Dynamic elements and kinetics: Most favorable conformations of peptides in solution with measurements and simulations

Gouri S. Jas [iD], Ricardo Vallejo-Calzada, Carey K. Johnson, and Krzysztof Kuczera [iD]

View Online   Export Citation   CrossMark

Dynamic elements and kinetics: Most favorable conformations of peptides in solution with measurements and simulations

View Online   Export Citation   CrossMark

Gouri S. Jas,[1,2,a)] (iD) Ricardo Vallejo-Calzada,[2] Carey K. Johnson,[3] and Krzysztof Kuczera[3,4] (iD)

AFFILIATIONS

[1] Department of Pharmaceutical Chemistry, The University of Kansas, Lawrence, Kansas 66047, USA
[2] School of Medicine, Central University, Bayamon, Puerto Rico 00956, USA
[3] Department of Chemistry, The University of Kansas, Lawrence, Kansas 66045, USA
[4] Department of Molecular Biosciences, The University of Kansas, Lawrence, Kansas 66045, USA

[a)] Author to whom correspondence should be addressed: gouri.jas@physicianscientists.org

ABSTRACT

Small peptides in solution adopt a specific morphology as they function. It is of fundamental interest to examine the structural properties of these small biomolecules in solution and observe how they transition from one conformation to another and form functional structures. In this study, we have examined the structural properties of a simple dipeptide and a five-residue peptide with the application of far-UV circular dichroism (CD) spectroscopy as a function of temperature, fluorescence anisotropy, and all-atom molecular dynamics simulation. Analysis of the temperature dependent CD spectra shows that the simplest dipeptide N-acetyl-tryptophan-amide (NATA) adopts helical, beta sheet, and random coil conformations. At room temperature, NATA is found to have 5% alpha-helical, 37% beta sheet, and 58% random coil conformations. To our knowledge, this type of structural content in a simplest dipeptide has not been observed earlier. The pentapeptide (WK5) is found to have four major secondary structural elements with 8% 3_{10} helix, 14% poly-L-proline II, 8% beta sheet, and 14% turns. A 56% unordered structural population is also present for WK5. The presence of a significant population of 3_{10} helix in a simple pentapeptide is rarely observed. Fluorescence anisotropy decay (FAD) measurements yielded reorientation times of 45 ps for NATA and 120 ps for WK5. The fluorescence anisotropy decay measurements reveal the size differences between the two peptides, NATA and WK5, with possible contributions from differences in shape, interactions with the environment, and conformational dynamics. All-atom molecular dynamics simulations were used to model the structures and motions of these two systems in solution. The predicted structures sampled by both peptides qualitatively agree with the experimental findings. Kinetic modeling with optimal dimensionality reduction suggests that the slowest dynamic processes in the dipeptide involve sidechain transitions occurring on a 1 ns timescale. The kinetics in the pentapeptide monitors the formation of a distorted helical structure from an extended conformation on a timescale of 10 ns. Modeling of the fluorescence anisotropy decay is found to be in good agreement with the measured data and correlates with the main contributions of the measured reorientation times to individual conformers, which we define as dynamic elements. In NATA, the FAD can be well represented as a sum of contributions from representative conformers. This is not the case in WK5, where our analysis suggests the existence of coupling between conformational dynamics and global tumbling. The current study involving detailed experimental measurements and atomically detailed modeling reveals the existence of specific secondary structural elements and novel dynamical features even in the simplest peptide systems.

Published under license by AIP Publishing. https://doi.org/10.1063/1.5131782

INTRODUCTION

Most biologically active molecules such as peptides and proteins, in solution, progress from the primary structure to a single, most stable, and unique native structure.[1–10] Upon achieving the native state, these biomolecules can transition between different conformers within a free-energy trough.[11–13] These conformational changes are often dictated by their functional responsibility. A quantitative insight into the mechanism of how these biomolecules assemble both *in vitro* and *in vivo* is important for a description of

their functional properties.[14–17] For this quantification to take place, we must examine not only the dynamical elements (formation of conformers and their interchange in solution) of these peptides and proteins in solution but also their kinetic pathways and transition state properties.

At the atomic level, no two trajectories are identical. As a protein or peptide progresses from an extended conformation to a properly folded state, it may take a substantially large number of pathways before achieving its native conformation. Optimally reducing them to a set of most favorable pathways with the application of theory and all-atom molecular dynamics (MD) simulation trajectory and comparing them with the experimental measurement is critical. Substantial efforts are made to develop theoretical model systems for gaining insight into the progression of the relaxation pathway of these reactions.[6,18–20] Applying these models to an all-atom molecular dynamics simulation trajectory and comparing with the experimental measurement is an ultimate test to validate theoretical approaches with reality, and a successful model may discover new features along the reaction pathway.

The solution environment plays an important role in the conformational stability as well as the function of biomolecules. The arrangement and orientation of water molecules with respect to neighboring residues exert influence on the solute conformational stability and intra- and inter-residue interactions.[21,22] The relationship between the water solvation shell and the water-driven ensemble average of preferred local conformations in the aqueous environment is essential for the functional aspects of peptides and proteins in the native state. More specifically, simpler systems such as dipeptides and pentapeptides with an accessible experimentally measured structure and dynamics combined with atomically detailed simulations and application of sophisticated theory in a reduced dimensional space can provide important new insights into the structure, transition state, and kinetic pathways of peptides and proteins in solution.

In this work, we present an anatomically detailed picture of the structure, dynamic elements, and kinetics pathways of dipeptides and pentapeptides in the aqueous environment with far-UV circular dichroism (CD) spectroscopy, fluorescence anisotropy measurements, all-atom molecular dynamics simulation, and application of optimally reduced dimensionality (ORD). Far-UV circular dichroism measurements are carried out for a fundamental dipeptide, N-acetyl-tryptophan-amide (NATA), and a five-residue peptide (WK5) in the aqueous environment. Both NATA and WK5 exhibited a conformational equilibrium, sampling a range of secondary structural elements in solution, including helical, turn, and extended forms. Fluorescence anisotropy measurements revealed a roughly three times faster reorientation time for NATA compared to WK5. All-atom molecular dynamics simulations of NATA and WK5 sampled a wide range of peptide conformations, including some helical states, in agreement with far-UV circular dichroism spectroscopy. Calculated indole group rotational correlation times were 31 ps and 105 ps, in good agreement with the measured data. To gain further insight into the reorientational motion, we performed a dynamic element analysis. This involves following reorientations in solution of representative peptide conformers in the presence of dihedral restraints, which eliminate large-scale conformational transitions. For NATA, the overall fluorescence anisotropy decay (FAD) signal can be well represented by a sum of contributions from representative conformers. This is not the case in WK5, where our analysis suggests an existence of coupling between local conformational dynamics and global tumbling.

MD trajectories were also employed to perform kinetic network analysis for these two systems, based on generation of a kinetic matrix in a discretized state space, followed by kinetic coarse graining through Optimal Dimensionality Reduction.[20] Our kinetic analysis showed that both peptides undergo relaxation dynamics in the nanosecond time scale. In NATA, the relaxation rate is about 1 ns and involves a coupled transition of the sidechain and the backbone dihedrals. In WK5, the relaxation rate is on the 10 ns timescale, exhibiting a transition between a family of coil-like states and a distorted helix. For WK5, both the ordered and disordered states are heterogeneous and allow the structural characterization of several metastable folding intermediates. This model also determines that the peptide relaxation occurs through multiple pathways, with the number of paths increasing with the resolution of the model. Our study with careful experimental measurements, atomically detailed molecular dynamics simulation, and application of sophisticated theory involving ORD can successfully specify the presence of structural conformation, exchange of conformations, reorientational motion of the structural elements, relaxation dynamics, rates, and pathways by kinetic coarse graining microstates of a dipeptide and a pentapeptide system in solution.

METHODS

Materials

N-Acetyl-tryptophan-amide (NATA) was purchased from Sigma. The five-residue heteropeptide Ac-WA$_3$K-NH$_2$ (WK5) was obtained from GenScript Corporation with greater than 95% purity. Experimental measurements on NATA and WK5 were performed by dissolving in 20 mM acetate buffer of pH 4.8. The concentration for each sample is determined with the absorbance of tryptophan at 280 nm with a molar extinction coefficient of 5690 M^{-1} cm^{-1}.

Circular dichroism

Far-UV circular dichroism spectroscopy as a function of temperature of NATA and WK5 at pH 4.8 is measured on a JASCO 815 spectropolarimeter in a cylindrical cell with a path length of 0.5 mm and at a concentration of ~170 μM. Scans between 266 and 360 K were recorded in 10° increments. Singular value decomposition (SVD) was applied on the far-UV, temperature dependent CD spectra in the range between 170 and 260 nm to resolve overlapping spectral components.

Anisotropy decay measurements

Time correlated single-photon counting (TCSPC) was employed in fluorescence anisotropy measurements. NATA and WK5 fluorescence was generated with excitation at 290 nm by the third-harmonic of a mode-locked, cavity-dumped Mira Optima 900F/Pulse Switch Ti:sapphire laser pumped by a 10 W Verdi Laser (Coherent, Inc., Santa Clara, CA), with harmonic generation in a 5–050 Ultrafast Harmonic Generator (Inrad, Northvale, NJ) Fluorescence was collected at 350 nm with an 8-nm band pass monochromator (model 9030, Sciencetech Inc., Concord, ON, Canada). As

described in detail elsewhere,[23] parallel and perpendicular fluorescence polarizations were collected simultaneously in a T-format with an instrument response time <30 ps. Since decays of fluorescence with polarization parallel and perpendicular to the excitation polarization were collected simultaneously in two different detection channels with different detection efficiencies, a correction factor was incorporated into the fitting equations. The fitting parameter had a value $g \approx 0.52$, where g is the relative detection efficiency of fluorescence with polarization perpendicular to the excitation polarization relative to fluorescence with polarization parallel to the excitation polarization. Fluorescence decays were measured for samples in aqueous solution. Employing in-house software, fluorescence decays for parallel and perpendicular polarizations were globally fit to a two-exponential intensity decay coupled to a double-exponential anisotropy decay. Rotational correlation times for NATA and WK5 were obtained by globally fitting the fluorescence decays with polarizations parallel and perpendicular to the excitation polarization as described elsewhere.[23]

COMPUTATIONAL

NATA

For Ac-Trp-NH$_2$ (NATA), the system was generated as described previously,[24] and a molecular dynamics trajectory length of 200 ns with OPLS/AA,[25] TIP4P[26] water at 300 K, in a 47.90 Å cubic box, with GROMACS 4.5.6[27] was generated for data analysis. There are 33 NATA atoms, 3596 TIP4P[25] water molecules, and a total of 14 417 atoms in the solvated system. There are 200 001 trajectory frames saved every 1 ps.

