

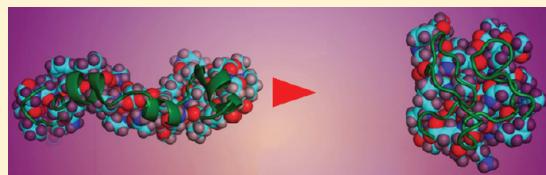
# Structural Dynamics of Free Proteins in Diffraction

Milo M. Lin, Dmitry Shorokhov, and Ahmed H. Zewail\*

Physical Biology Center for Ultrafast Science and Technology, Arthur Amos Noyes Laboratory for Chemical Physics, California Institute of Technology, Pasadena, California 91125, United States

 Supporting Information

**ABSTRACT:** Among the macromolecular patterns of biological significance, right-handed  $\alpha$ -helices are perhaps the most abundant structural motifs. Here, guided by experimental findings, we discuss both ultrafast initial steps and longer-time-scale structural dynamics of helix–coil transitions induced by a range of temperature jumps in large, isolated macromolecular ensembles of an  $\alpha$ -helical protein segment thymosin  $\beta_9$  ( $T\beta_9$ ), and elucidate the comprehensive picture of (un)folding. In continuation of an earlier theoretical work from this laboratory that utilized a simplistic structure-scrambling algorithm combined with a variety of self-avoidance thresholds to approximately model helix–coil transitions in  $T\beta_9$ , in the present contribution we focus on the actual dynamics of unfolding as obtained from massively distributed *ensemble-convergent* MD simulations which provide an unprecedented scope of information on the nature of transient macromolecular structures, and with atomic-scale spatiotemporal resolution. In addition to the use of radial distribution functions of ultrafast electron diffraction (UED) simulations in gaining an insight into the elementary steps of conformational interconversions, we also investigate the structural dynamics of the protein via the native ( $\alpha$ -helical) hydrogen bonding contact metric which is an intuitive coarse graining approach. Importantly, the decay of  $\alpha$ -helical motifs and the (globular) conformational annealing in  $T\beta_9$  occur consecutively or competitively, depending on the magnitude of temperature jump.



## 1. INTRODUCTION

Proteins are the cleaners, builders, motors, messengers, and transporters of biological cells.<sup>1</sup> The remarkable diversity of their biological functions defines the enormous complexity of their structures and dynamics, which in turn gives rise to the vast variety of length and time scales involved.<sup>2</sup> The polypeptide chain of a protein is flexible enough to adapt an exponentially large number of geometrical configurations, some of which are significantly more stable than the majority of others. Among the elements of such thermodynamically stable configurations (*native* and/or *misfolded structures* of biological significance),<sup>3</sup> right-handed  $\alpha$ -helices are, perhaps, the most abundant structural motifs. The remarkable functionality of these motifs stems from a unique pattern of hydrogen bonding that defines their spatial structure. Thus, in a right-handed  $\alpha$ -helix, the C=O group of an amino acid located at the position  $i$  in the polypeptide (backbone) chain of a protein forms a hydrogen bond with the N–H group of another amino acid that occupies the position  $i + 4$ , which results in a compact and mechanically robust molecular packing characterized by a well-defined pitch (rise per single turn of the helix) of  $\sim 5.5$  Å;<sup>4</sup> see Figure 1.

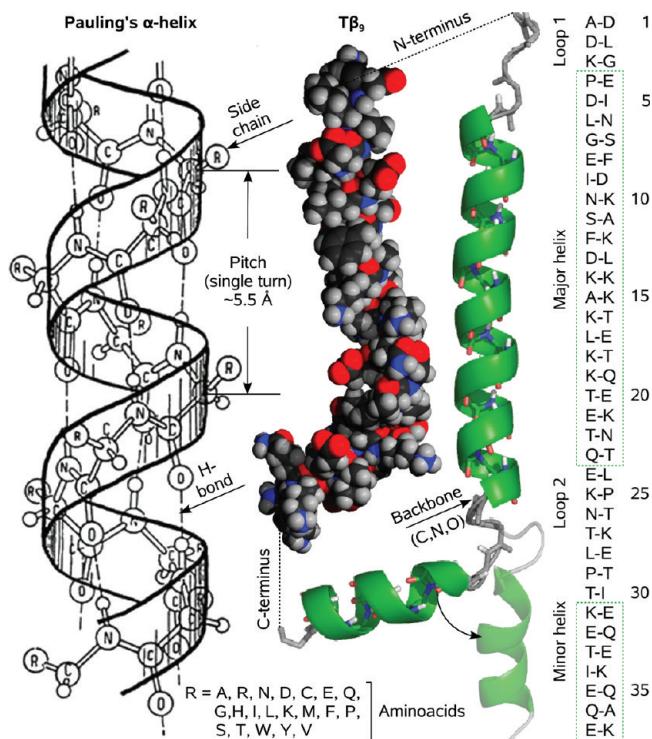
In 1951, based on the structures<sup>5</sup> of amino acids and the planar nature of the peptide bond, Pauling and Corey correctly proposed the  $\alpha$ -helix and the  $\beta$ -strand (to be the subject of another contribution in the series of works at hand) as the primary structural motifs in protein secondary structure.<sup>6</sup> The discovery of the spatial structure and chemical bonding patterns of  $\alpha$ -helices (Figure 1), which has become a cornerstone of structural

biology,<sup>7</sup> laid the groundwork for further investigations of their thermodynamic properties that are typically defined in the context of *order-disorder* (or helix–coil) transitions. Later, the actual rates of such transitions were studied both theoretically and experimentally,<sup>8</sup> measuring them on the microsecond time scale and prompting the association of the results with “ultrafast” dynamics.<sup>9</sup> It was not until recently that the (un)folding dynamics of  $\alpha$ -helices was determined with ultrafast temporal resolution.<sup>10</sup>

In these studies,<sup>10</sup> it was shown that the fastest folding events take place on the  $\sim 5$  ns time scale and that, even in the case of the shortest possible (five-residue) structures, the helix–coil transition involves multistep dynamics associated with a number of (transient) intermediate states. Importantly, the (un)folding dynamics takes place on a landscape of different pathways. Thus, the peptide may search for the correct hydrogen-bonded conformation through intermediate collapsed structures,<sup>11</sup> and different trajectories may be involved: those beginning from the unfolded peptides and forming intermediate collapsed structures prior to the helical structure, and those which directly form the helical structure from collapsed structures at shorter length and time scales. Such a bifurcation concept is a general feature of complex molecular reactions with different pathways.<sup>12</sup> A simple theoretical treatment involving the rotational diffusion of peptide bonds for the transformation from

Received: August 17, 2011

Published: October 05, 2011



**Figure 1.** Molecular structure of  $\alpha$ -helical peptides. In canonical right-handed  $\alpha$ -helices, the  $C=O$  group of an amino acid located at the position  $i$  in the backbone chain of a protein forms a hydrogen bond with the  $N-H$  group of another amino acid that occupies the position  $i + 4$ . Equivalent atomic positions recur every  $5.4 \text{ \AA}$  along the chain which defines the *pitch* of the helix. Each amino acid residue is known to advance the helix by  $1.5 \text{ \AA}$  along the helix axis which defines the *rise*. Consequently, in ideal  $\alpha$ -helices there are 3.6 amino acid residues per turn (left).<sup>4</sup> Also shown in a variety of representations is the molecular structure of the protein thymosin  $\beta_9$ , ( $T\beta_9$ ; PDB ID: 1HJ0; sequence: ADKPDLGEINSFDKAKLKKTTQEKNLPTKETIEQEKQAK) in solution of 40% (v/v) 1,1,1,3,3,3-hexafluoro-2-propanol- $d_2$  in water as obtained from 2D  $^1\text{H}$  NMR experimental data at  $T = 298 \text{ K}$ ,<sup>27</sup> 667 atoms of  $T\beta_9$ , which give rise to about 222,000 unique interatomic distances  $r_{ij}$ , are engaged into 37 individual  $i \rightarrow i + 4$  contacts ( $\alpha$ -helical and nonhelical). The ordered conformation of  $T\beta_9$  includes two  $\alpha$ -helical sequences, “major” and “minor”, which extend from contacts 4 and 31 to 23 and 37, respectively. The two helices are joined together by a poorly structured (loop) sequence (amino acids 28 to 30). Another loop is located at the N-terminus of the protein (right).

helical to nonhelical basins of the conformational space  $\{(\psi, \varphi)\}$ , as defined by a set of Ramachandran dihedral angles,<sup>13</sup> predicted folding time constants of about 2 ns at room temperature. MD simulations from this laboratory supported the multiple path scheme of folding. The picosecond (annealing) and nanosecond (diffusion) time scales as well as the ensemble-averaged helicity fractions thus obtained were wholly in agreement with those measured experimentally at the same temperature.<sup>10,13</sup>

In continuation of research on order-disorder transitions in biomolecules, which was triggered by our interest in helix-coil interconversions in polypeptides<sup>10,13,14</sup> and DNA/RNA,<sup>15</sup> and further extended to recent investigations of temperature-induced conformational changes in linear alkanes,<sup>16</sup> in what follows we report on the ultrafast unfolding dynamics of large (hundreds of trajectories) macromolecular ensembles of  $T\beta_9$  (Figure 1) in vacuo, the first effort to examine the dynamics of proteins in

diffraction. In so doing, we address a number of long-standing issues, such as the validity of the Kinetic Zipper model<sup>17</sup> and single sequence approximation of Schellman,<sup>18</sup> albeit in the absence of water, as well as elucidate representative molecular structures characteristic of isolated  $\alpha$ -helical polypeptides; conformational dynamics of  $T\beta_9$  in aqueous solutions will be the subject of another contribution. The focus will be on the different types of dynamical motions that arise at different length and time scales of the unfolding process, and how the relative time scales shift as a function of temperature. The implications of these results for experimental studies of dynamical transitions of macromolecules are succinctly highlighted.

