J|A|C|S
ARTICLES
Published on Web 05/18/2002

Peptide Loop-Closure Kinetics from Microsecond Molecular Dynamics Simulations in Explicit Solvent

In-Chul Yeh and Gerhard Hummer*

Contribution from the Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520

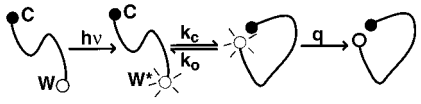
Received February 1, 2002

Abstract: End-to-end contact formation rates of several peptides were recently measured by tryptophan triplet quenching (Lapidus et al. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 7220). Motivated by these experiments, we study loop-closure kinetics for two peptides of different lengths, Cys-(Ala-Gly-Gln)$_n$-Trp (n = 1, 2), in multiple all-atom explicit-solvent molecular dynamics simulations with different initial conditions and force fields. In 150 simulations of approximately 20 ns each, we collect data covering 1.0 and 0.8 μs for the penta-peptide simulated with the AMBER and CHARMM force fields, respectively, and about 0.5 μs each with the two force fields for the octa-peptide. These extensive simulations allow us to analyze the dynamics of peptides in the unfolded state with atomic resolution, thus probing early events in protein folding, and to compare molecular dynamics simulations directly with experiment. The calculated lifetimes of the tryptophan triplet state are in the range of 50−100 ns, in agreement with experimental measurements. However, end-to-end contacts form more rapidly, with characteristic times less than 10 ns. The contact formation rates for the two force fields are similar despite differences in the respective ensembles of peptide conformations.

Introduction

Folded proteins are routinely characterized with increasingly powerful probes of their structural and physicochemical properties. In contrast, comparably little is known about unfolded proteins; yet quantification of protein folding thermodynamics, kinetics, and mechanisms will require an understanding of both sides of the equilibrium. A fundamental process in the dynamics of unfolded peptides is the formation of amino acid contacts. As the folding of an initially disordered polypeptide chain proceeds, contacts between amino acids continuously form and break. With remarkable advances in peptide labeling and time-resolved optical spectroscopy, the contact formation kinetics in biopolymers can now be measured with nanosecond time resolution,[1−10] opening up the possibility to compare theory and simulation[11−17] directly to experiments. By measuring the lifetime of the tryptophan triplet state, Lapidus et al.[7] recently studied the loop-closure kinetics of peptides Cys-(Ala-Gly-Gln)$_n$-Trp (n = 1−6). The tryptophan at one end of the peptide is optically excited into a triplet state that is quenched upon contact formation with the cysteine at the other end, as illustrated in Scheme 1. In the absence of any quencher, the tryptophan triplet state lives for ∼40 μs.[7] At a sufficiently high cysteine quenching rate q relative to the rate of contact breaking k_o, the tryptophan triplet decay rate is an excellent probe for the end-to-end contact formation rate k_c. The observed rates of tryptophan triplet quenching by cysteine are found to be in the range of 1/(100 ns). These are time scales getting within the reach of all-atom molecular dynamics (MD) simulations in explicit solvent.

Scheme 1



Here, we report the results of such simulations to provide an atomistic picture of motions in the unfolded state. For two state-of-the-art force fields, we determine the loop-closure kinetics of two peptides and compare contact formation rates directly

* To whom correspondence should be addressed. E-mail: hummer@helix.nih.gov.

(1) Haas, E.; Katchalski-Katzir, E.; Steinberg, I. Z. *Biopolymers* **1978**, *17*, 11.
(2) Hagen, S. J.; Hofrichter, J.; Szabo, A.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 11615.
(3) Volk, M.; Kholodenko, Y.; Lu, H. S. M.; Gooding, E. A.; DeGrado, W. F.; Hochstrasser, R. M. *J. Phys. Chem. B* **1997**, *101*, 8607.
(4) Bonnet, G.; Krichevsky, O.; Libchaber, A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 8602.
(5) Bieri, O.; Wirz, J.; Hellrung, B.; Schutkowski, M.; Drewello, M.; Kiefhaber, T. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9597.
(6) McGimpsey, W. G.; Chen, L.; Carraway, R.; Samaniego, W. N. *J. Phys. Chem. A* **1999**, *103*, 6082.
(7) Lapidus, L. J.; Eaton, W. A.; Hofrichter, J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 7220.
(8) Wallace, M. I.; Ying, L. M.; Balasubramanian, S.; Klenerman, D. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 5584.
(9) Hagen, S. J.; Carswell, C. W.; Sjolander, E. M. *J. Mol. Biol.* **2001**, *305*, 1161.
(10) Hudgins, R. R.; Huang, F.; Gramlich, G.; Nau, W. M. *J. Am. Chem. Soc.* **2002**, *124*, 556.
(11) Szabo, A.; Schulten, K.; Schulten, Z. *J. Chem. Phys.* **1980**, *72*, 4350.
(12) Zwanzig, R. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 2029.
(13) Zhou, H. X.; Szabo, A. *Biophys. J.* **1996**, *71*, 2440.
(14) Pastor, R. W.; Zwanzig, R.; Szabo, A. *J. Chem. Phys.* **1996**, *105*, 3878.
(15) Thirumalai, D. *J. Phys. Chem. B* **1999**, *103*, 608.
(16) Klenin, K. V.; Langowski, J. *J. Chem. Phys.* **2001**, *114*, 5049.
(17) Dua, A.; Cherayil, B. J. *J. Chem. Phys.* **2002**, *116*, 399.