WK5

For Ac-WA$_3$K-NH$_2$ (WK5), an extended structure was built with CHARMM.[28] Solvation, generation of Na$^+$ and Cl$^-$ ions for a neutral solution at 0.15M ionic strength, energy minimization, and 10 ns of equilibration under NPT conditions at 300 K and 1 bar were performed with GROMACS 5.1.[29] A trajectory length of 5000 ns with OPLS/AA,[25] TIP4P[26] water at 300 K, with 23 336 atoms in a 56 Å cubic box, with GROMACS 5.1[29] was generated for data analysis. There are 85 peptide atoms, 5804 TIP4P[26] water molecules, and 18 Cl$^-$ and 17 Na$^+$ ions, and lysine is protonated in the solvated system. There are 5 000 001 frames saved every 1 ps.

Additional computational details are described in the supplementary material. Hydrogen bonds were counted as formed when the O$_i$···N$_j$ distance was below 3.6 Å, with $j = i + 3$ for 3_{10}, $j = i + 4$ for α-helical, and $j = i + 5$ for π-helical structures. Additional secondary structure analysis was performed with DSSP[30] and STRIDE.[31] Root mean square deviation (RMSD) clustering was done with the GROMACS algorithm, and dihedral clustering was done with CHARMM.[28] Statistical error estimates were obtained by dividing data into consecutive blocks and calculating the standard error of the mean at the 95% confidence level. To analyze peptide kinetics, time autocorrelation functions $C_1(t) = \langle (x(0) - \langle x \rangle)(x(t) - \langle x \rangle) \rangle$ were calculated, with x being the radius of gyration, end-to-end distance, number of hydrogen bonds, RMSD, and surface area. These functions were fitted to two-exponential decays $B * \exp(-t/\tau_1) +$

$C * \exp(-t/\tau_2)$, and the longer time scale τ_2 was used to estimate the global relaxation time of the peptide.

Fluorescence anisotropy decay was modeled by calculating the autocorrelation function for reorientation of the tryptophan ^{1}L$_b$ axis,[24,32] $C_2(t) = 1/2 \langle 3\cos^2(\theta) - 1 \rangle$, with θ being the axis reorientation angle during time t, and fitting to a two-exponential decay as described above, with the longer time scale τ_2 compared to the experimental rotational correlation times. To analyze the molecular structures contributing to the FAD signal, we performed a dynamic element analysis. This involved generating dihedral-restrained MD trajectories corresponding to selected peptide structures obtained for NATA and WK5. In each case, the closest-lying solvated structure was extracted from the regular MD trajectory and a dihedral restraint with a force constant of 100 kJ mol^{-1} rad^{-2} applied, with reference values set to the dihedral cluster centers. In this way, eight 50 ns trajectories were generated for NATA, with restraints on the backbone ϕ, ψ and sidechain χ_1, χ_2, and six 500 ps trajectories for WK5 with restraints on all five pairs of backbone ϕ, ψ in this blocked pentapeptide. In these trajectories, the restrained dihedral exhibited standard deviations of about 8°–10° around the reference values. From each trajectory, the $C_2(t)$ correlation function of the Trp^{1}L$_b$ axis was calculated, and a rotational correlation time τ_2 was extracted by fitting to a biexponential decay.

For kinetic analysis, we started with dihedral clustering[33] of trajectory structures, using samples of 200 001 structures sampled every 1 ps for NATA and 100 001 structures sampled every 50 ps for WK5. Next, we performed discretization of the trajectory by assigning each frame to a cluster and counting transitions and cluster/microstate populations. This was followed by calculation of a kinetic matrix **K**, obtained by dividing the number of transitions by the residence time in the initial state. For the lowest dimensions, the populations, kinetic matrix, and the flux matrix were used to describe the kinetics directly. Eigenvalues of the kinetic matrix were calculated to describe the system kinetic lifetimes. Additionally, **K** and its eigenvalues and eigenvectors were used as the basis of reduced dimensionality analysis using the PCCA+ algorithm[34] combined with the optimal dimensionality reduction approach of Hummer and Szabo.[35] The discretization was performed using an adjustable core radius R$_c$, i.e., transitions were counted when the trajectory moved from the core region of the cluster/microstate j to the core region of the cluster/microstate i. R$_c$ was adjusted so that the lowest nonzero eigenvalue λ_2 of **K**, corresponding to the slowest kinetic relaxation time $\tau_2 = 1/\lambda_2$, agreed with the time scales associated with the analysis of autocorrelation functions of dynamic observables from MD.

EXPERIMENTAL RESULTS

Circular dichroism

Far-UV circular dichroism (CD) spectra of NATA and the pentapeptide WK5 at pH 4.8 as a function of temperature between −10 °C and 90 °C are presented in Fig. 1. Figure 1(a) shows the CD spectra of NATA as a function of temperature. Observed spectral features, with two minima at 210 and 198 nm and two maxima at 225 and 185 nm, represent a population with ordered structures in CD spectra that change as a function of increasing
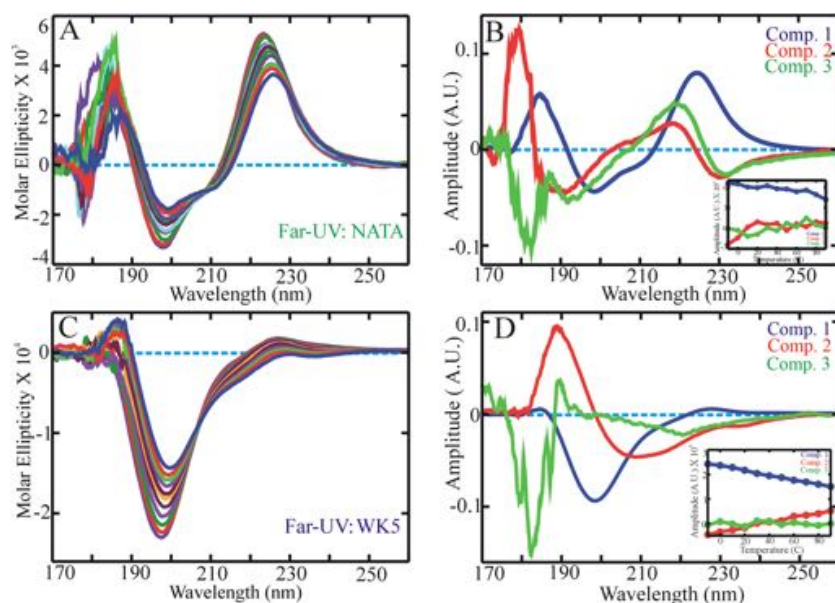
FIG. 1. Far-UV circular dichroism (CD) spectra and singular value decomposition (SVD) analysis of NATA and WK5. (a) Far-UV CD of NATA from −10 to 90 °C. (b) First three SVD components (blue, red, and green) of NATA temperature dependent CD. Inset: Amplitude vectors of the first three components (blue, red, and green) of NATA shown as a function of temperature. (c) Far-UV CD of WK5 from −10 to 90 °C. (d) First three SVD components (blue, red, and green) of the WK5 CD spectra. Inset: Amplitude vectors of the first three components from SVD analysis of WK5 shown as a function of temperature.

temperature. Fig. 1(b) shows the CD spectral component analysis of the CD spectra with the application of singular value decomposition (SVD)[36] on the complete set of temperature dependent CD spectra. Amplitudes of the first three components are presented with respect to wavelength. Presented in the inset of Fig. 1(b) are the amplitude vectors of first three components, with respect to temperatures.

Figure 1(c) shows the CD spectra of WK5 as a function of temperature with maxima at 227 and 185 nm and a minimum at 198 nm with an isodichroic point at 207 nm. There is a significant amplitude at 220 nm that changes as a function of temperature. The WK5 CD spectral component analysis with the application of SVD on the complete set of temperature-dependent CD spectra is presented in Fig. 1(d). Amplitudes of the first three components are presented with respect to wavelength. These three components show the presence of distinct populations of three primary secondary structural components.

Presented in the inset of Fig. 1(d) are the amplitude vectors of first three components shown in Fig. 1(d), with respect to temperature. The far-UV CD spectral signatures for both NATA and WK5 show the presence of a significant population of ordered structures. What is even more significant is that a small dipeptide like NATA exhibits an ordered structure in solution. When we examine the model CD spectra[37] of helix, hairpin, and random coil secondary structural elements, we find two minima at 222 nm and 210 nm and a maximum at 190 nm for a model helix. In a model hairpin, we find a minimum at 215 nm and a maximum at 195 nm, respectively. In a model random coil, we find the presence of slight maxima at 215 nm and 197 nm.

Previously, far-UV CD spectra of poly-L-proline, poly-L-glutamic acid, and poly-L-lysine in their charged states as a function of temperature were examined by Tiffany and Krimm.[38] The CD spectrum of poly-L-proline II exhibits a maximum at 225 nm

and a minimum at 198 nm. In poly-L-glutamic acid and poly-L-lysine, a maximum is present at 215 nm and a minimum at 198 nm. Based on the CD spectral changes with temperature, the authors assigned the polypeptides to an unordered chain conformation and speculated that there may be another state with an extended helical conformation. They further observed an isodichroic point and/or a region cross over at about 206–209 nm. These findings lead them to suggest that three states such as α-helix, the extended helix, and unordered conformations are the origin of their CD spectral features.

Here, application of SVD on the temperature dependent CD measurements of NATA resolved a dominant first component with minima at 212 nm and 200 nm and maxima at 225 nm and 180 nm. Two other components have a maximum at 220 nm and a minimum around 190 nm. Component 2 and component 3 intersect at 210 nm. These findings are consistent with the presence of a poly-L-proline II like conformation. Component 1 has a spectral feature similar to that of an α-helix but with a shift of the minima to shorter wavelength. Even in a dipeptide system as simple as NATA, we observe the possible presence of a distorted helix/extended helix as well as a disordered conformation. In the case of WK5, a pentapeptide, the application of SVD on the temperature dependent CD measurements resolved three components, while higher components are found to be noisy. The first component has a maximum at 225 nm and a minimum at 200 nm. This conformation is poly-L-proline II-like. The second component has a minimum with a small amplitude at around 235 nm and a pronounced minimum at 212 nm and a maximum at 19 2 nm. This spectral feature resembles a conformation like a model hairpin. The third component, nosier than the other two, shows a distinct minimum at 220 nm. The amplitude at 220 nm is small but easily distinguishable. These three components resolved from the spectral analysis of WK5 point to the presence of three distinct conformations in solution. Poly-L-proline II and hairpins are higher in

population with α-helix being lower. Application of DichroWeb[39] on the measured CD spectra of NATA and WK5 derived a percent population of the secondary structural components that are consistent with the SVD component analysis. For the dipeptide NATA, the specified components are an α-helical population of 5%, a beta sheet element population of 37%, and a random coil population of 58%, at 30 °C. For the pentapeptide WK5, DichroWeb derived percent population of components are a 3_{10} helix of 8%, PPII of 14%, unordered population of about 56%, turns of 14%, and beta-sheet of 8% at 30 °C. The analysis of CD data for NATA and WK5 confirms the presence of a rich heterogeneous population of secondary structural elements even in a simplest dipeptide and in a pentapeptide.