As opposed to the proof-of-principle study from this laboratory that was based on simplistic theoretical modeling,<sup>14</sup> ensemble convergent MD simulations are used here to picture the temporal evolutions of the ensemble-averaged structures of  $T\beta_9$  induced by nonequilibrium thermal perturbations (*temperature jumps*). Notably, the size and conformational flexibility of the studied ensembles of  $T\beta_9$  represent a real challenge to both experiment and theory because the number of possible structures and degrees of freedom to be considered is huge. Therefore, a number of coarse-graining approaches, such as the ensemble-averaged radial distribution functions and radii of gyration,<sup>14–16</sup> are used to account for the transitional dynamics of the entire macromolecular ensemble without sacrificing the crucial structural resolution. The studied ensembles of macromolecules were generated using distributed MD simulations, and the ensemble convergence was ensured by monitoring the ensemble-averaged radial distribution functions, helicity fractions, and unfolding trajectories.

Atomic-scale spatiotemporal resolution combined with a massive statistical redundancy, which is sufficient to achieve the desired degree of ensemble convergence, constitute the basis for accuracy and reproducibility of the numerical experiments presented below. Because direct experimental determination of detailed molecular structures of complex, flexible biopolymers both in *vacuo* and in solution is not feasible at present, theoretical methods remain the guiding force in exploring their conformational space in various environments.<sup>10,13–16</sup> Importantly, the consistency and credibility of our results are ensured by the data redundancy and signal-to-noise ratio comparable to those typically obtained, e.g., during the course of ultrafast electron diffraction (UED)<sup>19</sup> experiments. Furthermore, in a similar way to ultrafast electron crystallography (UEC)<sup>20</sup> experiments, which utilize the spatial and temporal coherence in the specimen to reveal the structural dynamics, the “computational microscopy” experiments reported here make use of both spatial resonance and ensemble averaging (see the subsequent Section) to extract accurate and reliable structural data at each particular point in time. Accordingly, the ensemble-convergent MD simulations presented here and elsewhere<sup>10,13,15</sup> provide statistically significant information on the nature of the transient behavior of large biological macromolecules.

## 2. DIFFRACTION OF FREE MACROMOLECULES

Of special interest in molecular biology is the influence of water on macromolecular functions. Studies of structural and conformational changes which are free of the additional effects of solvation, crystallization, or external ordering imposed on the specimen have received increased attention.<sup>21,22</sup> Although biological macromolecules are sensitive to their environment, some

aspects of the structure are relatively robust, persisting in a wide variety of environments. This gives rise to a number of questions of fundamental significance; in particular: what structural features are preserved in vacuo and are therefore “inherent” to the physics of the isolated macromolecule? Equally importantly, how does the absence of solvent affect conformational dynamics? Answering these questions may elucidate the extent to which biological feasibility relies on the aqueous environment.

Studies of isolated species have been made using various experimental methods, especially mass spectrometry in which the macromolecule is kept intact.<sup>22</sup> In contrast to spectrometric and spectroscopic investigations into gas-phase behavior, UED allows for direct probing of molecular geometry. For small isolated molecules, the success of UED in determining their structures<sup>23</sup> and dynamics<sup>24</sup> has been largely aided by the advances made in computation and detector sensitivity. Because the spatial periodicity that gives rise to *coherent* electron and X-ray scattering characteristic of crystalline specimens normally vanishes for free molecules, the UED approach utilizes quantum-chemical calculations as a starting point for conformational fitting. High-level ground- and excited-state electronic structure calculations followed by normal-mode analyses of the resulting force fields minimize guesswork and result in highly accurate predictions of structures and dynamics for relatively small to moderate-sized organic molecules.

In more complex cases, such as nonequilibrium structures on shallow potentials and/or multiple, highly correlated large-amplitude intramolecular motions, a Monte Carlo sampling algorithm, coupled with the internal coordinate representation, allows the fit structure to be vastly different from the starting-point model provided by quantum-chemical calculations.<sup>19a</sup> Last but not least, using powder diffraction data,<sup>25</sup> it has been shown recently that molecular structures and conformations can be determined from UED using genetic algorithms.<sup>26</sup> These and other structure refinement techniques form the basis of the UED structural search in large conformational space guided by experiment. However, both the size and conformational flexibility of biological macromolecules render molecular structure parameters, such as bond distances, valence angles, or torsional angles, inaccessible by UED. Because all internuclear scattering terms are projected onto a one-dimensional radial distribution function, the atomic-scale resolution must be replaced by other structural features (“fingerprints”).

The remarkable periodicity of the  $\alpha$ -helix architecture gives rise to a one-dimensional structural resonance analogous to that of three-dimensional crystals. We have exploited this property in our studies. To capture this resonance feature, we utilize the (time-dependent) *ensemble-averaged radial distribution function*,  $\langle f(r, t) \rangle_n$ , where  $r \cdot \langle f(r, t) \rangle_n$  is the distribution of internuclear distances  $r_{ij}$ ,  $i \neq j$ ,  $i, j \leq N$ , across the studied macromolecular ensemble,  $N$  is the number of atoms in the macromolecule, and  $n$  is the ensemble size. At each particular point in time  $t$ ,  $\Delta \langle f(r, t_m, t_k) \rangle_n = \langle f(r, t_m) \rangle_n - \langle f(r, t_k) \rangle_n$  represents a *natural* measure of the ensemble-averaged conformational change which occurs during the time interval  $\Delta t = t_m - t_k$ . With properly chosen frame-referencing, i.e., with an adequate selection of the two points in time that correspond to the two different states of the system under study, the spatiotemporal evolution of transient structures has been extensively explored for a variety of chemical processes.<sup>24</sup> The procedure of obtaining  $\langle f(r, t) \rangle_n$  from diffraction data has been described in detail in a number of sources.<sup>14,16,19</sup> In

this work,  $\langle f(r, t) \rangle_n$  was calculated from computed structures using a modified version of the UEDANA code.<sup>19a</sup>

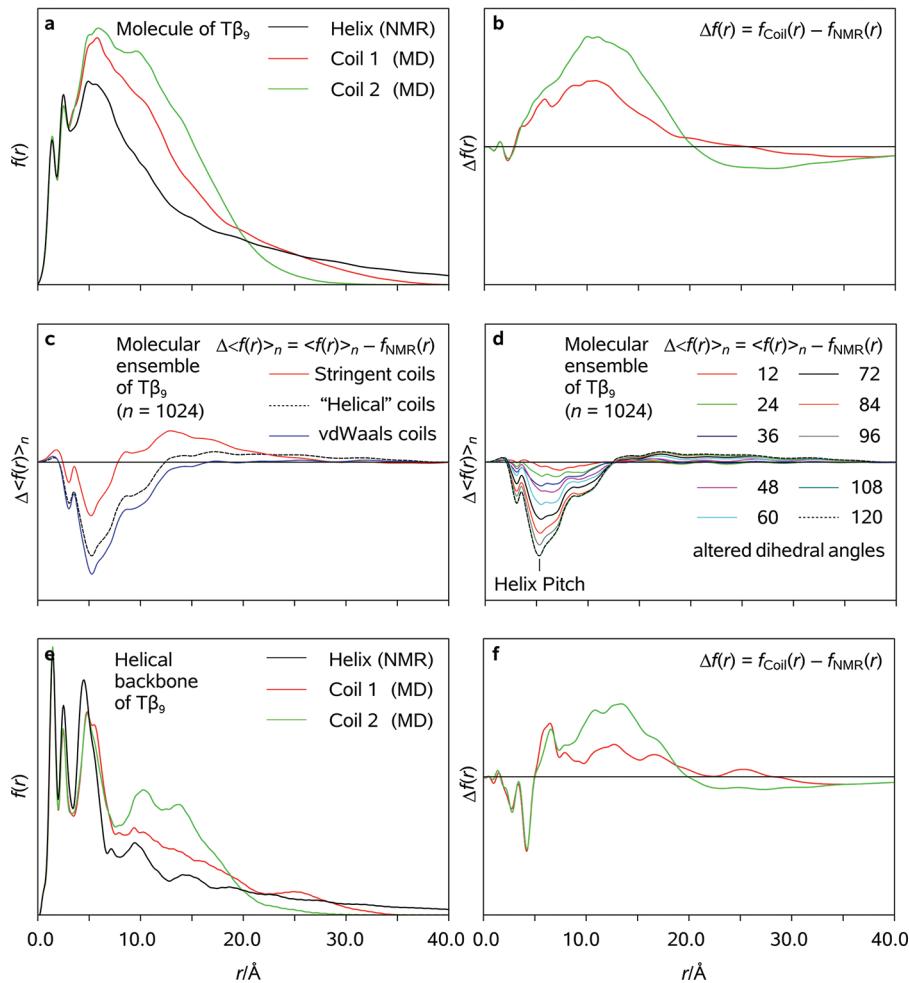
In what follows, we investigate the conformational dynamics of  $T\beta_9$  via the ensemble-averaged radial distribution functions of UED simulations and use native-contact metrics to parametrize and (locally) quantify the actual disruption of helicity. This novel approach provides a comprehensive insight into both local and globular structural changes associated with helix–coil transitions across the macromolecular ensemble under scrutiny. Although certain apparent implications for future UED experiments are briefly outlined in ref 14, the above-mentioned radial distribution functions are invoked here purely as an intuitive coarse-graining method for macromolecular order–disorder transitions in general. We note that any other definition of the ensemble-averaged radial distribution functions and their temporal differences (“diffraction differences” in the case of UED measurements), including simple pairwise internuclear distance histograms, is equally well suited to analyze the dynamics of macromolecules.

