delta e^{-\delta t} + \zeta\epsilon e^{-\zeta t}}{\alpha\beta(1-\gamma-\epsilon) + \gamma\delta + \zeta\epsilon}$$

and

$$k_{ring} = -\frac{-\alpha^2\beta^2(1-\gamma-\epsilon)e^{-(\alpha t)^{\beta}}(\alpha t)^{2\beta-2} + \alpha^2(\beta-1)\beta(1-\gamma-\epsilon)e^{-(\alpha t)^{\beta}}(\alpha t)^{\beta-2} - \gamma\delta^2 e^{-\delta t} - \zeta^2\epsilon e^{-\zeta t}}{\alpha\beta(1-\gamma-\epsilon)e^{-(\alpha t)^{\beta}}(\alpha t)^{\beta-1} + \gamma\delta e^{-\delta t} + \zeta\epsilon e^{-\zeta t}}$$

Power law

If

$$S_{eq}(t) = 1 - \gamma t^{\alpha}$$

then

$$S_{ring}(t; t \neq 0) = \alpha\gamma t^{\alpha-1}$$

and

$$k_{ring} = \frac{1-\alpha}{t}$$

$$k_{eq} = \frac{\alpha\gamma t^{\alpha-1}}{1-\gamma t^{\alpha}}$$

References

1. Yeung, C.; Friedman, B. A., Relation between cyclization of polymers with different initial conditions. *Europhys. Lett.* **2006,** *73* (4), 621-627.
2. Milanesi, L.; Waltho, J. P.; Hunter, C. A.; Shaw, D. J.; Beddard, G. S.; Reid, G. D.; Dev, S.; Volk, M., Measurement of energy landscape roughness of folded and unfolded proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109* (48), 19563-19568.
3. Volk, M.; Kholodenko, Y.; Lu, H. S. M.; Gooding, E. A.; DeGrado, W. F.; Hochstrasser, R. M., Peptide conformational dynamics and vibrational stark effects following photoinitiated disulfide cleavage. *J. Phys. Chem. B* **1997,** *101* (42), 8607-8616.
4. Fierz, B.; Satzger, H.; Root, C.; Gilch, P.; Zinth, W.; Kiefhaber, T., Loop formation in unfolded polypeptide chains on the picoseconds to microseconds time scale. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104* (7), 2163-2168.