Fluorescence anisotropy decay

The solvent induced motion of NATA and WK5 and their direct interaction with the immediate solution environment are monitored with time-resolved fluorescence anisotropy decay measurements. Fluorescence decays for parallel and perpendicular polarization and the corresponding anisotropy decays at 30 °C in aqueous solvent are shown in Figs. 2(a) and 2(b). In Fig. 2(a), the top panel shows a semilog plot of the fluorescence intensity in the

parallel (blue) and perpendicular (green) polarizations for NATA. A fit to the data is shown in red. The bottom panel of Fig. 2(a) shows the fluorescence intensity in the parallel (blue) and perpendicular (green) polarizations for WK5. The fluorescence decays for WK5 are significantly faster in both polarizations compared to those for NATA. The solvent conditions for both NATA and WK5 are identical. The experimental anisotropy is determined from the measured fluorescence intensity ratio shown in Fig. 2(a). The top panel of Fig. 2(b) (blue) shows the anisotropy for NATA at 30 °C in aqueous solution. A fit to the anisotropy decay is shown in red. The inset in the top panel of Fig. 2(a) shows the entire decay up to 12 ns for NATA. The bottom panel of Fig. 2(b) (green) shows the anisotropy for WK5 at 30 °C in aqueous environment. A fit to the anisotropy decay is shown in red. The inset in the bottom panel of Fig. 2(b) shows the entire decay up to 12 ns for WK5. The anisotropy decay time derived from the experimental decay is observed to be 45 ps for NATA and 120 ps for WK5. This analysis suggests that NATA has a significantly more compact shape and experiences less solvent interaction. WK5, on the contrary, has a reorientation that is slower by a factor of 2.7 compared to NATA, suggesting a significantly larger size.

Experimental summary

Application of far-UV circular dichroism, fluorescence lifetime measurements, and fluorescence anisotropy decay measurements for two simple peptide systems, NATA and WK5, presented a very interesting and consistent insight into the structural properties and dynamical behavior in solution. Structurally, it is observed that even a dipeptide as simple as NATA can occupy structures like α-helix (5%), beta-sheet (37%), and random coil (58%) conformations in a varying degree of percent population. Structural analysis of WK5 derived four ordered secondary structural elements like poly-L-proline II (14%), a 3_{10} helix (8%), turns (14%), beta sheet (8%), and an unordered population of conformers (56%).

Fluorescence anisotropy measurements of NATA and WK5 generated two very different reorientation times under identical temperature and solvent conditions. The reorientation time for WK5 is found to be slower by a factor about three compared to that for NATA. These results illustrate a large difference in shape and size of these two peptide systems in an identical solution condition, with WK5 clearly being larger compared to NATA.

In order to obtain an atomically detailed picture of these experimental observables and develop a model system with sophisticated theory that adequately characterizes these experimental observables, we have carried out an all-atom molecular dynamics (MD) simulation for both NATA and WK5 in solution.

NATA MD

The results of the NATA MD simulations are briefly outlined below and presented in Figs. 3–6. More details are described in the supplementary material. The conformations of NATA were described primarily by four dihedral angles—the backbone ϕ and ψ and sidechain χ_1 and χ_2. The probability distributions of these angles are analogous to those described in our previous work[24,30] which employed a shorter trajectory. Basically, the backbone samples four main conformations, which may be classified approximately as
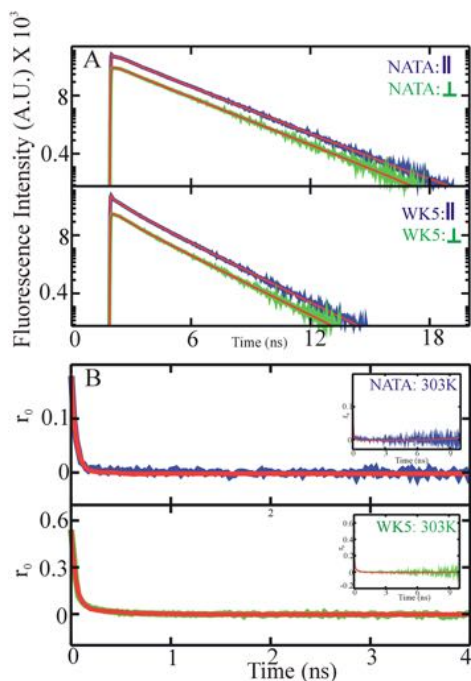
FIG. 2. Time-correlated single-photon counting is employed to measure fluorescence decays for NATA and WK5 in solution. (a) Fluorescence decay with polarization parallel to the excitation (blue) and perpendicular to the excitation (green), at 30 °C. The decays for fluorescence with polarization perpendicular to the excitation polarization are about 30% of the intensity of fluorescence with parallel polarization due to the different detection efficiencies in the two detection channels. Fit to the parallel and perpendicular fluorescence decays is in red. (b) Fluorescence anisotropy with inset showing the entire measured data.
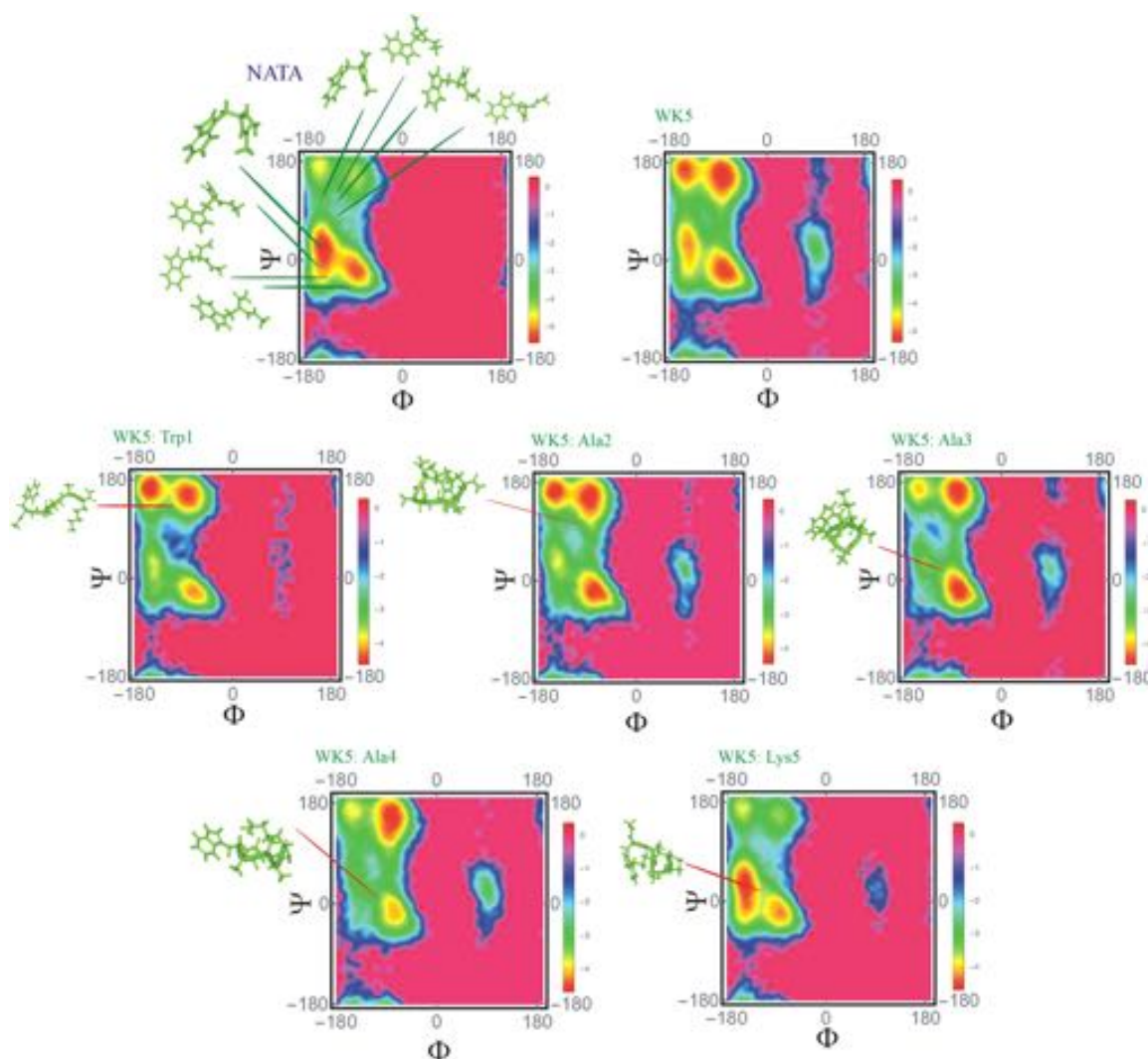
FIG. 3. MD of NATA and WK5. Backbone dihedral angle distribution is shown as potential of mean force $G(\phi, \psi) = RT \ln P(\phi, \psi)$ at T = 300 K. NATA: (ϕ, ψ) distribution from NATA MD. WK5: aggregate (ϕ, ψ) distribution for all five residues and individual residues, Trp1, Ala2, Ala3, Ala4, and Lys5.

α, β, PPII, and turn-like state $(\phi, \psi) = (-140°, 10°)$, see Fig 3. The sidechain χ_1 samples $g-$, t, and $g+$ and χ_2 samples the *syn-* and *anti*-conformations. The distribution of the χ_1 values is shown in Fig. 4(a), and other data are shown in the supplementary material.

The ranges of shapes and sizes explored by NATA are shown in Figs. 4(b) and 4(c). The radius of gyration, measuring the average molecular size, samples only a limited range [Fig. 4(a)]. However, the range of shapes, as measured by the δ^* parameter,[36] indicates that the system takes on a wide range of shapes, from spherical ($\delta^* = 0$) to elongated ($\delta^* = 1$), see Fig. 4(c).

NATA generally exhibits moderate interactions with water.[32] Samples of pair distribution functions [Fig. 5(a)] show that, in this system, water-water interactions are strongest. Relatively strong interactions exist between water and the peptide carbonyl oxygen

atoms, while interactions of the rest of NATA with the solvent are weaker. Specific solvation of the indole nitrogen by water can be seen in Fig. 6.

To estimate the time scale of the global dynamics of NATA, we analyzed the fluctuations of several quantities, including the four dihedrals discussed above, as well as the radius of gyration, RMSD from the starting structure, surface area, and the length of the internal hydrogen bond (supplementary material). Autocorrelation functions of the quantities were calculated and fitted to two-exponential decays. This led to the estimate of 1.0 ns as the longest relaxation time in the NATA dipeptide. Distributions of several quantities and their ACFs are shown in Fig. 4 and supplementary material.