with experiments. This is achieved by performing many ~20 ns simulations with different initial conditions, instead of one long run. The simulations with two different force fields cover 1.0 and 0.8 μs, respectively, for a penta-peptide, and about 0.5 μs each for an octa-peptide. Running multiple independent simulations results in trivial and thus ideal parallelization on low bandwidth computer clusters, and allows us to explore the configuration space of peptides efficiently. For exponential, that is, memoryless processes, the two approaches are practically identical.

Methods

The penta- and octa-peptides Cys-(Ala-Gly-Gln)$_n$-Trp (n = 1 and 2) are simulated in explicit aqueous solution at 300 K to calculate their end-to-end contact formation rates. To quantify the effects of force field approximations, we use two all-atom potential parameter sets for biomolecular simulations: AMBER 94[18] and CHARMM 22.[19] We perform multiple independent simulation runs for each parameter set with total accumulated times of 1 and 0.8 μs, respectively, for the short peptide (n = 1) and about 0.5 μs each for the long (n = 2) peptide. The Cys-Ala-Gly-Gln-Trp (CAGQW) peptide is solvated in 526 TIP3P water molecules.[20] As in the experiment,[7] the peptide is amidated at the C-terminus without capping at the N-terminus. One sodium (Na^+) ion and two chloride (Cl^-) ions are included in the system, resulting in a net-neutral system. Periodic boundary conditions are applied to the cubic simulation box with particle-mesh-Ewald electrostatics[21,22] on a 32 × 32 × 32 grid. In both the AMBER and the CHARMM simulations, we use a time step of 2 fs. The SHAKE algorithm is used to constrain bonds involving hydrogen atoms.[23] The Berendsen thermostat and barostat[24] are used for temperature and pressure control. Every picosecond, coordinates are saved for analysis. In averaging equilibrium properties, all saved conformations are given equal weight.

For the CAGQW runs with the AMBER 94 force field,[18] we use the SANDER module of the AMBER 6.0 program.[25] The 1−4 interactions are consistently scaled by a factor of (1/1.2). A 200 ps MD simulation at 300 K temperature and 1 bar pressure is used to estimate the box size of subsequent production runs at constant volume. A 1 ns constant-volume simulation at an elevated temperature of 700 K is then used to rapidly sample the configuration space. Fifty configurations from the 700 K run are used as initial configurations for subsequent production runs. Initial velocities are randomly drawn from Maxwell−Boltzmann distributions at 300 K. After 100 ps of equilibration, all 50 production runs last slightly more than 20 ns, resulting in a total accumulated production time of more than 1 μs. For the CAGQW runs with the CHARMM 22 force field,[19] we use the CHARMM program version 27b3.[26] Box size and initial configurations for 50 production runs are prepared in a similar manner as for AMBER runs. Excluding initial equilibration periods of 100 ps, 34 production runs cover about 20.6 ns, while the remaining 16 runs range from 4.6 ns up to 20 ns, resulting in a total accumulated time of 0.854 μs. We also carry out simulations for a longer peptide Cys-(Ala-Gly-Gln)$_2$-Trp [C(AGQ)$_2$W]. Here 1064 TIP3P water molecules with one Na^+ and two Cl^- counterions are used as solvent. Twenty-five simulations are performed for each force field. Most of the runs cover 21 ns with the AMBER force field and 18 ns with the CHARMM force field, resulting in total simulation times of 0.532 μs for AMBER and 0.445 μs for CHARMM. For simulations of C(AGQ)$_2$W with the CHARMM force field, we use the program NAMD.[27]

The end-to-end distance r_e of peptides C(AGQ)$_n$W (n = 1, 2) is defined as the distance between the sulfur atom of the cysteine side chain and the closest non-hydrogen atom of the tryptophan indole ring. The distribution and time dependence of r_e in each run are analyzed to estimate the end-to-end contact formation rate. For all saved conformations of each run, end-to-end distances are calculated and binned into histograms. We assume that the potential surface of the tryptophan triplet state is unchanged from that of the ground state. At a lower temperature of 293 K, the shear viscosity of a modified TIP3P water model[29] included in the CHARMM force field is smaller than the experimental value by a factor of 2.86.[30] We also calculate the shear viscosity of the original TIP3P water[20] at a temperature of 298 K and a number density of 33 water molecules per nm^3, corresponding to a pressure of 9 ± 2 bar. A 10.1 ns simulation of 2048 TIP3P water molecules analyzed by Green−Kubo relations gives an estimated shear viscosity of (3.1 ± 0.1) × 10^{-4} kg m^{-1} s^{-1}, smaller than the experimental value by a factor of 2.87. To compare the contact formation rates with experiment, we divide the calculated rates by 2.87, assuming a linear solvent viscosity dependence and neglecting "internal friction".[31−33]

Results and Discussion

End-to-End Distance Distributions. Distributions $P(r_e)$ of the end-to-end distance r_e from CHARMM and AMBER runs of the peptides CAGQW and C(AGQ)$_2$W are shown in Figure 1. For both force fields, we find a small peak near r_e = 4 Å, corresponding to an ensemble of contact configurations of the N- and C-terminal side chains. Figure 1 also shows that the CHARMM simulations have extended configurations more populated than AMBER. This is consistent with the observation by Samuelson et al.[34] that the AMBER 94 force field[18] favors helical backbone conformations, whereas the CHARMM force field favors extended ones in simulations of a blocked tri-alanine peptide in a vacuum and water.