### 3. DIFFRACTION OF $T\beta_9$ : A PROOF OF PRINCIPLE

In this Section, we illustrate the above-mentioned principles with simulated  $f(r)$  patterns of isolated macromolecules of  $T\beta_9$ . The 41 amino acid residues of  $T\beta_9$  form two (“major”) and (“minor”) right-handed  $\alpha$ -helices connected by a poorly structured loop; a short sequence of nonhelical amino acid residues is also located at the N-terminus of the protein (Figure 1). The radial distribution function of  $T\beta_9$ , as obtained from its NMR structure (PDB ID: 1HJ0) found in a water–alcohol solution<sup>27</sup> is depicted in Figure 2a; 667 atoms constituting  $T\beta_9$  give rise to about 222 000 unique scattering terms  $r_{ij}$ . Importantly,  $f(r)$  of the NMR structure of  $T\beta_9$  exhibits a sharp resonant peak at  $r \approx 5.5 \text{ \AA}$ , as well as some residual periodicity of  $\sim 5.5 \text{ \AA}$ , which correspond to coherent accumulation of  $r_{ij} \approx 5.5 \cdot n \text{ \AA}$  distances ( $n = 1, 2, \dots$ ) along the two helices of the macromolecule. Also shown in the Figure are corresponding patterns of individual structures of  $T\beta_9$  chosen at random from the gas-phase macromolecular ensemble equilibrated at  $T = 300 \text{ K}$  during the course of MD simulations (see Section 4). The differences between the two kinds of structures also apparent in  $\Delta f(r)$  patterns (Figure 2b) are indicative of a pronounced change in diffraction upon evaporation of the protein. As demonstrated in Figure 3, the initial helical ordering characteristic of  $T\beta_9$  in water–alcohol solution is significantly reduced in the gas phase. In addition, the (largely unwound) protein macromolecules tend to form highly compact globular structures stabilized by non-native hydrogen bonding. Hence, the population of the outermost (“end-to-end”)  $r_{ij}$  distances characteristic of the native structure of  $T\beta_9$  is depleted whereas the number of incoherently distributed  $r_{ij}$  distances in the inner area of the pattern increases accordingly.

Earlier we reported<sup>14</sup> a simplistic theoretical modeling of helix-to-coil transitions in  $T\beta_9$ . The electron diffraction simulations were carried out as follows. First,  $n = 1024$  (partially) randomized pseudoconformers of  $T\beta_9$  were generated using its NMR structure<sup>27</sup> as a starting point (no changes in the structure parameters other than the torsional angles determining the backbone conformation were allowed). The set of torsional angles,  $\{\varphi_i\}$ , of a given pseudoconformer was readjusted at random if it did not satisfy one of the following steric constraints: (i)  $r_{ij} > 2.6 \text{ \AA}$  (“stringent” constraint); (ii)  $r_{ij} > r_{ij}(\text{native})$  as obtained from the NMR

## Radial distribution functions of helix-coil transformations: From molecules to backbones

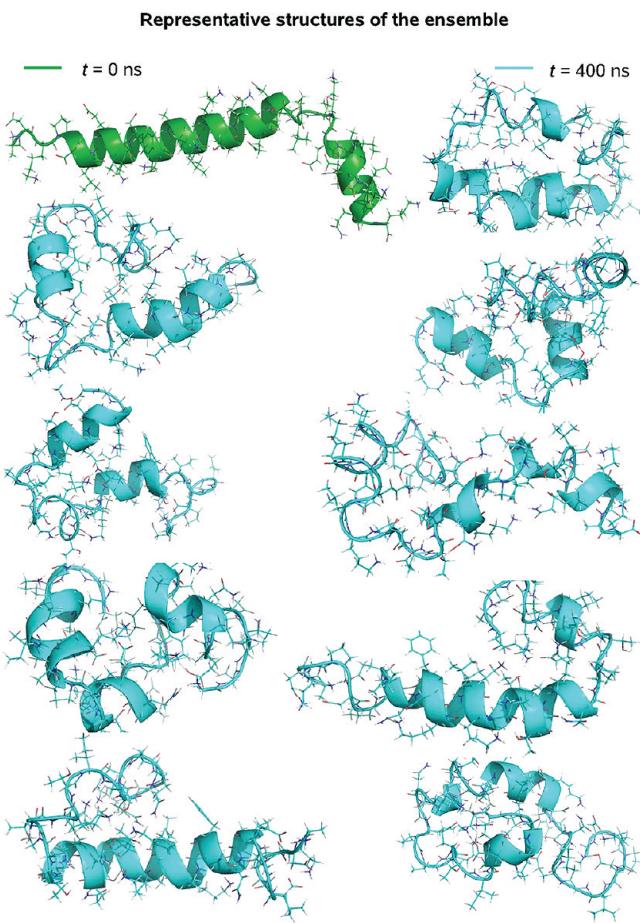


**Figure 2.** Modeling of helix-to-coil transformations: from molecules to backbones. Shown are radial distribution functions of the starting-point (NMR) structure<sup>27</sup> and two conformations of  $T\beta_9$ , chosen at random from a large ( $n = 200$ ) macromolecular ensemble equilibrated in vacuo at room temperature during the course of ensemble-convergent MD simulations ( $f(r)$ ; (a), as well as corresponding diffraction differences ( $\Delta f(r)$ ); (b) calculated with respect to the reference starting point structure. UED patterns representing {C, N, O} backbones of the same structures of  $T\beta_9$  (excluding side chains) are given by ( $f_B(r)$ ; (e) and ( $\Delta f_B(r)$ ; (f)). Ensemble-convergent diffraction differences with respect to the reference ensemble of NMR structures ( $\langle \Delta f(r) \rangle_n$ ,  $n = 2^{10}$ ) calculated for fully randomized (random coil) structures assuming various self-avoidance thresholds (c) and for partially randomized  $\alpha$ -helical structures in the assumption that  $T\beta_9$  was unwinding continuously from its C-terminus (d) are adapted from ref 14.

structure of  $T\beta_9$  ("helical" constraint), or (iii)  $r_{ij} > 3.1 \text{ \AA}$  (van der Waals-type constraint). Second, ensemble-averaged radial distribution functions  $\langle f(r, T) \rangle_n$  were calculated for the above-mentioned models at  $T = 300 \text{ K}$ . Third, and finally,  $\langle f_{\text{NMR}}(r, T) \rangle_n = f_{\text{NMR}}(r, T)$  of the native (NMR) structure of  $T\beta_9$  was calculated in a similar fashion, and further subtracted from  $\langle f(r, T) \rangle_n$  of the (partially) randomized macromolecular ensembles in order to obtain  $\Delta \langle f(r, T) \rangle_n$  characteristic of the structural change. Depicted in Figure 2c,d are diffraction difference patterns as obtained using the above simplistic approach.

From the results of Figure 2c,d it is apparent that, no matter which threshold nature imposes, there should be a clear distinction between  $\alpha$ -helical and random-coil structures of  $T\beta_9$ . Moreover, the gradual accumulation of the diffraction difference in the resonant ("pitch") area of the pattern may be used to monitor the residual helicity fraction across the ensemble at each particular point in time. However, the fact that the ensemble-averaged

helicity fraction is reduced to about 20% in the gas phase, which results in a substantial reduction of the diffraction difference signal strength, renders the diffraction data analysis significantly more difficult. A way to excise the majority of incoherently distributed scattering terms that will inevitably obscure helix-to-coil transitions in the gas phase is to perform UED simulations on the backbone atoms of  $T\beta_9$  (without side chains; see Figure 1) to obtain  $\langle f_B(r, t) \rangle_n$ . Indeed, an inspection of Figure 2e,f indicates that, when  $\langle f_B(r, t) \rangle_n$  and  $\Delta \langle f_B(r, t) \rangle_n$  are used instead of  $\langle f(r, t) \rangle_n$  and  $\Delta \langle f(r, t) \rangle_n$  (cf. Figures 2a,b), the resonant features become much sharper and more distinct, so much so, in fact, that the diffraction difference pattern predicted in ref 14 is partially recovered even for individual, randomly chosen structures of the MD ensemble. A similar approach was used in our recent computational study of helix-coil transitions in DNA, where individual phosphorus atoms of the sugar backbone were used to reveal the intimate details of coiling and strand separation.<sup>15a</sup>



**Figure 3.** Representative structures of  $T\beta_9$  as obtained from large-scale ( $n = 200$ ) ensemble-convergent MD simulations carried out in vacuo at room temperature. Shown in green is the reference NMR structure which represents the ensemble population at  $t = 0$ . Shown in blue are gaseous structures randomly chosen from the thermally equilibrated ensemble at  $t = 400$  ns. Note the globular structure formation and partial decay of  $\alpha$ -helical motifs characteristic of  $T\beta_9$  in the gas phase.