Several quantities calculated for NATA from the MD trajectory may be related to experimental observation. The backbone
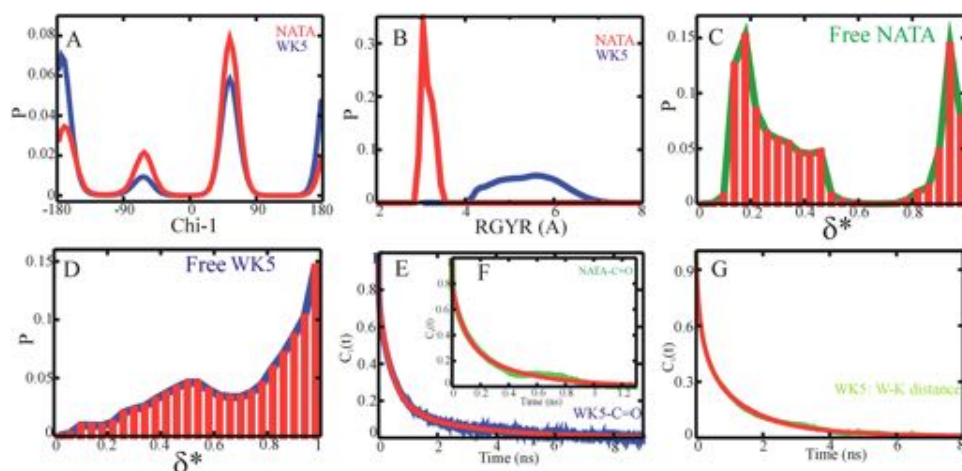
FIG. 4. MD of NATA and WK5. (a) Distribution of sampled values of Trpχ_1 dihedral. (b) Distribution of the radius of gyration values. (c) Distribution of the shape parameter δ^* for NATA. (d) Distribution of the shape parameter δ^* for WK5. (e) Autocorrelation of fluctuation of indole ring distance to Trp amide C=O in WK5. (f) Autocorrelation of fluctuation of indole ring distance to Trp amide C=O in NATA. (g) Autocorrelation of fluctuation of indole ring distance to lysine NZ atom in WK5.

conformations sampled in the MD, α, β, PPII, and turn (ϕ, ψ) = ($-140°$, $10°$), with populations of 31%, 10%, 9%, and 50%, respectively, are qualitatively similar to the CD measurement results.

The results for reorientational dynamics of NATA in the MD simulations are presented in Fig. 7. The predicted rotational correlation time from the MD trajectory τ_2 is 31 ps. This value is very close to our previous MD result, 30 ps, obtained from a shorter trajectory.[24,32] It is also comparable to the experimental measurement

of 45 ps described above. More details are presented in the section titled "Dynamic elements."

NATA kinetic modeling

Dihedral clustering was performed in the four-dimensional space of two backbone (ϕ and ψ) and two sidechain (χ_1 and χ_2) dihedrals of NATA, using the ART 2′ algorithm of CHARMM.[28] Here,
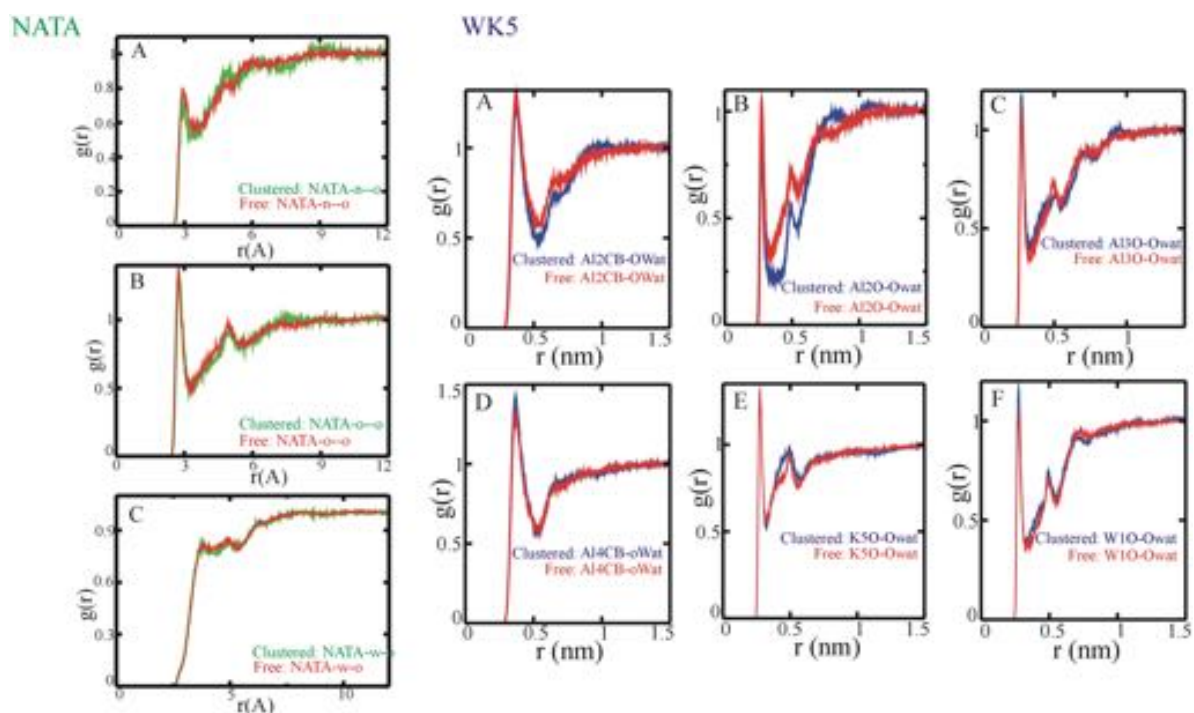
FIG. 5. MD of NATA and WK5—solvation described by radial distribution functions g(r). Results of full MD trajectory (Free) compared with dihedral-restrained trajectory for the most populated cluster structure (Clustered). NATA: (a) NATA peptide N···water O. (b) NATA peptide O···water O. (c) NATA indole ring···water O. WK5: (a) Ala2 CB···water O. (b) Ala2 O···water O. (c) Ala3 O···water O. (d) Ala 4 CB···water O. (e) Lys 5 O···water O. (f) Trp O···water O.
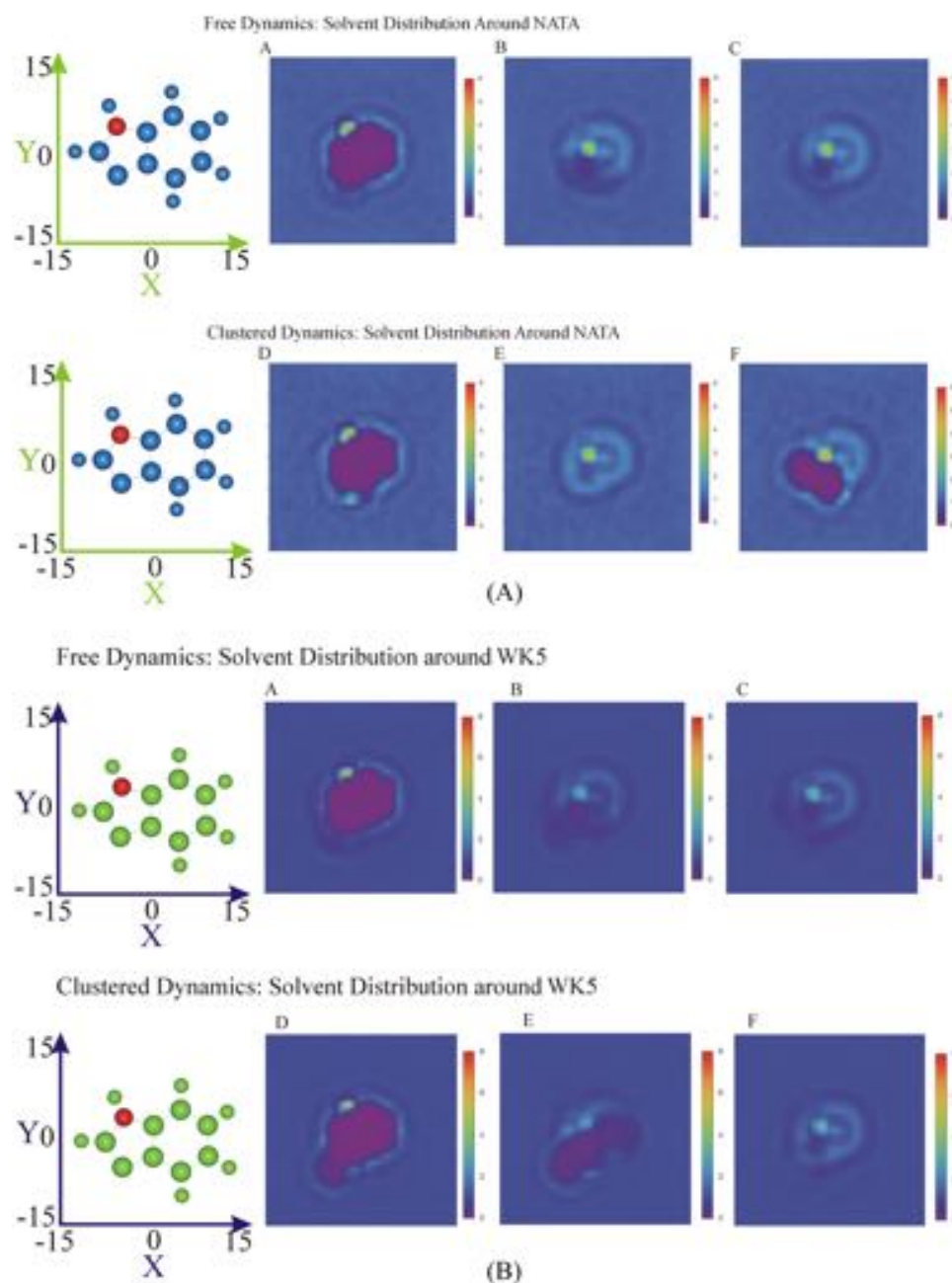
FIG. 6. 3D solvent distributions around the indole ring. Colors code for the ratio of local water density compared to average, from 0 (purple) to 6 (red). Side structure shows indole atom locations in the same coordinate system. Full MD trajectory of NATA: (a) in the indole plane, (b) 3 Å above the indole plane, and (c) 3 Å below the indole plane. Representative cluster restrained MD: (d) in the indole plane, (e) 3 Å above the indole plane, and (f) 3 Å below the indole plane. Full MD trajectory of WK5: (a) in the indole plane, (b) 3 Å above the indole plane, and (c) 3 Å below the indole plane. Representative cluster restrained MD of WK5: (d) in the indole plane, (e) 3 Å above the indole plane, and (f) 3 Å below the indole plane.

we describe the results of analysis with cluster radius R = 90°, which yielded eight clusters, N_c = 8. To reproduce the global dynamics rate of 1 ns found in the MD simulations, a core radius R_c of 45° was employed to count transitions using the lifetime-based approach.[35] More kinetic models are presented in the supplementary material.

The relaxation times generated by the full kinetic matrix K with N_c = 8 and the different reduced dimensionality models are presented in Table I. The initial eight clusters are

described in Table II, which shows their populations, kinetic lifetimes, and structural characteristics, and images are presented in Fig. 8.

Reduced dimensionality models for NATA are described in Figs. 9 and 10, as well as Tables I and III–VIII.