Contact Formation Kinetics. Consistent with the experimental analysis of Lapidus et al.,[7] we use a contact distance of d_c = 4 Å, which corresponds to the approximate position of the first peak in Figure 1. On the basis of this definition, we estimate $S(t)$, the survival probability of the tryptophan triplet state between time 0 and t. First, we consider diffusion-limited quenching, corresponding to an infinitely large quenching rate within the contact distance and no quenching beyond. The triplet state of the tryptophan at the C-terminus survives until it first

(18) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
(19) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.
(20) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(21) Darden, T. A.; York, D. M.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *98*, 10089.
(22) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.
(23) Ryckaert, J.; Ciccotti, G.; Berendsen, H. *J. Comput. Phys.* **1977**, *23*, 327.
(24) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
(25) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E., III; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1.
(26) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(27) Kale, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. *J. Comput. Phys.* **1999**, *151*, 283.
(28) Koradi, R.; Billeter, M.; Wüthrich, K. *J. Mol. Graphics* **1996**, *14*, 51.
(29) Durell, S. R.; Brooks, B. R.; Ben-Naim, A. *J. Phys. Chem.* **1994**, *98*, 2198.
(30) Feller, S. E.; Pastor, R. W.; Rojnuckarin, A.; Bogusz, S.; Brooks, B. R. *J. Phys. Chem.* **1996**, *100*, 17011.
(31) Ansari, A.; Jones, C. M.; Henry, E. R.; Hofrichter, J.; Eaton, W. A. *Science* **1992**, *256*, 1796.
(32) Klimov, D. K.; Thirumalai, D. *Phys. Rev. Lett.* **1997**, *79*, 317.
(33) Portman, J. J.; Takada, S.; Wolynes, P. G. *J. Chem. Phys.* **2001**, *114*, 5082.
(34) Samuelson, S.; Tobias, D. J.; Martyna, G. J. *J. Phys. Chem. B* **1997**, *101*, 7592.

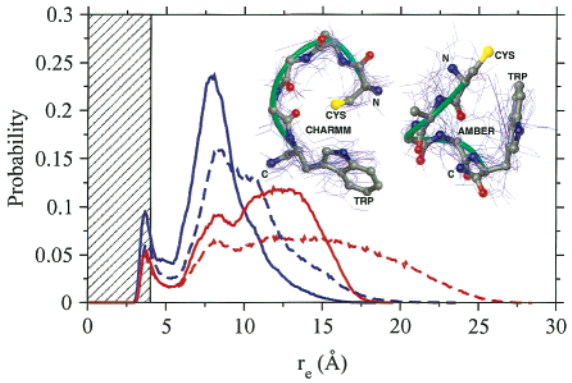


Figure 1. Probability distributions $P(r_e)$ of the end-to-end distance r_e for peptides CAGQW (solid lines) and C(AGQ)$_2$W (dashed lines) for AMBER (blue) and CHARMM (red). The shaded region indicates end-to-end contacts ($r_e < 4$ Å). The inset shows representative structures of the CHARMM (left) and AMBER (right) contact ensembles of the CAGQW peptide (figure prepared with the program MOLMOL[28]). The green tube traces the α-carbon backbone. Fifteen randomly selected structures are shown with thin blue lines to indicate variations in the contact ensembles.

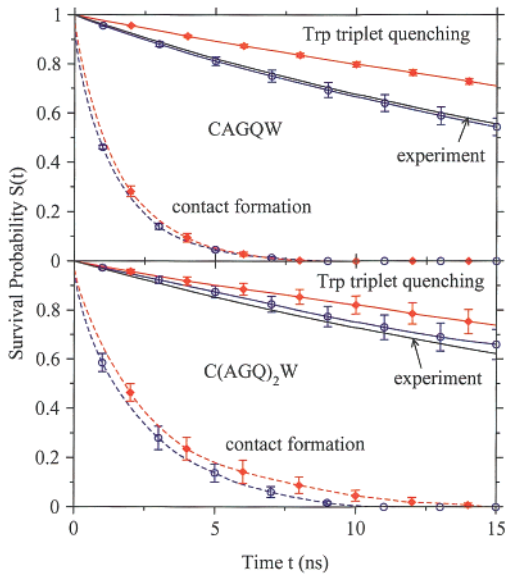


Figure 2. Contact formation and tryptophan triplet quenching kinetics. Survival probabilities $S(t)$ of the tryptophan triplet state are shown for peptides CAGQW (upper panel) and C(AGQ)$_2$W (lower panel) calculated from MD simulations using AMBER (blue; ○) and CHARMM force fields (red; ◆). Dashed lines are $S(t)$'s for diffusion-limited quenching corresponding to the kinetics of actual contact formation. The tryptophan triplet quenching kinetics are calculated with a cysteine quenching rate of $q = 0.8$ ns^{-1} and shown as solid lines. $S(t)$'s calculated with the experimental decay rate, scaled by a viscosity factor of 2.87, are shown as solid black lines.