#### 4. DIFFRACTION OF $T\beta_9$ : MOLECULAR DYNAMICS SIMULATIONS

The previous section describes the proof-of-principle of using UED resonance peaks to monitor helix–coil dynamics. We note that the  $\Delta\langle f(r, T) \rangle_n$  patterns of Figure 2d were calculated with respect to an ensemble of NMR (*solution-phase*) structures of  $T\beta_9$ ,<sup>14</sup> i.e., they did not reflect the decay in helicity associated with dehydration and loss of the stabilizing alcohol buffer (Figure 3). Furthermore, the structure of the coil state was estimated using a simplistic theoretical model. In the present Section, we analyze the actual behavior of  $T\beta_9$  in gas phase by feeding the data from ensemble-convergent all-atom MD simulations into our coarse-graining methodology.

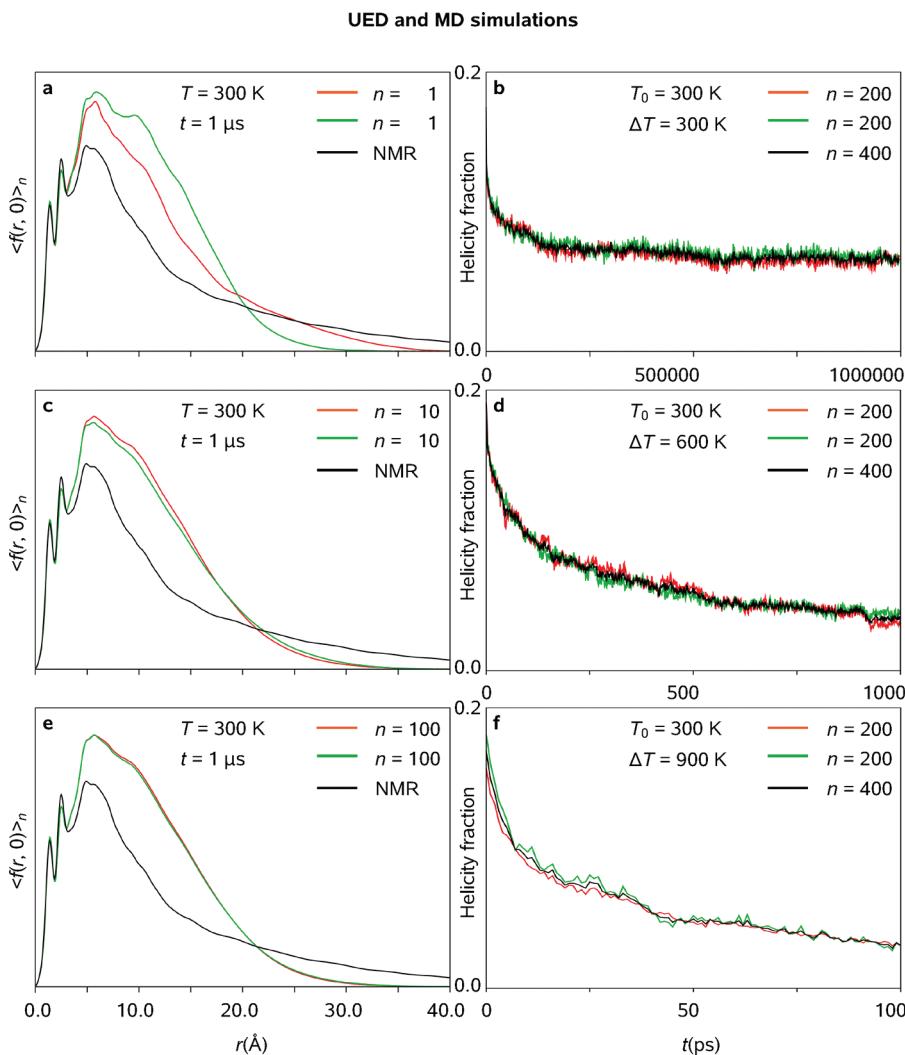
**4.1. Ensemble Convergence.** Shown in Figure 4 is the convergence behavior of the ensemble-wide radial distribution function  $\langle f(r, 0) \rangle_n$  at “time zero” ( $t = 0$ ; panels a, c, e) and the decay of the residual helicity fraction characteristic of the studied ensemble of  $T\beta_9$  (panels b, d, f) as obtained in vacuo from our MD simulations at  $T_0 = 300$  K and following the temperature jumps of  $\Delta T = 300, 600$ , and  $900$  K, respectively. The ensemble size  $n$  is indicated in each panel of the Figure. In

regard to  $\langle f(r, t) \rangle_n$ , averaging over  $n = 100$  independent MD trajectories allows for the convergence of dynamical information into a clear picture of ensemble-level behavior (Figure 4a,c,e). However, the inherent fuzziness of the ensemble-wide residual helicity profiles derived from the backbone hydrogen-bond metric precluded any reasonable assessments of the time scales involved in the unfolding transition for  $n = 100$ . Therefore, an additional set of  $n = 200$  macromolecular trajectories was calculated, and further used to obtain the results of Figure 4b,d,f. Thus, because of the fleeting nature of the hydrogen-bond contacts, the temporal profiles of residual helicity across the macromolecular ensemble of  $T\beta_9$  are at least twice as sensitive to the (random) noise level in ensemble-convergent MD simulations as those of  $\langle f(r, t) \rangle_n$ . Even when the ensemble size was increased to  $n = 400$ , the noise level remained significant, but, fortunately, the signal-to-noise ratio improved to an extent sufficient to permit reliable determination of features pertinent to the time scales involved in the studied structural transformations.

**4.2. Gas Phase Equilibration at Room Temperature.** As outlined above, an inspection of UED patterns characteristic of isolated macromolecular ensembles of  $T\beta_9$  (Figure 4a,c,e) reveals that the spatial resonance associated with  $\alpha$ -helical ordering in the protein is somewhat weakened upon vaporization, which may be rationalized in terms of formation of highly ordered ( $\alpha$ -helical) local domains separated by distorted, poorly structured moieties such as loops (cf. Figure 3). Our MD simulations indicate that these structures are *transient* in nature. Despite nonnative intramolecular hydrogen-bonding contacts that arise at  $300 \leq T \leq 600$  K in the absence of a hydrating water shell, which is known to efficiently compete with intramolecular hydrogen bonding in aqueous solutions, the conformational dynamics of nonhelical moieties remains significant not only due to their inherent flexibility, but also because of the continuous rupture and recombination of the newly formed hydrogen bonds. The excessive intramolecular hydrogen bonding and entropy-driven contraction of highly extended macromolecular conformations lead to the dominance of more compact (“globular”) structures across the ensemble, which are characterized by dramatically reduced radii of gyration.<sup>14</sup> However, despite the transition from “canonical” (NMR-like) to partially randomized, somewhat more globular macromolecular structures which takes place upon vaporization of  $T\beta_9$ , significant helicity is preserved in the gas phase (see, for example, Figure 2e). Thus, regardless of the degree of hydration, multiple sets of spatially coherent intramolecular distances (“scattering terms”) associated with robust  $\alpha$ -helical motifs give rise to unique resonant features in both  $\langle f(r, t) \rangle_n$  and  $\langle f_B(r, t) \rangle_n$  as obtained at “time zero” ( $t \equiv 0$ ) for macromolecular ensembles of  $T\beta_9$  equilibrated at room temperature ( $T_0 = 300$  K).

**4.3. Temperature Jump Dynamics and Temperature Dependence.** As evidenced from the results of Figure 5 and, more importantly, Figure 6, the temporal decay of spatial resonance in  $\langle f(r, t) \rangle_n$  and  $\langle f_B(r, t) \rangle_n$  that is observed following a temperature jump ( $\Delta T > 0$ ;  $t > 0$ ) provides direct insight into the details of order–disorder transitions throughout the studied macromolecular ensemble of  $T\beta_9$ . In what follows we monitor the temporal variation of the sharp helix resonance of  $\langle \Delta f_B(r, t) \rangle_n$  centered at the pitch distance ( $r \approx 5.5$  Å) to study the helix-to-coil transition across the ensemble (Figures 6b,d,f) for a range of temperature jumps.

For  $\Delta T = 900$  K, the (local) helix “unzipping” and (global) coiling of  $T\beta_9$  in the gas phase include multiple stages, each of which is characterized by its own specific timing. Thus, the native