Figs. 9(a) and 10(a) and Tables III and IV show the state properties and effective rate matrix for the reduced dimensionality model with N = 2 mesoscopic states. State A consists of clusters 1, 2, 3, and 7, while state B includes clusters 4, 5, 6, and 8 (Table II). Thus, the
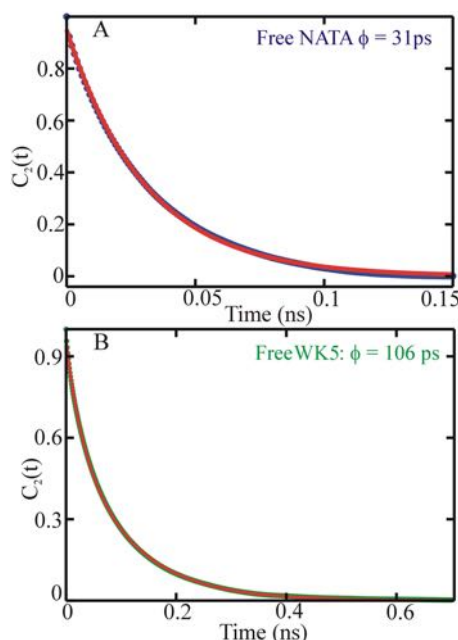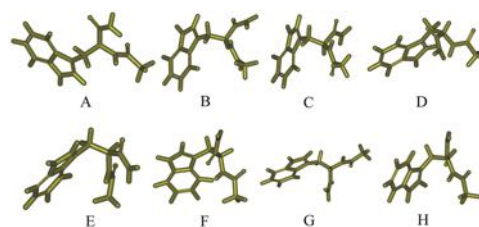
FIG. 7. $C_2(t)$ correlation functions of Trp^{1}L$_b$ axes from MD simulations. (a) NATA. (b) WK5. Simulation results are in red, and single-exponential fits are in blue.

FIG. 8. NATA MD simulations. Central structures of eight dihedral clusters. Structures (a)–(h) correspond to states 1–8 in Table II.
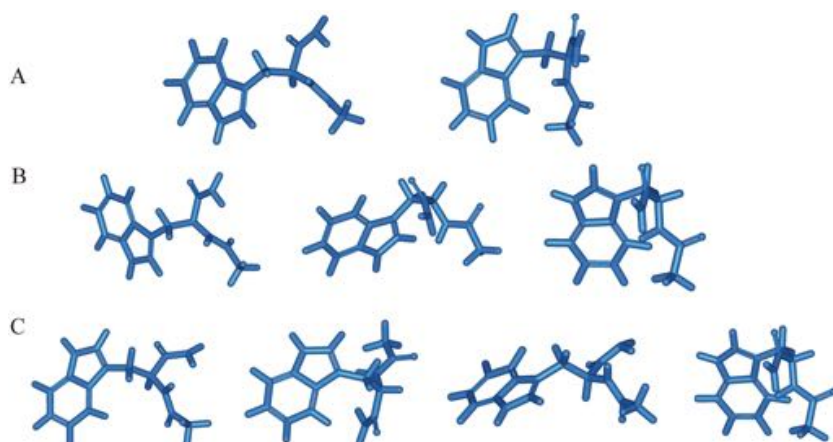
slowest transition is due to conformational changes of the trypto-phan χ_1 dihedral, between the t/g− conformers and g+. The A–B transition also involves a coupled change in backbone conformation, h/p to tl/b.

Figs. 9(b) and 10(b) and Tables V and VI show the state properties and effective rate matrix for the reduced dimensionality model with N = 3 mesoscopic states. State A consists of clusters 1, 2, 3, and 7 (same as state A for N = 2), state B consists of clusters 4 and 8, and state C consists of clusters 5 and 6. Thus, the second slowest transition is primarily due to conformational changes of the tryptophan χ_2 dihedral between syn and anti states.

Figs. 9(c) and 10(c) and Tables VII and VIII show the state properties and effective rate matrix for the reduced dimensionality model with N = 4 mesoscopic states. In this case, state A includes clusters 2 and 3, state B clusters 4 and 8, state C clusters 1 and 7, and state D clusters 5 and 6, i.e., each of the two states obtained for N = 2 has been split into two halves. Thus, it appears that the third slowest transition is related to coupled transitions of χ_1, χ_2 and backbone dihedrals.

Kinetic summary

This small model peptide samples multiple conformations on the time scale of 1 ns. The backbone explores helix, turn, beta, and PPII structures. Using our approach, the dynamics may be well described by kinetic models of dimensionality 2–8. The slowest motions involve transitions between the sidechain dihedral conformers coupled to backbone changes.

TABLE I. NATA dihedral clustering: slowest kinetic lifetimes, ns. Full kinetic matrix (column 1) and reduced dimensionality models for N = 2, 3, 4.

R = 90°, N_c = 8, R_c = 45°	N = 2	N = 3	N = 4
0.96	0.93	0.94	0.96
0.50		0.48	0.50
0.32			0.32
0.12			
0.10			

TABLE II. Relative free energies and kinetic lifetimes—WK5 dihedral clustering. Raw clusters, N_c = 8. Notation for backbone conformations: helix or "h" $(\phi, \psi) \approx (-100°, -10°)$; beta or "b" $(\phi, \psi) \approx (-140°, 150°)$; PPII or "p" $(\phi, \psi) \approx (-90°, 130°)$; turn-like or "tl" $(\phi, \psi) \approx (-140°, 10°)$. Notation for sidechains: g−, t, g+ for $\chi_1 \sim -60°$, 180°, and +60°, respectively, "syn" for $\chi_2 < 0$, and "anti" for $\chi_2 > 0$.

State	ΔG (kcal/mol)	Lifetime = $-1/K_{ii}$ (ns)	$\phi, \psi, \chi_1, \chi_2$ (deg)	Type
1	0.42	0.26	−101, −14, −114, 85	ht/g− syn
2	0.45	0.12	−107, −11, −130, −94	h t anti
3	1.05	0.052	−114, 114, −162, −104	P t anti
4	0.40	0.40	−131, 0, 53, −89	tl g+ anti
5	0.00	0.41	−134, 7, 55, 87	tl g+ syn
6	1.13	0.088	−137, 154, 52, 89	b g+ syn
7	1.12	0.13	−106, 125, −138, 72	p t syn
8	1.50	0.14	−132, 154, 35, −84	b g+ anti

FIG. 9. NATA kinetic models. Representative structures for the (a) two-state model, N = 2, (b) three-state model, N = 3, and (c) four-state model, N = 4.

WK5 MD

The results of the WK5 MD simulations are briefly presented below and illustrated in Figs. 3–5. More details are described in the supplementary material. For WK5, we focused on the backbone ϕ and ψ of the individual residues. As shown in Fig. 3, the residue conformations involved mostly three conformers—helix, β, and PPII in all cases. Additionally, a turn structure was found to have the highest population for lysine (K5), and some left-handed helical structures appeared for the three alanine residues (areas with positive ϕ and ψ). In Fig. 4, we also plotted the distribution of the WK5 tryptophan χ_1, which is different from that found in NATA.

Secondary structure analysis with DSSP[30] showed the presence of bend (7%), turn (7%), and 3_{10} helix (1%) populations in the MD trajectory of WK5. The peptide is thus mostly unstructured, with a very small contribution from the 3_{10} helical state.

The ranges of shapes and sizes explored by WK5 are shown in Figs. 4(b) and 4(c). Both the radius of gyration, measuring the average molecular size [Fig. 4(b)], and the δ^* parameter,[40,41] measuring the shape [Figs. 4(c) and 4(d)], sample a wide range of values, showing that WK5 is much more flexible than NATA.

Interactions of WK5 with water are analyzed in Fig. 5(b). In this system, water-water interactions are strongest, followed by water-lysine NZ and water-peptide C=O. As in the case of NATA, interactions of the remaining parts of WK5 with the solvent are relatively weak. Specific solvation of the indole nitrogen by water in WK5 is quite similar to that seen in NATA. For comparison, these data are shown in the WK5 section of Fig. 6.

To estimate the time scale of the global dynamics of WK5, we analyzed the fluctuations of several quantities, including the end-to-end distance, the radius of gyration, RMSD from model α, 3_{10}, and π helices, and populations of the related hydrogen bonds. Autocorrelation functions of the quantities were calculated and fitted to two-exponential decays. This led to the estimate of 9 ns as the longest relaxation time in WK5. Distributions of several quantities and their ACFs are shown in Fig. 4 and supplementary material.

Several quantities calculated from the MD trajectory may be related to experimental data. The MD distribution of χ_1 dihedral values has three peaks, with populations of 8%, 48%, and 44% for the

$g-$, t, and $g+$ conformers,[42] respectively. The backbone conformations sampled in the MD were mostly in the helix, β, PPII, and turn $(\phi, \psi) \approx (-140°, 20°)$ types, with populations of 25%, 20%, 34%, and 19%, respectively, with a small contribution of ca. 1% from the left-handed region. These results are qualitatively similar to the results from CD measurements.

The results for reorientational dynamics of WK5 are presented in Fig. 7. The predicted rotational correlation time from the MD trajectory τ_2 is 105 ps, very close to the experimental measurement of 120 ps described above. More details are presented in the section titled "Dynamic elements."

Kinetic analysis

Dihedral clustering was performed in the ten-dimensional space of backbone (ϕ, ψ) dihedrals of the blocked peptide WK5, using the ART 2′ algorithm of CHARMM.[28,33] Here, we describe the results of analysis with cluster radius R = 98°, which yielded six clusters, $N_c = 6$. To reproduce the global dynamics rate of 9 ns found in the MD simulations, a core radius R_c of 42° was employed to count transitions using the lifetime-based approach.[35] More kinetic models are presented in the supplementary material.

The kinetic relaxation times obtained with the optimal model employing all six microstates are shown in Table IX. Relative free energies, lifetimes, and structures of the six states are presented in Table X, and images are shown in Fig. 10 (Table XI).

The structures may be categorized as extended-type (1, 2, 4, and 5) and turn-type (3 and 6), with structure 3 exhibiting one 3_{10} helical hydrogen bond (Trp1 C=O···H−N Ala4). Based on the analysis of backbone conformations presented above, these are representative structures for WK5 in the MD trajectory. The raw state kinetic lifetimes are in the 2.9–8.8 ns range. The most stable form is the extended conformer from cluster 1. The kinetics involves transitions between all six states, with the majority of the flux due to the four most populated states, 1, 2, 4, and 6.