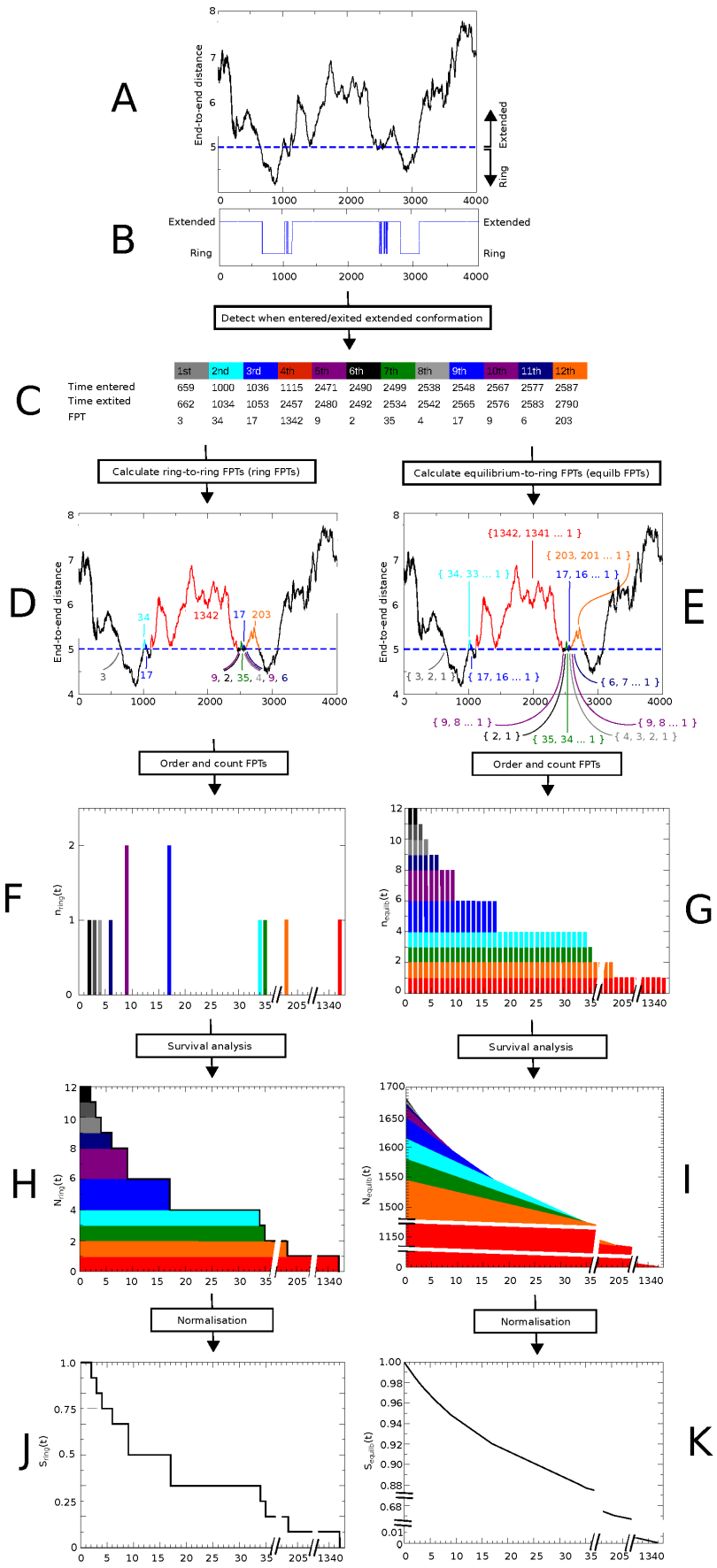
A
End-to-end distance
Extended
Ring
B
Extended
Ring
Detect when entered/exited extended conformation
C
1st 2nd 3rd 4th 5th 6th 7th 8th 9th 10th 11th 12th
Time entered 659 1000 1036 1115 2471 2490 2499 2538 2548 2567 2577 2587
Time extited 662 1034 1053 2457 2480 2492 2534 2542 2565 2576 2583 2790
FPT 3 34 17 1342 9 2 35 4 17 9 6 203
Calculate ring-to-ring FPTs (ring FPTs)
Calculate equilibrium-to-ring FPTs (equilb FPTs)
D
E
Order and count FPTs
Order and count FPTs
F
G
Survival analysis
Survival analysis
H
I
Normalisation
Normalisation
J
K

Figure S2: **Graphical illustration of the relationship between equilibrium and ring survival probabilities. (A)** End-to-end distance of a polypeptide chain during a short trajectory. The contact distance r_c is shown as a dashed line; above this line the peptide is in its open conformation, below in its closed or ring conformation. **(B)** The state of the peptide during the simulation: a total of 12 ring-to-ring events can be enumerated. **(C)** By keeping track of the time points at which r becomes smaller than r_c the ring-to-ring FPT durations can be calculated. These are the set: [3, 34, 17, 1342, 9, 2, 35, 17, 4, 9, 6, 203]. For the illustration of how equilibrium-to-ring FPTs can be calculated from ring-to-ring FPTs and the relationship between $n(t)$, $N(t)$ and $S(t)$ these have been assigned colors depending on their value. **(D)** End-to-end distance with the extended ring-to-ring FPTs indicated. 12 such events are shown each beginning with a ring conformation $r(t) < r_c$. **(E)** Equilibrium-to-ring FPTs can be enumerated from the individual ring-to-ring FPTs. For example the first ring-to-ring FPT has duration of 3 time units (in grey in D) this results in three equilibrium-to-ring FPTs with durations 3, 2 and 1 time-units (shown in grey in (E)). **(F)** Frequency distribution of extended ring-to-ring FPTs, $n_{ring}(t)$. The total integral is equal to the number of ring FPT events. **(G)** Frequency distribution of extended equilibrium-to-ring FPTs, $n_{eq}(t)$. Here the coloring shows the originating ring-to-ring FPTs used to generate the equilibrium-to-ring FPTs. For example the three "grey" FPTs described in above are indicated to contribute to the frequency bars for $n_{eq}(1)$, $n_{eq}(2)$ and $n_{eq}(3)$. **(H)** $N_{ring}(t)$ is the population of surviving extended polypeptide conformations, that have yet to undergo loop closure after a time *t*. $N_{ring}(t)$ is calculated from the $n_{ring}(t)$ using an integral relationship $N_{ring}(t) = \int_t^\infty n_{ring}(t')dt'$. This is also clear from the coloring of the distribution, for example the two ring-to-ring FPT with duration 17 (blue) signifies two conformation that "survive" up to 17 time units, and thus contributes to the $N_{ring}(t)$ for every $t < 17$ time units. However at $N_{ring}(17)$ there is a drop consistent with $-\frac{dN_{ring}(17)}{dt} = n_{ring}(17) = -2$. The form of $N_{ring}(t)$ and $n_{eq}(t)$ are the same, except when calculated from the raw equilibrium-to-ring FPTs the $n_{eq}(t)$ is discrete and $N_{ring}(t)$ is continuous. **(I)** Unnormalized survival function of the equilibrium-to-ring FPTs, $N_{eq}(t)$. This is calculated as $N_{ring}(t)$, except that $n_{eq}(t)$ is used as the input FPT event density function. Thus the corresponding relationship holds, $-\frac{dN_{eq}}{dt} \propto n_{eq}$ and because $n_{eq}(t)$ and $N_{ring}(t)$ are equivalent (see G and H) we see that $-\frac{dN_{eq}}{dt} \propto N_{ring}$. **(J)** Normalized survival functions, $S_{ring}(t)$, *i.e.* the probability that a polypeptide initially in the ring conformation has not yet re-entered the ring state at time *t*. **(K)** Normalized survival functions, $S_{eq}(t)$, *i.e.* the probability that a polypeptide initially in a random equilibrium conformation will not yet have re-entered the ring state at time *t* calculated as $S_{eq}(t) = \frac{N_{eq}(t)}{N_{eq}(0)}$ and related to $S_{ring}(t)$ by the relationship $-\frac{dS_{eq}}{dt} \propto S_{ring}$ in the same manner that $-\frac{dN_{eq}}{dt} \propto N_{ring}$ as described in (I).