encounters the cysteine at the N-terminus. The survival probability is then $S(t) = 1$ for $t < t_c$, where t_c is the first-contact time, and $S(t) = 0$ for $t \geq t_c$. We assume that every saved noncontact configuration ($r_e > d_c$) for each MD simulation run is a valid starting point. $S(t)$'s averaged over all starting points of the respective simulations are shown in Figure 2. By integration,[11]

$$k_c^{-1} = \int_0^\infty S(t)\, dt \qquad (1)$$

we find contact formation rates k_c (k_{D+} in ref 7) for the CAGQW peptide of 1/(1.4 ± 0.1) and 1/(1.6 ± 0.1 ns) for AMBER and CHARMM, respectively, corresponding to diffusion-limited quenching. Errors correspond to one standard deviation of the mean, and are estimated by subdividing production runs into multiple groups assumed to be independent. Essentially the same contact rates are obtained if 10 ns instead of 100 ps at the beginning of each trajectory are treated as equilibration. Interestingly, the contact formation rates from simulations with two different force fields, AMBER and CHARMM, agree within their combined errors despite the apparent discrepancy in the end-to-end distance distributions (Figure 1). An explanation for this is that the equilibration in the open state ($r_e \gg d_c$) is rapid as compared to contact formation, which faces approximately the same barrier in the two force fields, as estimated roughly from the potential of mean force along the end-to-end distance coordinate,[11] $-k_BT \ln[P(r_e)/r_e^2]$. After correcting the MD results for solvent viscosity, we thus consistently estimate contact formation times of about 5 ns for the penta-peptide CAGQW.

Tryptophan Triplet Quenching Kinetics. The diffusion-limited rates of contact formation are significantly faster than the experimentally measured decay rate of about 1/(73 ns) of the tryptophan triplet state, even after considering the almost 3-fold lower viscosity of the MD solvent. This large discrepancy between experimental and simulation rates suggests that the cysteine-induced quenching of the tryptophan triplet state may not be diffusion-limited. This was also concluded from measurements using lipoate instead of cysteine as a quencher of tryptophan triplet states,[7] and from experiments with 2,3-diazabicyclo[2.2.2]oct-2-ene (DBO) as a fluorophore and tryptophan as a quencher attached to peptides of comparable length.[10] To account for slower quenching rates, we write the survival probability as a path integral[13]

$$S(t) = \langle \exp\{-\int_0^t \kappa[r_e(t')]\, dt'\} \rangle \qquad (2)$$

where $\kappa[r_e(t)]$ is the quenching rate by cysteine, assumed to depend only on the end-to-end distance (r_e) which is itself a function of time t, and $\langle ... \rangle$ signifies an average over the equilibrium ensemble of starting configurations. As in the experimental analysis,[7] we assume that the contact pair is only reactive if $r_e < d_c$:

$$\kappa(t) = \kappa[r_e(t)] = \begin{cases} q \text{ if } r_e(t) < d_c \\ 0 \text{ if } r_e(t) \geq d_c \end{cases} \qquad (3)$$

Lapidus et al.[7] estimated the quenching rate q at contact to be 0.8 ns^{-1} for cysteine, with a contact distance of $d_c = 4$ Å. With this rate, we calculate $S(t)$ by using eqs 2 and 3. In averaging over multiple runs, we assume that every saved configuration of each MD simulation run is a valid starting point and that the tryptophan is in the triplet state at the starting point. The calculated survival probabilities $S(t)$ show single-exponential decay, as shown in Figure 2, with tryptophan triplet decay rates of $\tau^{-1} = 1/(24.6 \pm 2$ ns) and 1/(44.4 ± 2 ns) for AMBER and CHARMM runs, respectively. If we divide these values by the viscosity factor 2.87 (the ratio of experimental and calculated viscosity coefficients), they become 1/(70.6 ± 6 ns) and 1/(127 ± 6 ns) for AMBER and CHARMM, respectively. This assumes a linear viscosity dependence of all processes relevant for tryptophan triplet quenching. Considering that d_c, $\kappa[r_e(t)]$, and the viscosity effects are not precisely defined, these values are in agreement with the experimentally measured decay rate of 1/(73 ns). This suggests that triplet quenching of CAGQW is

mostly reaction-controlled. We can thus estimate the observed lifetime of the tryptophan triplet state as the sum of the equilibrium lifetime τ_{eq} and a diffusion lifetime τ_d:[35,36]

$$\tau_{app} \approx \tau_{eq} + \tau_d \tag{4}$$

τ_{eq} is the reaction-controlled lifetime (i.e., the lifetime in the limit of fast diffusion):