To provide more insight into the conformational dynamics of WK5, we generated three reduced-dimensionality models, with N = 2, 3, and 4. These models are described in Figs. 12 and 13 and Tables IX and XII–XVII.
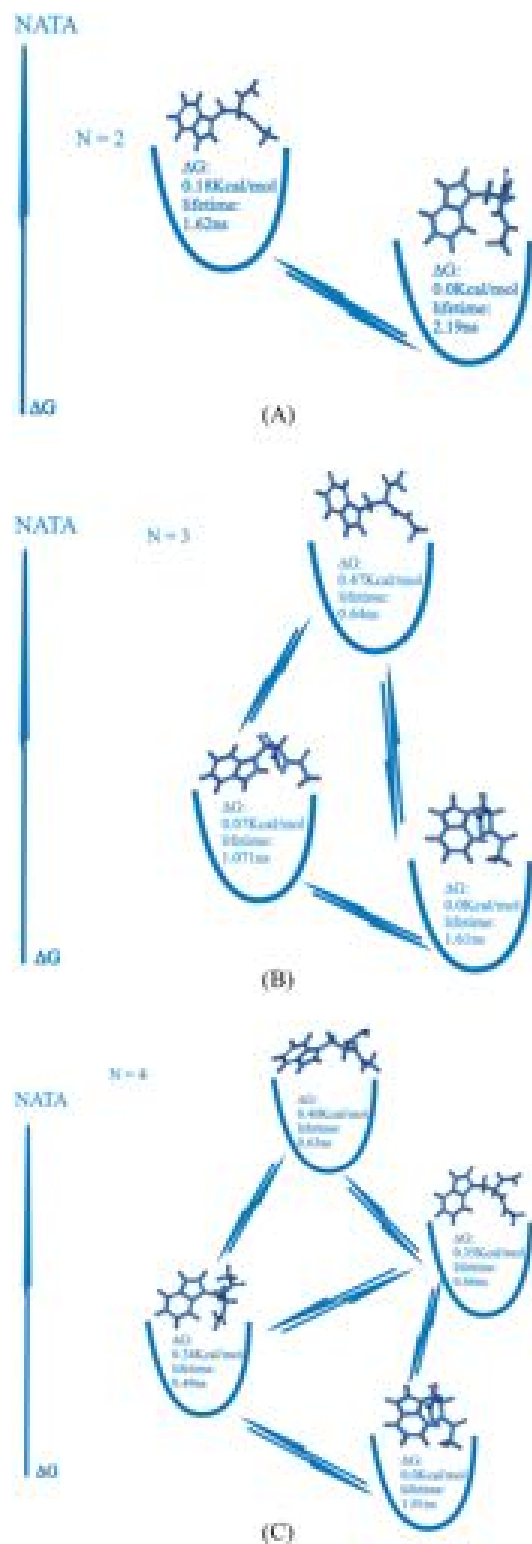
FIG. 10. Free energy diagrams for NATA kinetic models. (a) N = 2. (b) N = 3. (c) N = 4.

TABLE III. Reduced dimensionality model for NATA dihedrals, N_c = 8. Aggregate state properties, N = 2. The effective relaxation time is 0.93 ns.

State	A	B
No. of clusters	4	4
ΔG (kcal/mol)	0.18	0.00
Lifetime (ns)	1.62	2.19
Type	χ_1 = t, g—	χ_1 = g+

TABLE IV. Reduced rate matrix **R** for NATA dihedrals, N_c = 8, and N = 2. Diagonal elements are related to lifetimes $t_i = 1/R_{ii}$, off-diagonal elements are rates, and R_{IJ} represents the rate of J to I transition. Rows and columns ordered by states: A, B.

-0.616 658	0.456 190
0.616 658	−0.456 190

Aggregate state properties for N = 2 are presented in Tables XII and XIII and Figs. 11(a) and 13(b). In this case, state B is made up of a single cluster (cluster 3), which includes structures with turns and 3_{10} helix type hydrogen bonds, while state A includes the five remaining clusters. According to this model, there are a very slow "folding" transition from the relatively ordered turn/3_{10} helix state B to the unfolded ensemble A and a relatively fast "unfolding" of B to A. The lifetime of state 2 is 209 ns, which is quite long, much longer than any of the individual clusters. This long lifetime is due to the low population of the "folded" state B relative to A. The prediction here is that a long time is spent by the system on transitions within the substates of A.

Aggregate state properties for N = 3 are presented in Tables XIV and XV and Figs. 12(b) and 13(b). In this case, the single-cluster states are A (cluster 2, extended) and B (cluster 3, with turn and 3_{10} helix content), while state C involves four clusters,

TABLE V. Reduced dimensionality models for NATA dihedrals, N_c = 8. Aggregate state properties, N = 3. The effective relaxation times are 0.94 and 0.48 ns.

State	A	B	C
No. of clusters	4	2	2
ΔG (kcal/mol)	0.00	0.47	0.07
Lifetime (ns)	1.61	0.64	1.07
Type	χ_1 = t, g—	g+ anti	g+ syn

TABLE VI. Reduced rate matrix **R** for NATA dihedrals, N_c = 8, and N = 3. Diagonal elements are related to lifetimes $t_i = 1/R_{ii}$, off-diagonal elements are rates, and R_{IJ} represents the rate of J to I transition. Rows and columns ordered by states: A, B, C.

-0.619 417	0.548 754	0.412 005
0.250 751	−1.573 144	0.523 118
0.368 666	1.024 390	−0.935 123

TABLE VII. Reduced dimensionality models for NATA dihedrals, $N_c = 8$. Aggregate state properties, $N = 4$. The effective relaxation times are 0.96, 0.50, and 0.32 ns.

State	A	B	C	D
No. of clusters	2	2	2	2
ΔG (kcal/mol)	0.35	0.40	0.34	0.00
Lifetime (ns)	0.66	0.63	0.49	1.01
Type	h/p t anti	tl/b g+ anti	h/pt/g— syn	tl/b g+ syn

TABLE VIII. Reduced rate matrix **R** for NATA dihedrals, $N_c = 8$, and $N = 4$. Diagonal elements are related to lifetimes $t_i = 1/R_{ii}$, off-diagonal elements are rates, and R_{IJ} represents the rate of J to I transition. Rows and columns ordered by states: A, B, C, D.

−1.523 581	0.384 554	1.133 704	0.013 601
0.353 315	−1.582 563	0.129 096	0.539 261
1.145 796	0.142 009	−2.039 730	0.436 434
0.024 471	1.056 000	0.776 930	−0.989 296

mostly of extended type. In this model, the states spend a relatively long time exploring substates in the most populated state C, while state A is an intermediate on the path to the relatively structured state B.

Aggregate state properties for $N = 4$ are presented in Tables XVI and XVII and Figs. 12(c) and 13(c). In this case, the single-cluster states are A (cluster 3, with turn and 3_{10} helix content), B (cluster 4, extended), and C (cluster 2, extended), while state D involves three clusters, two extended and one turn type. The aggregate state D with the largest population, which we can call the "coil," also includes the

TABLE IX. WK5 dihedral clustering: slowest kinetic lifetimes, ns. Full kinetic matrix (column 1) and reduced dimensionality models for $N = 2, 3, 4$.

$R = 98°$, $N_c = 6$, $R_c = 42°$	$N = 2$	$N = 3$	$N = 4$
8.4	8.4	8.4	8.4
5.4		5.4	5.4
4.2			4.1
3.3			
2.9			

TABLE X. Relative free energies and kinetic lifetimes—WK5 dihedral clustering. Raw clusters, Nc = 6.

State	ΔG (kcal/mol)	Lifetime = −1/K_{ii} (ns)	Type
1	0.00	6.8	Extended
2	0.69	6.3	Extended
3	1.47	8.8	Turn/3–10
4	1.24	4.3	Extended
5	2.10	2.9	Extended
6	0.36	5.0	Turn

TABLE XI. Raw kinetic matrix **K** for WK5 dihedral clustering, $N_c = 6$; unit: ns^{-1}. Diagonal elements are related to lifetimes $t_i = 1/K_{ii}$, off-diagonal elements are rates, and K_{ij} represents the rate of j to i transition.

−0.146 264	0.101 802	0.034 759	0.156 480	0.185 721	0.159 106
0.031 851	−0.158 061	0.019 862	0.031 969	0.021 429	0.020 803
0.002 934	0.005 358	−0.114 209	0.008 413	0.007 143	0.006 934
0.019 488	0.012 725	0.012 414	−0.232 196	0.000 000	0.008 090
0.005 448	0.002 009	0.002 483	0.000 000	−0.342 869	0.006 934
0.086 543	0.036 167	0.044 690	0.035 334	0.128 576	−0.201 869

TABLE XII. Reduced dimensionality models for WK5 dihedrals, $N_c = 6$, with $N = 2$ aggregate states.

State	A	B
No. of clusters	5	1
ΔG (kcal/mol)	0.00	1.89
Lifetime (ns)	209.0	8.8
Type	Extended	Turn/3–10

largest number of clusters (high entropy) and has the longest lifetime. In this model, the states spend a relatively long time exploring substates in the most populated state D, while states B and D are intermediates on the path to the relatively structured state A. The slowest system relaxation time corresponds roughly to the lifetime of the "folded" turn/3_{10} state in all the models.

Kinetic summary

This is a highly flexible peptide sampling multiple conformations on the time scale of several nanoseconds. The peptide is mostly unstructured, sampling extended structures of β and PPII types, as well as turns. It also exhibits a small population of helical states,

TABLE XIII. Reduced rate matrix **R** for WK5 dihedrals, $N_c = 6$, and $N = 2$. Diagonal elements are related to lifetimes $t_i = 1/R_{ii}$, off-diagonal elements are rates, and R_{IJ} represents the rate of J to I transition. Rows and columns ordered by states: A, B.

−0.004 781	0.113 901
0.004 781	−0.113 901

TABLE XIV. Reduced dimensionality models for WK5 dihedrals, $N_c = 6$, with $N = 3$ aggregate states.

State	A	B	C
No. of clusters	1	1	4
ΔG (kcal/mol)	1.01	1.79	0.00
Lifetime (ns)	6.3	8.8	30
Type	Extended	Turn/3–10	Extended/turn

TABLE XV. Reduced rate matrix **R** for WK5 dihedrals, $N_c = 6$, and N = 3. Diagonal elements are related to lifetimes $t_i = 1/R_{ii}$, off-diagonal elements are rates, and R_{IJ} represents the rate of J to I transition. Rows and columns ordered by states: A, B, C.

−0.157 535	0.019 324	0.028 070
0.005 213	−0.113 905	0.004 702
0.152 322	0.094 581	−0.032 772

TABLE XVI. Reduced dimensionality models for WK5 dihedrals, $N_c = 6$, with N = 4 aggregate states.

| State	A	B	C	D
No. of clusters	1	1	1	3
ΔG (kcal/mol)	1.74	1.51	0.96	0.00
Lifetime (ns)	8.8	4.3	6.3	21
Type	Turn/3–10	Extended	Extended	Extended/turn

which appear to be mostly of 3_{10} type. Using our approach, the dynamics may be well described by kinetic models of dimensionality 2–6. In our analysis, the slowest motion involves transitions between the helix and remaining states. In this coarse-grained picture, the sampled states are very heterogeneous—each corresponds to a range of conformers. However, the microscopic analysis clearly shows a folding transition from mostly extended states to a 3_{10} helix.