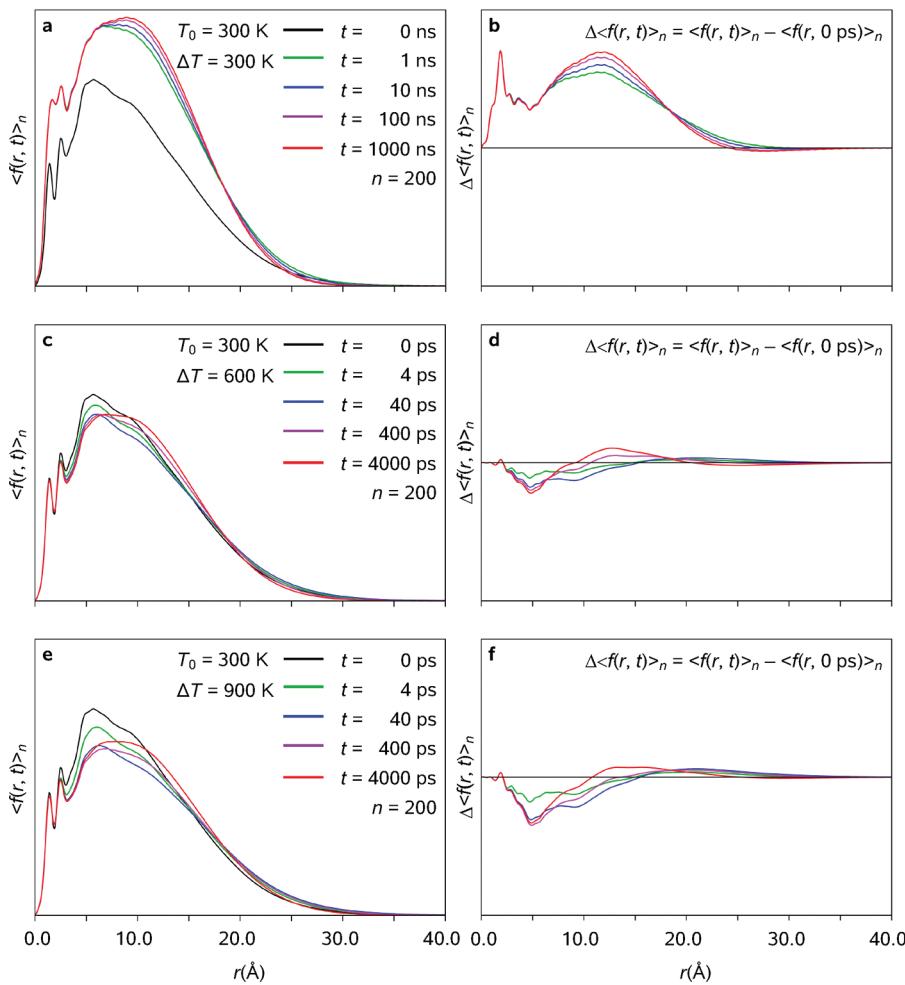
**Figure 4.** Assessment of ensemble convergence in UED and MD simulations. Shown are radial distribution functions ( $\langle f(r, 0) \rangle_n$ ; a, c, e) that were averaged over  $n = 1, 10$ , and  $100$  individual conformations constituting two equally large macromolecular ensembles of  $T\beta_9$  as obtained using large-scale ( $n = 200$ ) microsecond ensemble-convergent MD simulations carried out in vacuo at room temperature. Importantly, the ensemble size of  $n = 100$  suffices to picture convergent behavior across the ensemble;  $\langle f(r, t) \rangle_n$ ,  $n = 1, 10$ , and  $100$ , of NMR structures of  $T\beta_9$  in aqueous solution<sup>27</sup> are plotted for comparison. Also presented are temporal variations of the ensemble-averaged residual helicity fraction,  $\langle h(t) \rangle_n$ , as obtained using large ( $n = 400$ ) macromolecular ensemble equilibrated in vacuo at room temperature and further subjected to  $300$  K (b),  $600$  K (d), and  $900$  K (f) temperature jumps during the course of ensemble-convergent MD simulations. Note that the size of the macromolecular ensemble under study had to be increased by a factor of 2 in order to obtain a signal-to-noise ratio suitable for time scale determination from the temporal profiles of  $\langle h(t) \rangle_n$ . In each panel of the Figure, red and green lines represent equally large, independent trajectory sets; black lines in panels (b,d,f) represent ensemble averaging over the two sets.

( $\alpha$ -helical) hydrogen bonding contacts are broken “dynamically” on an ultrafast time scale; i.e., the hydrogen-bond rupture starts to prevail over the bond recombination throughout the entire ensemble of  $T\beta_9$  within a few picoseconds. As a result, the inherent stiffness of  $\alpha$ -helices that constitutes the basis for their remarkable mechanical properties<sup>28</sup> weakens, and the overall structure of the helices becomes more *diffuse* (hence the gradual broadening, and displacement toward higher values of  $r_{ij}$  of the characteristic helicity fingerprints in UED patterns). Notably, the overall shape of  $\langle f(r, t) \rangle_n$  and  $\langle f_B(r, t) \rangle_n$  patterns remains intact for at least  $40$  ps (Figures 5e,f and 6e,f). However, by  $t = 400$  ps, the deterioration of  $\alpha$ -helical motifs in  $T\beta_9$  becomes overwhelming, which is evidenced by the virtual disappearance of any resonant features in both  $\langle f(r, t) \rangle_n$  and  $\langle f_B(r, t) \rangle_n$ . On longer time scales, the molecule undergoes global conformational diffusion such that by

$t = 4$  ns the formation of a globular coil is virtually accomplished, and further (nonresonant) transient structural changes average out across the studied ensemble. Thus, the high temperature jump induces helix unzipping on the picosecond time scale, whereas the global structure of the macromolecule changes 2 orders of magnitude more slowly; the overall shape can be considered “frozen” during local unfolding.

It is, perhaps, instructive to emphasize here that when the tight atomic packing characteristic of  $\alpha$ -helices, which manifests itself through a resonant accumulation of relatively short ( $5$ – $10$  Å) interatomic distances across the macromolecular ensemble, becomes fully (or even partially) scrambled, the diffuse packing of the (collapsed) globule structure gives rise to a broad, smooth, hump-like incoherent feature centered at about  $12$  Å in  $\langle f_B(r, t) \rangle_n$  (Figure 6e; a similar effect is observed in  $\langle f(r, t) \rangle_n$  as well, but

## Helicity loss: Radial distribution functions of thymosin



**Figure 5.** Unfolding dynamics: radial distribution functions in vacuo. Shown are temporal evolutions of macromolecule-wide radial distribution functions ( $\langle f(r, t) \rangle_n$ ; a, c, e) and corresponding diffraction differences ( $\Delta \langle f(r, t) \rangle_n$ ; b, d, f) as obtained using large ( $n = 200$ ) macromolecular ensemble equilibrated in vacuo at room temperature and further subjected to 300 K (a,b), 600 K (c,d), and 900 K (e,f) temperature jumps during the course of ensemble-convergent MD simulations. Note the dramatic difference between the ensemble-averaged unfolding behaviors characteristic of lower ( $\Delta T = 300 \text{ K}$ ) and higher ( $\Delta T = 600, 900 \text{ K}$ ) temperature jumps.

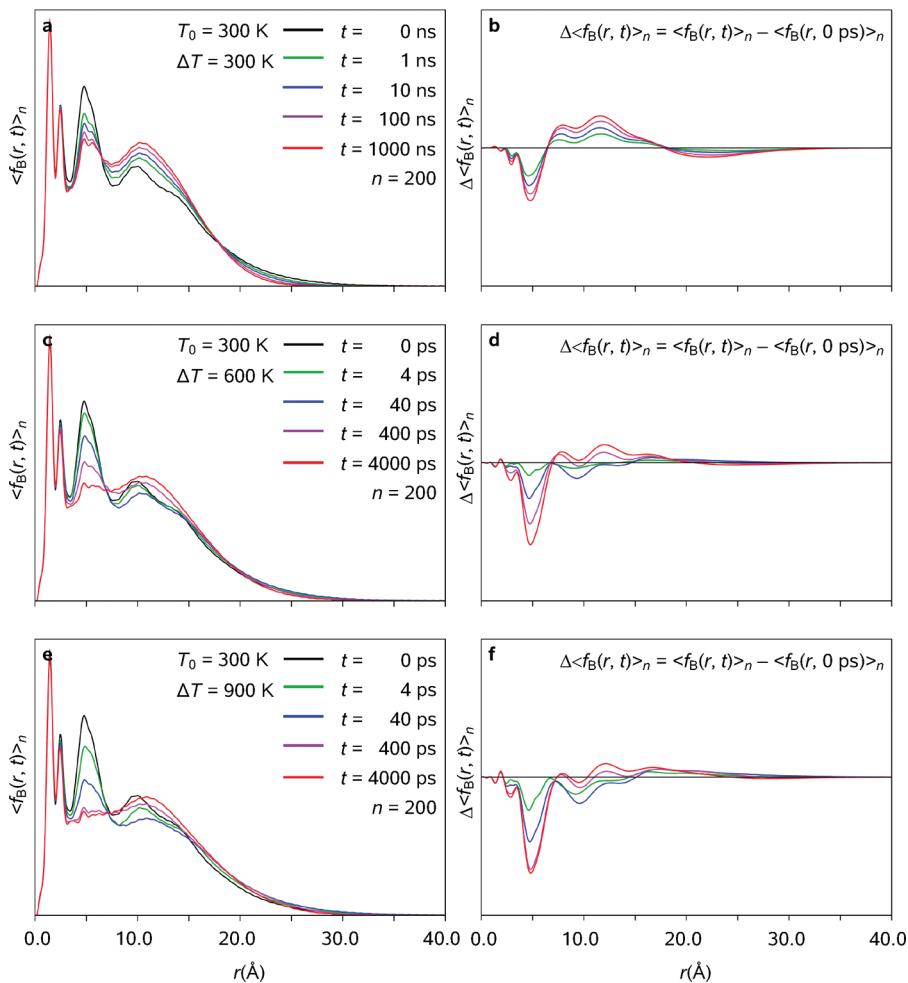
it is less pronounced at higher temperatures which preclude the formation of stable intramolecular hydrogen bonds in the gas phase, *cf.* Figure 5e). At  $\Delta T = 600 \text{ K}$  (Figure 6c), the globule formation becomes slightly more pronounced than at  $\Delta T = 900 \text{ K}$  (Figure 6e) because lower temperatures facilitate enthalpy-driven contraction of the globule, as opposed to the more extended ensembles at higher-temperatures. However, the overall pattern of structural changes as obtained for  $\Delta T = 600 \text{ K}$  (Figures 5c,d and 6c,d), despite being somewhat less dramatic, closely resembles that characteristic of  $\Delta T = 900 \text{ K}$  (Figures 5e,f and 6e,f), with the only exception of higher residual-helicity fractions being preserved in the globule at longer times.