$$\tau_{eq}^{-1} = k_{eq} = \int_0^\infty \kappa(r_e) P(r_e)\, dr_e \tag{5}$$

τ_d is defined by the autocorrelation function of the distance-dependent rate

$$\tau_d = \int_0^\infty \frac{\langle \delta\kappa(t)\delta\kappa(0)\rangle}{k_{eq}^2} dt \tag{6}$$

where $\delta\kappa(t) = \kappa[r_e(t)] - k_{eq}$. Calculated lifetimes τ_{app}, τ_{eq}, and τ_d are 26.3, 20.3, and 6.0 ns, respectively, for AMBER, and 43.9, 38.4, and 5.5 ns for CHARMM. The lifetimes τ_{app} from eq 4 agree well with the decay times of $S(t)$ from eq 2: $\tau_{app} =$ 26.3 ns versus $\tau = 24.6$ ns for AMBER, and $\tau_{app} = 43.9$ ns versus $\tau = 44.4$ ns for CHARMM. The lifetime τ_{app} is dominated by the contribution from τ_{eq} in both AMBER and CHARMM runs, and thus determined largely by the end-to-end distance distributions $P(r_e)$. About $6.2 \pm 0.6\%$ and $3.3 \pm 0.4\%$ of configurations from AMBER and CHARMM runs, respectively, have their end-to-end distances within the contact distance, $d_c = 4$ Å. This difference in contact populations largely accounts for the difference in the tryptophan triplet decay rates for the two force fields.

Lifetime of Contacts. These results suggest that triplet quenching of CAGQW is mostly reaction-controlled. This is a consequence of the relatively fast decay of contact states to open, noncontact states, with a contact breaking rate k_o on the order of 10 ns^{-1}, as compared to a quenching rate of $q = 0.8$ ns^{-1}. In a simplified kinetic model corresponding to Scheme 1

$$O \underset{k_o}{\overset{k_c}{\rightleftharpoons}} C \xrightarrow{q} Q \tag{7}$$

C, O, and Q signify unquenched contact states ($r_e < d_c$), open noncontact states ($r_e > d_c$), and quenched states, respectively. For equilibrium initial conditions [$C(t = 0) = k_c/(k_o + k_c)$, $O(0) = k_o/(k_o + k_c)$, and $Q(0) = 0$], the lifetime τ of unquenched states is given by

$$\tau = \int_0^\infty [1 - Q(t)]\, dt = \frac{(k_o + k_c)^2 + k_o q}{k_c(k_o + k_c)q} \tag{8}$$

Note that for infinitely fast quenching ($q \to \infty$), τ becomes equal to the equilibrium fraction of noncontact states, $k_o/(k_o + k_c)$, divided by the rate of contact formation, k_c. With diffusion-limited rates k_c of contact formation from eq 1, decay times τ of $S(t)$ from eq 2, and the quenching rate q of the experimental analysis,[7] we estimate contact breaking rates of $k_o \approx 1/(80$ ps) (AMBER) and 1/(50 ps) (CHARMM), reasonably reproducing the observed equilibrium populations of contact states [$k_c/(k_o + k_c) \approx 0.054$ versus 0.062 ± 0.006 observed in the AMBER simulations, and 0.029 versus 0.033 ± 0.004 for CHARMM]. The kinetic scheme eq 8 also provides a framework for a future, more detailed analysis of the solvent viscosity dependence, with the possibility that the three rates scale differently with solvent viscosity.

Peptide Length Dependence. As the length of the polypeptide chain increases, different configurational preferences of individual amino acids in force fields become amplified. This should be reflected in the end-to-end distance distribution. To study the effects of chain length, we analyze trajectories of $C(AGQ)_2W$. The end-to-end distance distributions (Figure 1) again reflect a preference for extended conformations in CHARMM as compared to AMBER. The contact formation rates are estimated to be $1/(2.25 \pm 0.3$ ns) and $1/(2.85 \pm 0.4$ ns) for AMBER and CHARMM, respectively, corresponding to diffusion-limited quenching of tryptophan triplet states. These values are similar to each other, and about 2 times slower than those for the shorter peptide. Corrected for the solvent viscosity, we thus estimate contact formation times of about 6−8 ns for the octa-peptide $C(AGQ)_2W$.

Lifetimes of tryptophan triplet states calculated with the same quenching rate q as used for CAGQW are 36 ± 10 and 50 ± 9 ns for AMBER and CHARMM, respectively. If we correct these values by the viscosity factor 2.87, they become 103 ± 29 and 143 ± 26 ns for AMBER and CHARMM, respectively. The experimental value for the lifetime is about 90 ns, in good agreement with the MD simulations. The rate of contact breaking k_o in the longer peptide, $C(AGQ)_2W$, is comparable to that of the shorter peptide, and significantly faster than the cysteine quenching rate q. On the basis of the kinetic model, eq 7, we estimate $k_o = 1/(90$ ps) for AMBER and 1/(80 ps) for CHARMM, again reproducing the populations of contact states (0.037 versus 0.037 ± 0.005 observed for AMBER, and 0.027 versus 0.030 ± 0.005 observed for CHARMM). Interestingly, the CHARMM simulations show only a small decrease in the relative fraction of contact states from 0.033 ± 0.004 to 0.030 ± 0.005 as the peptide length increases. A possible explanation is that chain stiffness leads to a reduction in the population of contact configurations for short peptides.[37] A nonmonotonic dependence of the contact formation rate was indeed observed for short peptides by Hudgins et al.[10] The chain-stiffness effect is expected to be more pronounced for the CHARMM force field which results in relatively more extended conformations than AMBER (Figure 1).