Dynamic elements

Fluorescence anisotropy decay measurements follow molecular reorientations and are a sensitive probe of structure, motions, and solute-solvent interactions. To provide further insight into
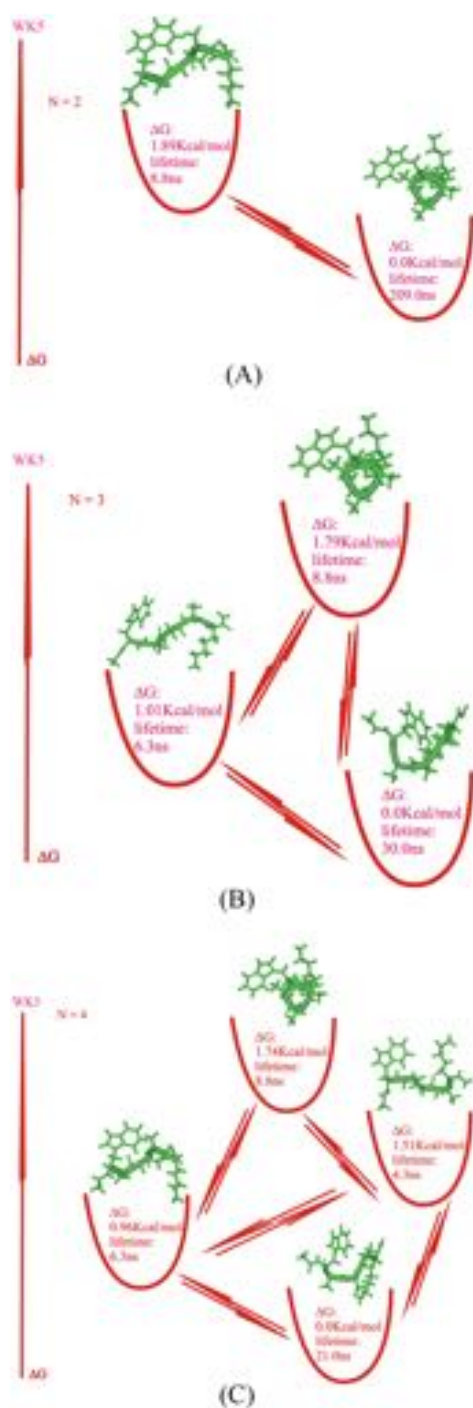
TABLE XVII. Reduced rate matrix **R** for WK5 dihedrals, $N_c = 6$, and N = 4. Diagonal elements are related to lifetimes $t_i = 1/R_{ii}$, off-diagonal elements are rates, and R_{IJ} represents the rate of J to I transition. Rows and columns ordered by states: A, B, C, D.

−0.113 973	0.007 943	0.005 183	0.004 455
0.011 721	−0.230 778	0.013 237	0.015 007
0.019 215	0.033 254	−0.157 581	0.027 675
0.083 037	0.189 581	0.139 161	−0.047 137
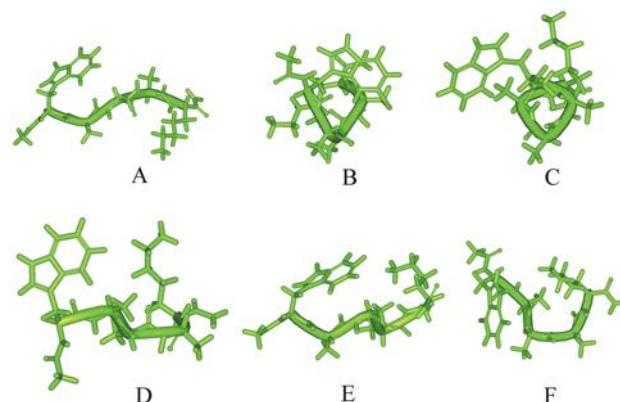
FIG. 11. WK5 MD simulations. Central structures of six dihedral clusters. Structures (a)–(f) correspond to states 1–6 in Table X.

the solution behavior of NATA and WK5, we have performed a dynamic element analysis. This procedure starts with generation of dihedral-restrained MD trajectories describing the motions of the central cluster structures. This is followed by calculation of the
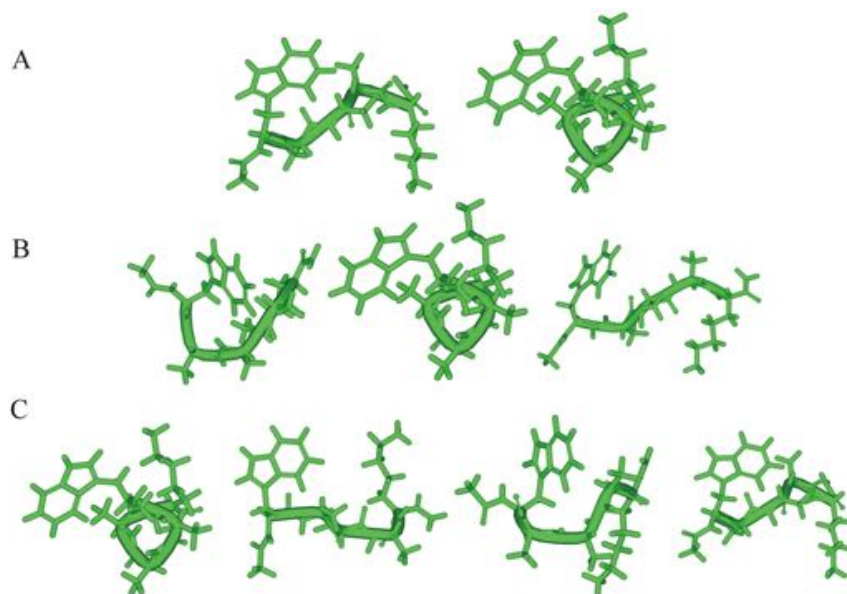
FIG. 12. WK5 kinetic models. Representative structures for the (a) two-state model, N = 2, (b) three-state model, N = 3, and (c) four-state model, N = 4.
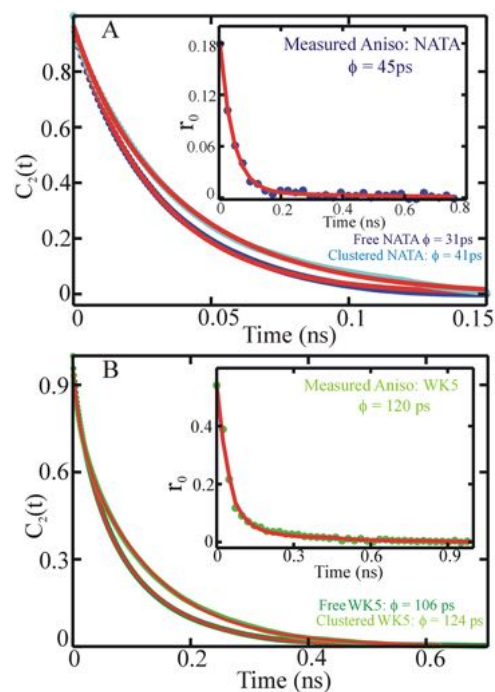
TABLE XVIII. Dynamics element rotational correlation times in ps.

NATA		WK5	
Cluster	τ_2	Cluster	τ_2
1	38	1	154
2	40	2	109
3	38	3	107
4	29	4	123
5	29	5	148
6	30	6	126
7	44		
8	34		
Average	34	Average	129
Most stable	29	Most stable	154

For NATA, we calculated rotational correlation times from dihedral restrained trajectories for the eight dihedral clusters in Table II. The results are presented in Fig. 12(a) and Table XVIII. The conformer relaxation times τ_2 are in the range of 29–44 ps, bracketing the free MD value of 31 ps. The population-weighted average τ_2 value is 41 ps, and the value for the most populated conformer

FIG. 13. Free energy diagrams for WK5 kinetic models. (a) N = 2. (b) N = 3. (c) N = 4.

reorientational times of these structures and comparison with the free MD and experimental data. Thus, we attempt to see how individual conformations sampled by these flexible peptides contribute to the overall observed spectroscopic signal.

FIG. 14. Dynamic elements. (a) NATA: comparison of $C_2(t)$ correlation functions of Trp^1L_b axes from free MD (dark blue) and a representative restrained MD trajectory (light blue). Inset: experimental NATA anisotropy decay. (b) WK5: comparison of $C_2(t)$ correlation functions of Trp^1L_b axes from free MD (dark green) and a representative restrained MD trajectory (light green). Inset: experimental WK5 anisotropy decay.

(cluster 5) is 29 ps. Thus, both the average and most favored cluster contributions represent the dynamics of NATA quite well. Interestingly, the NATA reorientational diffusion time scale, about 100 ps (i.e., three times τ_2), is comparable to some of the conformer lifetimes, which span 52–410 ps (Table II). However, possibly due to the relative narrow range of conformer reorientational times, we find that the overall dynamics is well represented by the contributing structures.

For WK5, we followed reorientations in dihedral-restrained trajectories for the six dihedral clusters in Table X. The results are presented in Fig. 12(b) and Table XVIII. The conformer rotational correlation times are in the 107–154 ps range. The ratio of highest to lowest times is about 1.4, comparable to 1.5 found for NATA. The population weighted correlation time τ_2 for WK5 is 129 ps, and the value τ_2 for the most stable conformer (cluster 1) is 154 ps. Thus, the WK5 conformer contributions generally tend to be higher than the free MD result τ_2 of 105 ps. Since introduction of dihedral constraints tends to slow down the reorientations of WK5, it appears that the conformational transitions, which are inhibited by the dihedral restraints, play an important role in determining these motions. Large scale dihedral fluctuations leading to transitions between clusters occur on much longer time scales, with lifetimes of the conformers falling in the 2.9–8.8 ns range (Table X) (Fig. 14).

DISCUSSION

Driven by the curiosity of determining the conformations and interchange rates of small peptides in solution, we carried out a combined range of experimental measurements, atomically detailed simulation, and kinetic analysis for a dipeptide and a pentapeptide system.

Structural, dynamic, and spectroscopic properties are examined experimentally with far-UV circular dichroism, nanosecond fluorescence lifetime measurements, and fluorescence anisotropy measurements. Data analysis of the experimental observations of a dipeptide (NATA) and a pentapeptide (WK5) shows that, in solution, they both adopt a range of structures. What is more revealing is that we can confirm experimentally that even a dipeptide as small NATA can occupy helical, β, and coil conformations in solution. The structural analysis of WK5 derived two dominant structural populations, PPII and β, as well as a less populated helical structure.

Anisotropy decay measurements show reorientation times of 45 ps for NATA and 120 ps for WK5. The larger peptide reorientations are thus about three times slower than those of NATA, indicating that the differences in size and sequence induce a very different morphology and interactions with the immediate environment.