As opposed to vibrationally hot ensembles of  $T\beta_9$  following  $\Delta T = 600 \text{ K}$  and  $\Delta T = 900 \text{ K}$  temperature jumps (see above), the ensemble-convergent behavior associated with the helix-to-coil transition triggered by  $\Delta T = 300 \text{ K}$  appears to be strikingly different insofar as the massive formation of (nonnative) intramolecular hydrogen-bonding contacts is no longer precluded by rapid vibrations and contortions of the macromolecule. The

increased number of stable nonnative hydrogen bonds manifests itself through a dramatic increase in the heights of the two innermost peaks of  $\langle f(r, t) \rangle_n$  (Figure 5a,b) which are associated with nearest-neighbor (bonded) and second-nearest-neighbor intramolecular distances (we note that corresponding peaks in  $\langle f_B(r, t) \rangle_n$  remain virtually unchanged because, unlike the {C, N, O} atoms forming side chains of  $T\beta_9$ , the atoms that constitute the backbone chain do not appear to participate in formation of the new hydrogen bonds; see Figure 6a,b).

From the overall shape of  $\langle f(r, t) \rangle_n$  (Figure 5a) one may conclude by mistake that the new hydrogen-bonded structures formed across the macromolecular ensemble must be completely random. However, an examination of  $\langle f_B(r, t) \rangle_n$  reveals something unexpected. The results of Figure 6a,b demonstrate with clarity that, despite somewhat increased contraction of the globule triggered by formation of the new, non-native hydrogen bonds, the helical-fingerprint motif in  $\langle f_B(r, t) \rangle_n$  remains virtually intact! The “stickiness” (or self-bonding bias) characteristic of  $T\beta_9$  in the absence of water induces formation of a substantial

## Helicity loss: Radial distribution functions of thymosin backbone



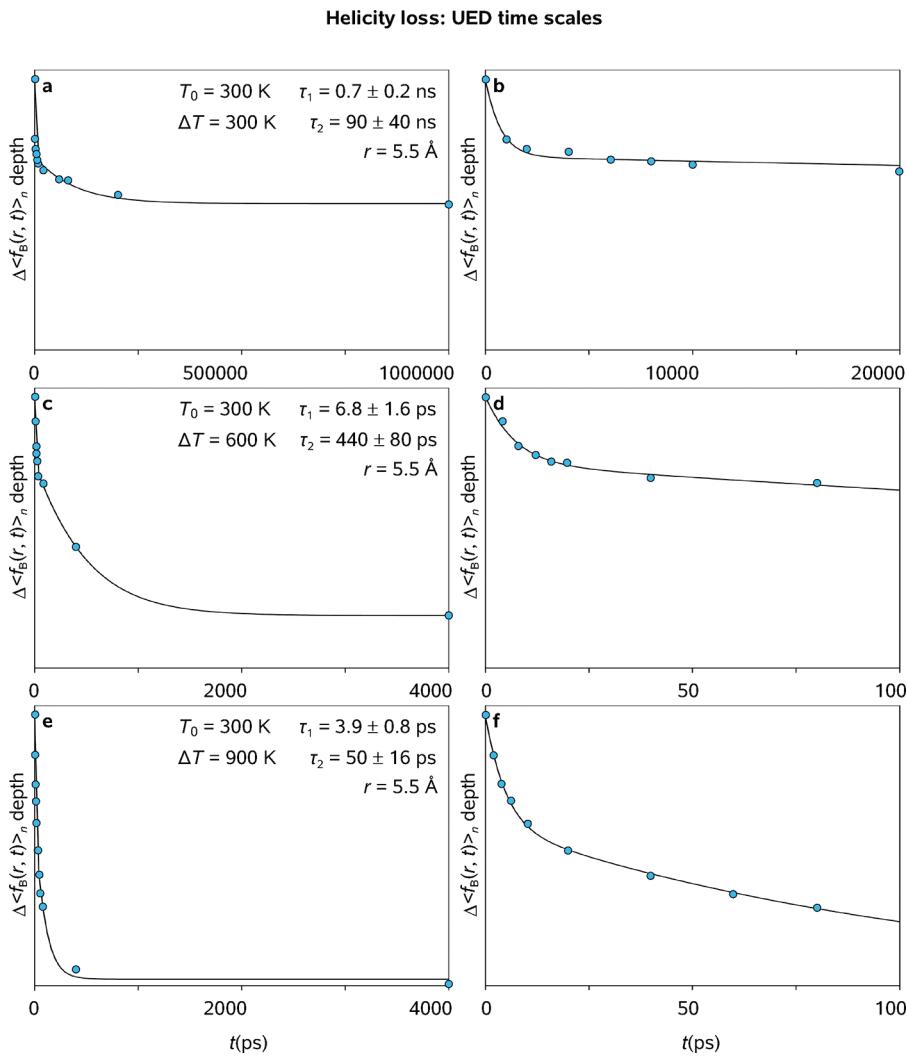
**Figure 6.** Unfolding dynamics: radial distribution functions in vacuo. Shown are temporal evolutions of backbone-specific radial distribution functions ( $\langle f_B(r, t) \rangle_n$ ; a, c, e) and corresponding diffraction differences ( $\Delta \langle f_B(r, t) \rangle_n$ ; b, d, f) as obtained using large ( $n = 200$ ) macromolecular ensemble equilibrated in vacuo at room temperature and further subjected to 300 K (a,b), 600 K (c,d), and 900 K (e,f) temperature jumps during the course of ensemble-convergent MD simulations. Note the dramatic difference between the ensemble-averaged unfolding behaviors characteristic of lower ( $\Delta T = 300 \text{ K}$ ) and higher ( $\Delta T = 600, 900 \text{ K}$ ) temperature jumps.

number of non-native, side-chain-dominated, hydrogen bonds even in the presence of fairly long  $\alpha$ -helices. The latter distort, but nevertheless retain their characteristic structural features throughout the whole simulation window, which is, indeed, remarkable. From comparing  $\langle f(r, t) \rangle_n$  with  $\langle f_B(r, t) \rangle_n$ , it is clear that global reorganization across the ensemble is fully accomplished within a few nanoseconds (Figure 5a), whereas the backbone helicity peak (Figure 6a) indicates that the helix is still unzipping on this time scale. Therefore, for  $\Delta T = 300 \text{ K}$ , local unfolding occurs at longer times than global unfolding and (non-native) refolding.

Similarly to the conformational and structural changes reported earlier for temperature-induced helix-to-coil transitions in DNA,<sup>15a</sup> the above-mentioned ensemble-convergent behavior arises because the native ( $\alpha$ -helical) hydrogen-bonding contact disruption and the global conformational change are enthalpy- and entropy-driven processes, respectively. At lower temperatures ( $\Delta T \leq 300 \text{ K}$ ), conformational diffusion happens even if few disruptions occur. Unlike DNA macromolecules, which are

highly negatively charged despite protonation,<sup>29</sup> isolated macromolecules of  $\text{T}\beta_9$  are neutral, which causes them to be even stickier in the absence of water. However, a temperature increase across the ensemble can unravel the intramolecular motions that facilitate formation of a globular coil stabilized by non-native hydrogen bonds. Since the barrier-crossing-type hydrogen-bond disruptions speed up exponentially with increasing temperature while the diffusive conformational interconversions are only weakly dependent on temperature, at a certain threshold temperature (on the interval  $300 \leq \Delta T \leq 600 \text{ K}$ ), the time scales for these two kinds of processes cross and the  $\alpha$ -helicity disappears prior to the characteristic diffusion time.

The results suggest that it is important to perform UED temperature jumps at  $\Delta T \gg 300 \text{ K}$  to prevent global dynamics from convoluting local helix unzipping in the  $\langle f(r, t) \rangle_n$ . It is also noteworthy that the admixture of residual helicity cannot be reliably detected using  $\langle f(r, t) \rangle_n$  alone because the spatial resonance associated with it is simply not strong enough to compete with incoherently distributed scattering terms associated with side



**Figure 7.** Unfolding dynamics: UED assessment of time constants in vacuo. Shown are temporal evolutions of resonant diffraction difference features ( $\langle \Delta f_B(r, t) \rangle_n, r \approx 5.5 \text{ \AA}$ ) as obtained using large ( $n = 200$ ) macromolecular ensemble equilibrated in vacuo at room temperature and further subjected to 300 K (a,b), 600 K (c,d), and 900 K (e,f) temperature jumps during the course of ensemble-convergent MD simulations. Note the dramatic difference between the ensemble-averaged unfolding behaviors characteristic of lower ( $\Delta T = 300 \text{ K}$ ) and higher ( $\Delta T = 600, 900 \text{ K}$ ) temperature jumps. Also, note the striking differences between the bond breaking ( $\tau_1$ ) and the helicity loss ( $\tau_2$ ) time scales.