Distance Dependence of Quenching Rate. Tryptophan triplet decay rates depend not only on the formation of end-to-end contacts, but also on details of the cysteine quenching mechanism. To compare simulations with experiments, we invoked the simple model eq 3 used before in the interpretation of the measurements.[7] Lapidus et al.[38] recently introduced a refined distance-dependent quenching rate, $\kappa(r_e) = \kappa_0 \exp[-\beta(r_e - r_0)]$. This model was fitted to triplet-state quenching data measured for a room-temperature glass containing tryptophan and cysteine at low and high concentrations, respectively. In this "frozen solution," tryptophan and cysteine are effectively immobilized. The quenching rate $\kappa(r_e)$ can thus be probed directly, given the positional distribution of tryptophan

(35) Wilemski, G.; Fixman, M. *J. Chem. Phys.* **1974**, *60*, 866.
(36) Bicout, D. J.; Szabo, A. *J. Chem. Phys.* **1997**, *106*, 10292.
(37) Camacho, C. J.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 1277.
(38) Lapidus, L. J.; Eaton, W. A.; Hofrichter, J. *Phys. Rev. Lett.* **2001**, *87*, 258101.

and cysteine. Under the simplifying assumption of a uniform distribution outside an exclusion radius of $r_0 = 4$ Å, the experimental triplet decay could be fitted with $\kappa_0 = 0.52$ ns^{-1} and $\beta = 3.23$ Å^{-1} if the solution was assumed to be completely frozen, and $\kappa_0 = 4.2$ ns^{-1} and $\beta = 4.0$ Å^{-1} if diffusion was allowed.[38] In adapting this model for use with simulation data, we assume that $\kappa(r_e < r_0) = \kappa(r_0) = \kappa_0$ to account for the observed population of end-to-end distances below 4 Å (Figure 1). We obtained lifetimes of 28 ns (AMBER) and 50 ns (CHARMM) for CAGQW, and 39 ns (AMBER) and 57 ns (CHARMM) for C(AGQ)$_2$W with $\kappa_0 = 0.52$ ns^{-1} and $\beta = 3.23$ Å^{-1}, in good agreement with the values calculated with the simpler step-function model, eq 3. With $\kappa_0 = 4.2$ ns^{-1} and $\beta = 4.0$ Å^{-1}, we obtained shorter lifetimes of 6 ns (AMBER) and 9 ns (CHARMM) for CAGQW, and 9 ns (AMBER) and 12 ns (CHARMM) for C(AGQ)$_2$W. This highlights a sensitivity to the quenching model, and the need either to refine the distance-dependent quenching rates, possibly by using a more realistic tryptophan-cysteine distance distribution in the analysis of the glass solution data or to extrapolate the experimental rates into the diffusion-limited regime.

Probe-Quencher Pair Dependence. Faster quenchers help avoid the difficulties associated with modeling quenching mechanisms in the interpretation of the experiments. Diffusion-limited quenchers would probe the contact formation rate directly. In a recent study, tryptophan at one end of a peptide was used as a quencher of DBO at the other end.[10] The DBO/Trp pair has a 4 times higher quenching rate as compared to that of the Trp/Cys pair, bringing DBO/Trp close to diffusion-limited quenching. Fluorescence quenching of DBO also requires the formation of a van der Waals contact with the quencher, and can thus be used to probe contact formation rates. The experiments by Hudgins et al.[10] show decay times of DBO fluorescence of 10−30 ns for peptides with 2−10 peptide units, that is, with lengths comparable to those studied here. To compare our results for the CAGQW and C(AGQ)$_2$W peptides with the peptides Trp-(Gly-Ser)$_m$-DBO-NH$_2$ studied by Hudgins et al.,[10] we increase the cysteine quenching rate q in our model eq 3 by a factor of 4 to account for the increase in quenching rate constants for the DBO/Trp pair relative to the Trp/Cys pair.[10] If we correct, in addition, for the 25% higher viscosity of D_2O as compared to H_2O, we estimate fluorescence decay rates of 30 μs^{-1} (AMBER) and 18 μs^{-1} (CHARMM) for CAGQW, and 21 μs^{-1} (AMBER) and 16 μs^{-1} (CHARMM) for C(AGQ)$_2$W. This is somewhat slower than the measured values of 68, 49, and 31 μs^{-1} for $m = 1$, 2, and 4 of Trp-(Gly-Ser)$_m$-DBO-NH$_2$, but in reasonable agreement considering the differences in peptide sequence. From these experiments and our simulations, we thus conclude that end-to-end contact formation can occur on time scales of 10 ns or less.