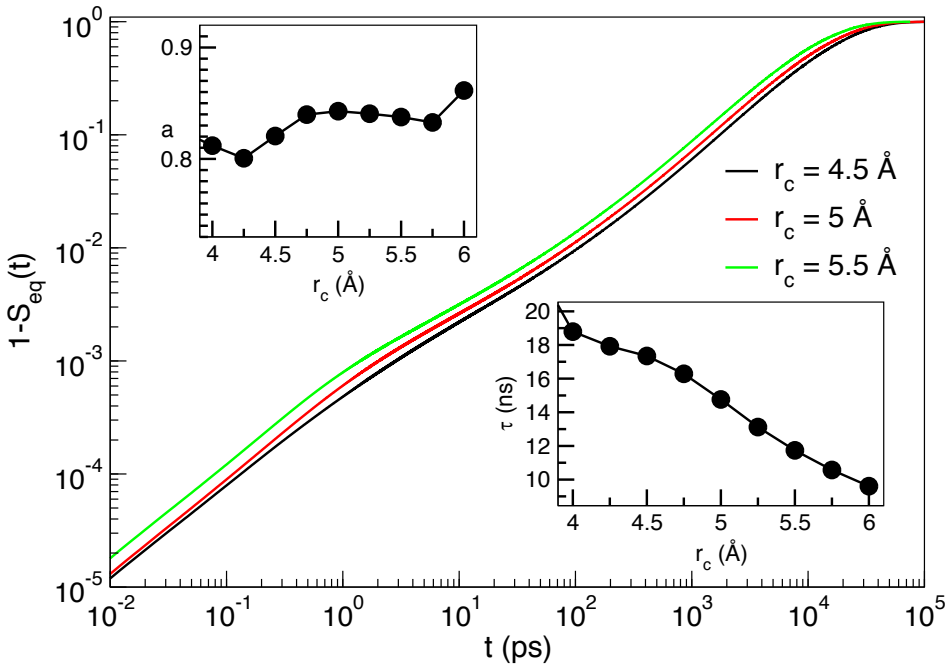

Figure S3. Effect of the contact radius r_c on S_{eq} and fitting parameters for the A12 peptide with the original force field. The parameter 'a' is the exponent of the power law used to fit $1 - S_{eq}$ for $t < 0.3$ ps and τ is the time constant for the single exponential fit of S_{eq} for $t > 2000$ ps. The exponent a does not significantly depend on r_c while τ depends weakly in an expected manner, i.e., decreases with increasing r_c.