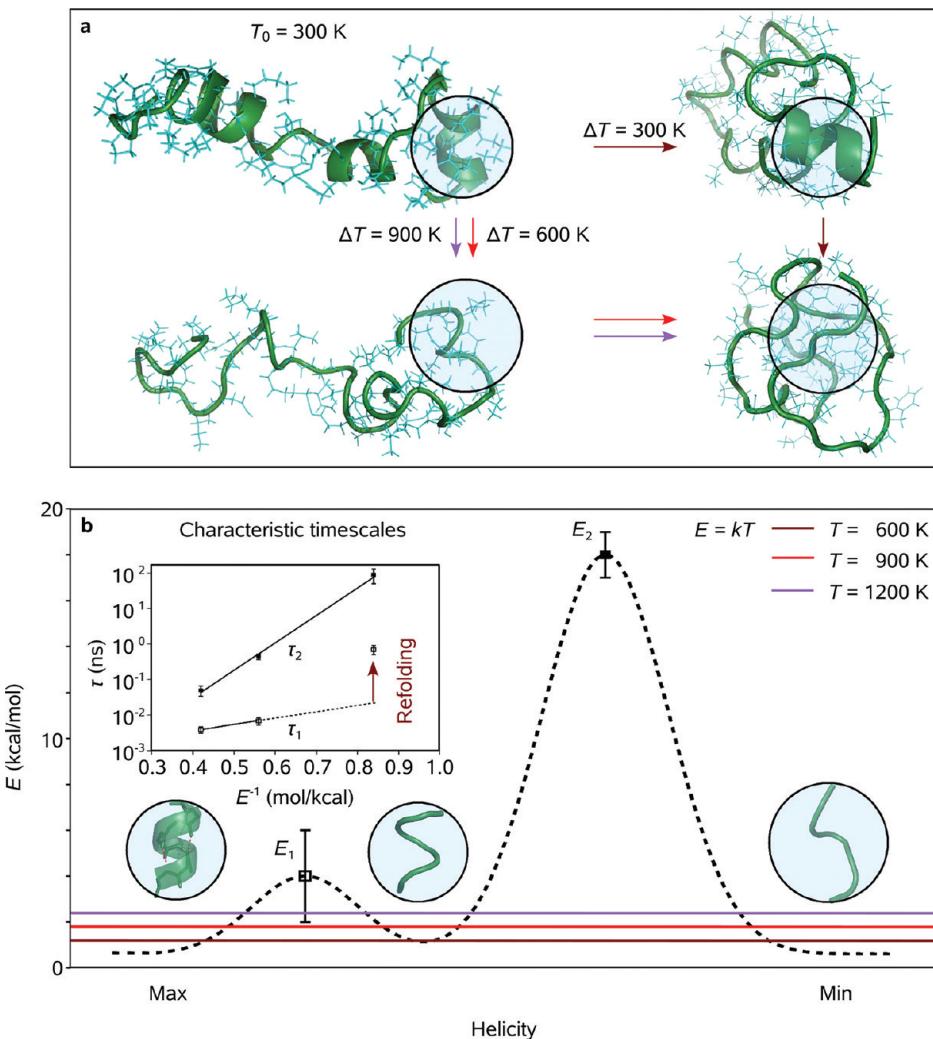
chains. Because backbone-related diffraction differences have proven to be sensitive to the subtle local (and global) structural changes that constitute elementary steps of order-disorder transitions,  $\langle f_B(r, t) \rangle_n$  and its varieties constitute a novel coarse graining approach which is significant for conformational analysis of such transitions in biomolecules (see also ref 15a).

**4.4. Rate Constants of the Helix-to-Coil Transition.** What are the actual time scales characteristic of the processes involved? First, we note that the heights of the diffuse (incoherent) nonresonant feature at  $\sim 12 \text{ \AA}$  and the “pitch” resonance peak at  $\sim 5.5 \text{ \AA}$  increase and decrease almost *quasilinearly* with the logarithmic time steps of Figures 5 and 6, which is indicative of the *exponential* rise and decay at long and short time scales, respectively. As explained earlier in this Section, the former variation is associated with the incoherent elimination of (residual) helical motifs, whereas the latter one is due to coherent rupture of native hydrogen-bonding contacts along  $\alpha$ -helices. The disruptions in *quasilinear*ity that are apparent in Figure 6 may be caused by an interplay of the above two processes

(e.g., at intermediate times) and/or by a statistically significant admixture of a third process, such as spontaneous non-native hydrogen bond formation that takes place during the first several nanoseconds at  $T = 600 \text{ K}$  (Figure 6a,b). Second, we note that unlike  $\langle f_B(r, t) \rangle_n$  that exhibits shoulders etc. in the nearest vicinity of the pitch-related resonance on intermediate time scales, a single, smooth, well-defined feature characteristic of  $\langle \Delta f_B(r, t) \rangle_n$  in that area provides a much cleaner, and numerically more accurate, measure of helicity change across the ensemble.

The results of biexponential fitting based on as many as about a dozen time-dependent profiles of  $\langle \Delta f_B(r) \rangle_n$  are presented in Figure 7, which provides a numerical quantification of the actual time constants involved in the temperature-induced helix-to-coil transitions reported here. Importantly, helix unzipping involves two distinct phases, each characterized by its own time scale. Rupture of the native hydrogen-bonding contacts, which becomes an ultrafast process above a certain temperature threshold (see above), is characterized by  $\tau_{300} = 0.7 \pm 0.2 \text{ ns}$ ,  $\tau_{600} =$

## Helicity loss: Time scales and energetics of structural interconversions in thymosin

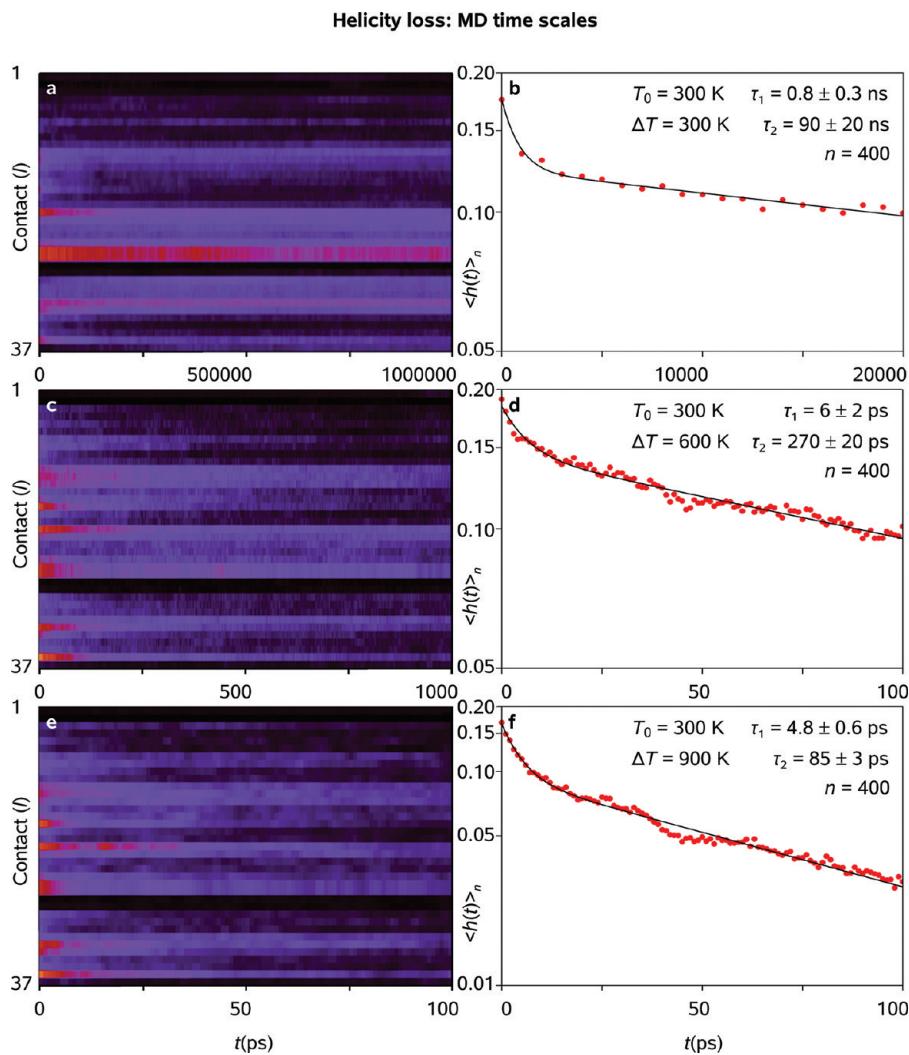


**Figure 8.** Global and local structural reorganization: following the temperature jump, the initial ensemble undergoes both local denaturation of helical motifs (blue circular highlight) as well as global coiling. For  $\Delta T = 300\text{ K}$ , the unfolding pathway (dark red arrows) involves global coiling and enthalpy-driven contraction followed by loss of local helicity, whereas for  $\Delta T = 600$  and  $900\text{ K}$  (light red and magenta arrows) the order of these two processes is reversed (a). The loss of local helical structure is further separated into two distinct time scales corresponding to fast bond disruption ( $\tau_1$ ) and slower loss of structural helicity ( $\tau_2$ ) which are separated by up to 2 orders of magnitude in their time scales (b inset). The unraveling of the helix therefore involves surmounting two barriers,  $E_1$  and  $E_2$  which are separated by the kinetic intermediate structure (b). Note that at  $\Delta T = 300\text{ K}$ , bond disruption is reversible, which significantly slows the dynamics (red arrow in the inset).