End-to-End Diffusion. Following Haas et al.,[1] we can estimate effective diffusion coefficients for the end-to-end motion of peptides. We assume a potential of mean force for the end-to-end distance, $W(r_e) = -k_BT \ln[P(r_e)/r_e^2]$, and solve Smoluchowski's diffusion equation[39] in polar coordinates with an absorbing boundary on a sphere with radius $r_e = d_c = 4$ Å. This assumes that r_e is a suitable reaction coordinate for contact formation and that the diffusion coefficient for end-to-end motion is independent of position, which are both approximations. By varying the effective diffusion coefficient for end-to-end motion, we can match the contact formation rates of the Smoluchowski diffusion model with those of the actual simulation data. This gives effective diffusion coefficients of 3.25×10^{-6} cm^2 s^{-1} (AMBER) and 8.9×10^{-6} cm^2 s^{-1} (CHARMM) for the CAGQW peptide. If we assume that the diffusion coefficients scale inversely with the solvent viscosity, we obtain diffusion coefficients of 1.1×10^{-6} cm^2 s^{-1} (AMBER) and 3.1×10^{-6} cm^2 s^{-1} (CHARMM) that are larger by factors of 2 (AMBER) and 7 (CHARMM) as compared to those reported by Haas et al.[1] from fluorescence-decay measurements for a peptide with five N^5-(2-hydroxyethyl)-L-glutamine repeating units. Considering the differences in the chemical composition, in particular the absence of a flexible glycine residue in the peptide, and the assumptions in the analysis, the agreement is reasonable. In a recent study of the effects of chain stiffness on loop formation of the C(AGQ)$_n$W peptides, Lapidus et al.[40] use a wormlike-chain model for the peptide to estimate end-to-end diffusion coefficients. The experimental measurements are consistent with a linear viscosity dependence of the end-to-end diffusion coefficient. At the experimental viscosity of water, Lapidus et al.[40] estimate an end-to-end diffusion coefficient of 2×10^{-6} cm^2 s^{-1}, a value bracketed by our estimates for the AMBER and CHARMM force fields after correction for the low viscosity of the TIP3P water model.

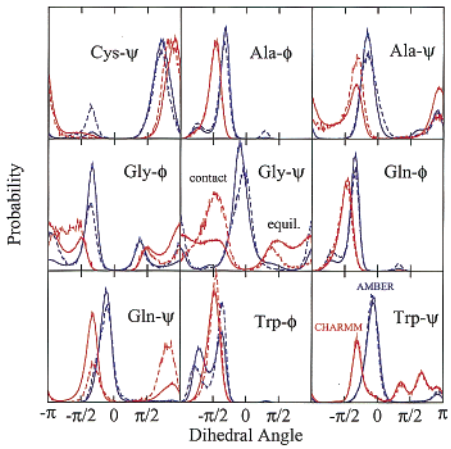


Figure 3. Distributions of backbone dihedral angles of the CAGQW peptide from AMBER (blue) and CHARMM (red). Solid and dashed lines represent distributions from the equilibrium and contact ensembles ($r_e < 4$ Å), respectively.

Peptide Conformations. To examine structural changes associated with the end-to-end contact formation of the peptide CAGQW, we show in Figure 3 distributions of peptide-backbone dihedral angles from configurations with end-to-end contacts and compare them with the corresponding distributions in the equilibrium ensemble. Overall, backbone dihedral angle distributions from AMBER runs show prominent peaks at positions corresponding to α helical structures, while those from CHARMM runs show a significant population of extended conformations, which is consistent with distributions of r_e in Figure 1. Cys-ψ, Gly-ϕ, Gly-ψ, and Trp-ϕ backbone dihedral angle distributions from AMBER runs and Ala-ψ, Gly-ϕ, Gly-ψ, and Gln-ψ distributions from CHARMM runs change significantly upon end-to-end contact formation, with the

(39) Bicout, D. J.; Szabo, A. *J. Chem. Phys.* **1998**, *109*, 2325.

(40) Lapidus, L. J.; Steinbach, P. J.; Eaton, W. A.; Szabo, A.; Hofrichter, J. *J. Phys. Chem. B* **2002**, submitted.

backbone dihedral angles of other residues less affected by the constraint of an end-to-end contact. Not unexpectedly, the backbone dihedral angles of the central Gly residue show the most pronounced effects upon end-to-end contact formation in both AMBER and CHARMM runs. However, the backbone dihedral angle distributions calculated from the two force fields differ significantly. For instance, the distribution of the Gly-ψ angle from AMBER is peaked near $\psi = 0$ where CHARMM shows a minimum. It is clear from the comparison of dihedral angle distributions from AMBER and CHARMM runs that there is little similarity between AMBER and CHARMM configurations even in the restricted ensembles of peptides with end-to-end contacts. This observation is confirmed by principal-component axes and mean-square distance analyses[41,42] on randomly selected samples of structures. The inset in Figure 1 shows representative structures of the contact ensembles ($r_e < d_c$) from the AMBER and CHARMM simulations; 55% and 30% of the CHARMM and AMBER contact structures are within 2 Å root-mean-square distance of their respective reference structure, calculated for all heavy atoms, including side chains. In contrast, only about 2% are within the same distance of the reference structure of the other force field, highlighting the small overlap of the two contact ensembles.