All-atom molecular dynamics simulations presented comprehensive models of the structure and dynamics of NATA and WK5 that are consistent with the experimental observations and produced complementary findings on microscopic details of the peptide behavior. For both peptides, we generated long MD trajectories and analyzed the system structure and dynamics. Especially useful were the kinetic network analysis using optimal dimensionality reduction and dynamic element analysis of the peptide reorientations. In MD, NATA populates helix, β, PPII, and turn backbone

conformations, as seen experimentally in CD. The dipeptide explores the four-dimensional space of (ϕ, ψ, χ_1, χ_2) on a 1 ns time scale. Somewhat surprisingly, the slowest motions appear to be driven by sidechain dihedral transitions coupled to backbone motions. Overall, indole sidechain reorientations exhibit a correlation time of ca. 30 ps, close to the experimental FAD value. This corresponds to a rotational diffusion time about 100 ps. Thus, even for this small system, the reorientations are much faster than the slowest conformational transitions. Dynamic element analysis showed that, for NATA, the overall rotational diffusion could be well modeled by the population-weighted average of conformer contributions, or even by the most populated state by itself. For WK5, sampled backbone conformations included extended (β and PPII), turn, and helix, in accord with CD experiments. Conformational dynamics of WK5 occurred on a timescale of ca. 10 ns, with the slowest process involving folding from mostly extended forms to a 3_{10}-type helix, with faster transitions involving interconversions between different unfolded forms. Indole sidechain reorientations had a correlation time of ca. 100 ps, similar to the experimental FAD value and about three times longer than that of the smaller NATA. Thus, WK5 rotational diffusion occurs on a timescale of ca. 300 ps, much faster than its global folding. Dynamic element analysis for WK5 showed that reorientation times for the dihedral-restrained conformers tend to be longer than those for the free peptide. This indicates that conformational transitions are important for determining the reorientation rate in the flexible WK5 peptide. In contrast to NATA, the FAD signal of WK5 cannot be well represented by average contributions from conformers, which lack the conformational flexibility exhibited in the free MD.

Our work indicates that even for such simple systems as the NATA dipeptide and WK5 pentapeptide, we can detect interesting conformational and kinetic effects, with sampling of multiple structures over a wide range of lifetimes. Especially interesting was the finding, in both experiment and modeling, of seeds of the ordered structure, as well as the coupling of internal dynamics to global reorientation. Both structural and dynamical observables were well reproduced by the computer models. Additionally, simulations yielded valuable atomistic insights. A combined study with experimental measurements and an all-atom molecular dynamics simulation with the application of sophisticated theory thus provides an atomically detailed picture of structural properties, most favorable conformations, dynamics, and interactions with the environment of biomolecular systems that would otherwise be difficult to achieve.

SUPPLEMENTARY MATERIAL

See the supplementary material for additional figures and tables describing details of the simulation.

ACKNOWLEDGMENTS

REFERENCES

[1] R. F. Service, "Problem solved* (*sort of)," Science **321**, 784–786 (2008).

[2] P. G. Wolynes and W. A. Eaton, "The physics of protein folding," Phys. World **12**, 39–44 (1999).

[3] D. Voet, J. G. Voet, and C. W. Pratt, *Fundamentals of Biochemistry* (Wiley, New York, 1999), p. 154.

[4] N. Go, "Theoretical studies of protein folding," Annu. Rev. Biophys. Bioeng. **12**, 183–210 (1983).

[5] J. D. Bryngelson and P. G. Wolynes, "Spin glasses and the statistical mechanics of protein folding," Proc. Natl. Acad. Sci. U. S. A. **84**, 7524–7528 (1987).

[6] P. E. Leopold, M. Montal, and J. N. Onuchic, "Protein folding funnels: A kinetic approach to the sequence-structure relationship," Proc. Natl. Acad. Sci. U. S. A. **89**, 8721–8725 (1992).

[7] U. Mayor, N. R. Guydosh, C. M. Johnson, J. G. Grossmann, S. Sato, G. S. Jas, S. M. V. Freund, D. O. V. Alonso, V. Daggett, and A. R. Fersht, "The complete folding pathway of a protein from nanoseconds to microseconds," Nature **421**, 863–867 (2003).

[8] H. S. Chung, K. McHale, J. M. Louis, and W. A. Eaton, "Single-molecule fluorescence experiments determine protein folding transition path times," Science **335**, 981–984 (2012).

[9] A. Matouschek, J. T. Kellis, Jr., L. Serrano, and A. R. Fersht, "Mapping the transition state and pathway of protein folding by protein engineering," Nature **340**, 122–126 (1989).

[10] T. Cellmer, M. Buscaglia, E. R. Henry, J. Hofrichter, and W. A. Eaton, "Making connections between ultrafast protein folding kinetics and molecular dynamics simulations," Proc. Natl. Acad. Sci. U. S. A. **108**, 6103–6108 (2011).

[11] R. Zwanzig, "Simple model of protein folding kinetics," Proc. Natl. Acad. Sci. U. S. A. **92**, 9801–9804 (1995).

[12] R. B. Best, Y.-G. Chen, and G. Hummer, "Slow protein conformational dynamics from multiple experimental structures: The helix/sheet transition of arc repressor," Structure **13**, 1755–1763 (2005).

[13] C. Charlier, T. R. Alderson, J. M. Courtney, J. Ying, P. Anfinrud, and A. Bax, "Study of protein folding under native conditions by rapidly switching the hydrostatic pressure inside an NMR sample cell," Proc. Natl. Acad. Sci. U. S. A. **115**, E4169–E4178 (2018).

[14] G. S. Jas, W. A. Hegefeld, C. R. Middaugh, C. K. Johnson, and K. Kuczera, "Detailed microscopic unfolding pathways of an α-helix and a β-hairpin: Direct observation and molecular dynamics," J. Phys. Chem. B **118**, 7233–7246 (2014).

[15] G. S. Jas, C. R. Middaugh, and K. Kuczera, "Non-exponential kinetics and a complete folding pathway of an α-helical heteropeptide: Direct observation and comprehensive molecular dynamics," J. Phys. Chem. B **118**, 639–647 (2014).

[16] M. E. McCully, D. A. C. Beck, and V. Daggett, "Multimolecule test-tube simulations of protein unfolding and aggregation," Proc. Natl. Acad. Sci. U. S. A. **109**, 17851–17856 (2012).

[17] M. Guo, Y. Xu, and M. Gruebele, "Temperature dependence of protein folding kinetics in living cells," Proc. Natl. Acad. Sci. U. S. A. **109**, 17863–17867 (2012).

[18] J. D. Bryngelson and P. G. Wolynes, "Intermediates and barrier crossing in a random energy model (with applications to protein folding)," J. Phys. Chem. **93**, 6902–6915 (1989).

[19] D. J. Bicout and A. Szabo, "Entropic barriers, transition states, funnels, and exponential protein folding kinetics: A simple model," Protein Sci. **9**, 452–465 (2000).

[20] G. S. Jas and K. Kuczera, "Helix-coil transition courses through multiple pathways and intermediates: Fast kinetic measurements and dimensionality reduction," J. Phys. Chem. B **122**, 10806–10816 (2018).

[21] Y. In-Chul and G. Hummer, "System-size dependence of diffusion coefficients and viscosities from molecular dynamics simulations with periodic boundary conditions," J. Phys. Chem. B **108**, 15873–15879 (2004).

[22] J. J. Prompers and R. Brüschweiler, "General framework for studying the dynamics of folded and nonfolded proteins by NMR relaxation spectroscopy and MD simulation," J. Am. Chem. Soc. **124**, 4522–4534 (2002).

[23] K. Kuczera, J. Unruh, C. K. Johnson, and G. S. Jas, "Reorientations of aromatic amino acids and their side chain models: Anisotropy measurements and molecular dynamics simulations," J. Phys .Chem. A **114**, 133–142 (2010).

[24] G. S. Jas, E. C. Rentchler, A. M. Słowicka, J. R. Hermansen, C. K. Johnson, C. R. Middaugh, and K. Kuczera, "Reorientation motion and preferential interactions of a peptide in denaturants and osmolyte," J. Phys. Chem. B **120**, 3089–3099 (2016).

[25] G. Kaminski and W. L. Jorgensen, "Performance of the AMBER94, MMFF94, and OPLS-AA force fields for modeling organic liquids," J. Phys. Chem. **100**(46), 18010–18013 (1996).

[26] W. L. Jorgensen and C. Jenson, "Temperature dependence of TIP3P, SPC and TIP4P water from NPT Monte Carlo simulations: Seeking temperatures of maximum density," J. Comput. Chem. **19**, 1179–1186 (1998).

[27] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, "GROMACS 4: Algorithms for highly efficient, load-balanced and scalable molecular simulation," J. Chem. Theory Comput. **4**, 435–447 (2008).

[28] B. R. Brooks, C. L. Brooks, A. D. MacKerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch *et al.*, "CHARMM: The biomolecular simulation program," J. Comput. Chem. **30**, 1545–1614 (2009).

[29] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen, "GROMACS: A message-passing parallel molecular dynamics implementation," Comput. Phys. Commun. **91**, 43–56 (1995).

[30] W. Kabsch and C. Sander, "Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features," Biopolymers **22**, 2577–2637 (1983).

[31] D. Frishman and P. Argos, "Knowledge-based protein secondary structure assignments," Proteins **23**, 566–579 (1995).

[32] G. S. Jas, C. R. Middaugh, and K. Kuczera, "Probing selection mechanism of the most favorable conformation of a dipeptide in chaotropic and kosmotropic solution," J. Phys. Chem. B **120**, 6939–6950 (2016).

[33] M. E. Karpen, D. T. Tobias, and C. L. Brooks III, "Statistical clustering techniques for analysis of long molecular dynamics trajectories. I: Analysis of 2.2 ns trajectories of YPGDV," Biochemistry **32**, 412–420 (1993).

[34] S. Kube and M. Weber, "A coarse graining method for the identification of transition rates between molecular conformations," J. Chem. Phys. **126**, 024103 (2007).

[35] G. Hummer and A. Szabo, "Optimal dimensionality reduction of multistate kinetic and Markov state models," J. Phys. Chem. B **119**, 9029–9037 (2015).

[36] G. S. Jas and K. Kuczera, "Deprotonation of a single amino acid residue induces significant stability in an α-helical heteropeptide," J. Phys. Chem. B **122**, 11508–11518 (2018).

[37] C. R. Cantor and C. R. Schimmel, *Biophysical Chemistry, Part 2: Techniques for the Study of Biological Structure and Function*, 1st ed. (W. H. Freeman and Company, 1980).

[38] M. L. Tiffany and S. Krimm, "Effect of temperature on the circular dichroism spectra of polypeptides in the extended state," Biopolymers **11**, 2309–2316 (1972).

[39] A. Lobley, L. Whitmore, and B. A. Wallace, "DICHROWEB: An interactive website for the analysis of protein secondary structure from circular dichroism spectra," Bioinformatics **18**, 211–212 (2002).

[40] M. O. Steinhauser, "A molecular dynamics study of universal properties of polymer chains in different solvent qualities. Part I. A review of linear chain properties," J. Chem. Phys. **122**, 094901 (2005).

[41] H. T. Tran, A. Mao, and R. V. Pappu, "Role of backbone-solvent interactions in determining conformational equilibria of intrinsically disordered proteins," J. Am. Chem. Soc. **130**, 7380–7392 (2008).

[42] J. M. Beechem, E. Gratton, M. Ameloot, J. R. Knutson, and L. Brand, "The global analysis of fluorescence intensity and anisotropy decay data: Second-generation theory and programs," in *Topics in Fluorescence Spectroscopy*, edited by J. R. Lakowicz (Plenum Press, 2002), Vol. 2, pp. 241–305.