Conclusions

From 150 MD simulations with different initial conditions, each covering about 20 ns, with total times between about 0.5 and 1 μs per peptide and force field, we are able to extract a detailed description of the motions in the unfolded state of two short peptides. We find that characteristic times for peptide end-to-end contact formation are remarkably short, less than 10 ns for the penta- and octa-peptides studied here. A comparison of the calculated contact formation rates with experimental measurements[7] suggests that the measured rate of tryptophan triplet quenching of the peptides $C(AGQ)_nW$ ($n = 1, 2$) is not diffusion-limited, but reaction-controlled. This follows from the short lifetime of contacts, with the rate of contact breaking for both peptides being on the order of 10 ns^{-1}, more than 1 order of magnitude larger than the estimated cysteine quenching rate of $q = 0.8$ ns^{-1}.

Our MD simulation data based on AMBER and CHARMM force fields predict similar values for the rate of end-to-end contact formation, despite substantial structural differences both in the respective equilibrium ensembles and in configurations forming end-to-end contacts. Moreover, estimates of the decay rate of tryptophan triplet states from MD simulations agree well with experimentally measured decay rates, based on a contact distance of 4 Å and a cysteine quenching rate of $q = 0.8$ ns^{-1}, as used before in the interpretation of the experimental data,[7] and the assumption of linear scaling with viscosity. The insensitivity of the contact formation kinetics to conformational preferences can be explained if the equilibration in the open state is rapid as compared to contact formation, and the free-energy barrier to contact formation is similar for the two force fields, as is indeed the case along the end-to-end distance coordinate.

Unlike the contact formation rate, the calculated tryptophan-quenching kinetics depend sensitively on the model used for the distance dependence of the cysteine quenching rate. Using faster quenchers helps avoid model dependences. Recent experiments by Hudgins et al.[10] for peptides of similar lengths but with faster quenchers indeed show decay rates of 1/(10) to 1/(20 ns), suggesting contact formation rates near 1/(10 ns) or faster, in good correspondence with our simulation data for different peptide sequences.

Differences in the ensembles of coil structures between the two force fields suggest that conformational properties of the unfolded states of proteins and peptides are sensitive to small perturbations in the interactions. This can be rationalized within an energy landscape framework. The unfolded ensemble consists of structures moving rapidly between a large number of shallow (~ 1 k_BT) minima on the free energy surface, and small perturbations (~ 1 k_BT) of the free energy surface lead to significant changes in the population of individual minima. In contrast, the folded state of a protein occupies a deep free energy minimum ($\gg 1$ k_BT), rendering it relatively insensitive to small perturbations, such that force fields with different parametrizations can perform comparably well for proteins in their native state. The observation of a sensitivity of the unfolded state to perturbations, such as changes in force-field parametrizations and possibly solvent conditions, is relevant in particular for future work on "ab initio" folding using MD.

Combined experimental and simulation studies on longer peptides and peptides of different amino acid sequences are expected to produce valuable information about the structural and kinetic aspects of peptide dynamics and contact formation. The dynamics of unfolded polymer chains are relevant not only in protein folding but also in cell adhesion.[43] Furthermore, probing the unfolded state of proteins will provide essential guidance for the refinement of force field models used in MD simulations. Short peptides are ideally suited for this task because they fully relax within the time scale of MD simulations.[42,44−46] However, consideration of both structure and kinetics is critical, as highlighted by the good agreement of kinetic data for different force fields with each other and experiments despite large variations in conformational preferences. This work also underscores the difficulties of comparing MD simulations to experiments, often requiring simplifying descriptions of the experimental measurements, such as the assumption of an explicitly distance-dependent quenching of tryptophan triplet states by cysteine. Nevertheless, such detailed comparisons of peptide measurements and simulations provide valuable insights into the properties of unfolded proteins, and will be instrumental for the success of MD simulations in protein folding and structure refinement.

Acknowledgment. We want to thank Attila Szabo, Jim Hofrichter, William A. Eaton, and Lisa J. Lapidus for many helpful discussions. This study utilized the high-performance computational capabilities of the Biowulf/LoBoS3 cluster at the National Institutes of Health, Bethesda, MD.

JA025789N

(41) García, A. E. *Phys. Rev. Lett.* **1992**, *68*, 2696.
(42) Hummer, G.; Garcıa, A. E.; Garde, S. *Proteins: Struct., Funct., Genet.* **2001**, *42*, 77.
(43) Jeppesen, C.; Wong, J. Y.; Kuhl, T. L.; Israelachvili, J. N.; Mullah, N.; Zalipsky, S.; Marques, C. M. *Science* **2001**, *293*, 465.
(44) Daura, X.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *J. Mol. Biol.* **1998**, *280*, 925.
(45) Daura, X.; Gademann, K.; Schäfer, H.; Jaun, B.; Seebach, D.; van Gunsteren, W. F. *J. Am. Chem. Soc.* **2001**, *123*, 2393.
(46) Hummer, G.; Garcıa, A. E.; Garde, S. *Phys. Rev. Lett.* **2000**, *85*, 2637.