$6.8 \pm 1.6\text{ ps}$  and  $\tau_{900} = 3.9 \pm 0.8\text{ ps}$ , as assessed for the temperature jumps of  $\Delta T = 300, 600$ , and  $900\text{ K}$ , respectively, using UED data obtained from the corresponding numerical experiments. Equally importantly, the loss of residual helical structure is characterized by  $\tau_{300} = 90 \pm 40\text{ ns}$ ,  $\tau_{600} = 440 \pm 80\text{ ps}$ , and  $\tau_{900} = 50 \pm 16\text{ ps}$ . Figure 8b inset shows the (log-scale) unfolding times of these two processes as a function of inverse temperature, for which the slopes of the lines give the energy barriers. The slow ( $\tau_2$ ) helicity loss process displays Arrhenius behavior over the entire temperature range with a barrier of  $18 \pm 1\text{ kcal/mol}$ . In contrast, the bond disruption process ( $\tau_1$ ), which is up to 2 orders of magnitude faster, involves crossing a  $4 \pm 2\text{ kcal/mol}$  barrier as evidenced from the  $\Delta T = 600$  and  $900\text{ K}$  time scales. We note that  $\tau_1$  at  $\Delta T = 300\text{ K}$  is significantly longer than that predicted for barrier crossing because, unlike the irreversible bond breaking process at higher

temperatures, bond disruption at this temperature is a dynamical interplay of bond breaking and reformation on the nanosecond time scale.

The barrier of structural helicity loss, which is much higher than that of bond disruption, indicates that the helical structural motif is much more persistent than its bond energetics would suggest. As such, helix unfolding is a 3-state process in which the helical-but-unbound state is a well-defined ensemble of structures which, depending on the temperature, may be a long-lived state during the unfolding dynamics. In fact, at  $\Delta T = 300\text{ K}$  this local structural motif is much longer-lived than the overall global structural relaxation. This paradoxical result arises because global changes in the shape require only small changes in the backbone degrees of freedom that can avoid crossing the highest steric and bond-breaking barriers. As such, the global reorganization rate is entropy-driven and much less sensitive to temperature than the



**Figure 9.** Unfolding dynamics: MD assessment of time constants in vacuo. Shown are temporal evolutions of the residual helicity of 37 individual  $i \rightarrow i + 4$   $\alpha$ -helical and nonhelical contacts in the backbone chain of  $T\beta_9$ ,  $\langle h_i(t) \rangle_n$  (dark violet: 0% helicity, light orange: 60% helicity; cf. the temporal profiles of the ensemble-averaged residual helicity fraction,  $\langle h(t) \rangle_n = \sum_i \langle h_i(t) \rangle_n$ , of Figures 4b,d,e) as obtained using large ( $n = 400$ ) macromolecular ensemble equilibrated in vacuo at room temperature and further subjected to 300 K (a), 600 K (c), and 900 K (e) temperature jumps during the course of ensemble-convergent MD simulations. Relevant time scales given in panels (b,d,f) are in good agreement with those obtained from UED simulations (see Text and Figure 7). Also, note the role of proline residues (P; contacts 4 and 29 in Figure 1) in hindering the formation of up to 4 upstream  $\alpha$ -helical contacts in  $T\beta_9$ ; see Text and panels (a,c,e).

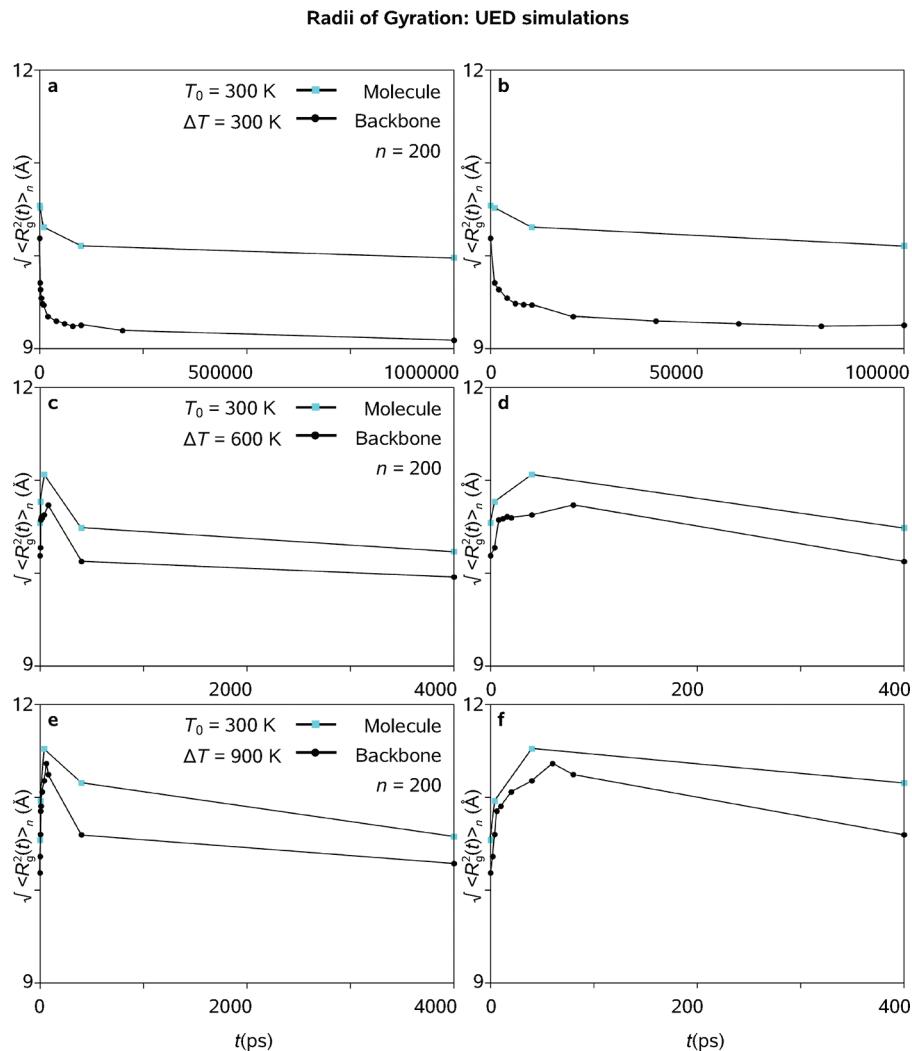
enthalpy-driven local unzipping rate which decreases exponentially with increasing inverse temperature.

**4.5. Native-Contact Separation: Backbone Hydrogen-Bonding Dynamics.** To independently assess the validity of findings made using sharp resonant features in  $\langle \Delta f_B(r) \rangle_n$  as well as to obtain sequence-sensitive dynamics of native-contact rupture in gaseous  $T\beta_9$ , we constructed temporal contact maps of the native ( $\alpha$ -helical) backbone hydrogen bonds using the atomic-resolution Cartesian coordinate sets generated from our MD simulations. For this analysis a larger ( $n = 400$ ) ensemble was required to achieve a satisfactory signal-to-noise ratio, i.e., ensemble convergence (we note that a fraction of this ensemble,  $n = 200$ , was used to obtain the UED patterns of Figures 5 and 6).

It is noteworthy that the first three nonhelical  $i \rightarrow i + 4$  amino acid contacts in  $T\beta_9$ , A-D, D-L, and K-G, give rise to a poorly defined *loop 1*, as evidenced from the NMR molecular structure of the protein found experimentally in a water–alcohol solution

(Figure 1). An inspection of Figure 9a,c,e reveals that this is also true in vacuo. Moreover, the second randomly structured moiety, *loop 2*, that has the LPT sequence and joins together the two extended helices in the NMR structure of  $T\beta_9$ , is also preserved in the gas phase throughout the MD simulation window. This behavior is due to the two proline residues, P, participating in the amino-acid contacts 4 and 29 (Figure 1). The proline residues lack NH groups and sterically block additional (up to four) upstream  $\alpha$ -helical hydrogen-bonding contacts, which renders proline “helix-breaking”. As a result, the two (“major” and “minor”) amino acid sequences of  $T\beta_9$ , that are known to form  $\alpha$ -helices in solution are readily isolated in the results of Figure 9a,c,e as areas exhibiting temporal variations.

It has long been assumed that  $\alpha$ -helices tend to unwind from the ends.<sup>18</sup> Operating from the Lifson–Roig model,<sup>30</sup> which is a refinement of the Zimm–Bragg model that recognizes that an  $\alpha$ -helix is stabilized by a hydrogen bond once three consecutive



**Figure 10.** Unfolding dynamics: radii of gyration in vacuo. Shown are temporal profiles of the ensemble-averaged, root-mean-square radii of gyration,  $\langle\langle R_g^2(t)\rangle_n\rangle^{1/2}$  (see Text), as obtained using UED data for a large ( $n = 200$ ) macromolecular ensemble equilibrated in vacuo at room temperature and further subjected to 300 K (a,b), 600 K (c,d), and 900 K (e,f) temperature jumps during the course of ensemble-convergent MD simulations. Importantly, the  $\langle\langle R_g^2(t)\rangle_n\rangle^{1/2}$  profiles of  $T\beta_9$  and its backbone exhibit the same trend, but it is somewhat more pronounced for the latter. Note the dramatic difference between the ensemble-averaged unfolding behaviors characteristic of lower ( $\Delta T = 300$  K) and higher ( $\Delta T = 600, 900$  K) temperature jumps.

residue torsional angles have adopted the helical conformation,<sup>31</sup> Poland and Scheraga generated all possible initial reactions and found that the reactions of helix ends dominated the initial rate.<sup>32</sup> However, the actual direction of unwinding (i.e., from the N versus the C terminus) will largely depend on the  $\alpha$ -helix propensities of the individual contacts involved. In the results of Figure 9a,c,e it can be readily seen that (i)  $T\beta_9$  tends to unwind from the N-terminus of the protein and (ii) certain  $\alpha$ -helical contacts are more persistent in time than the others.

For example, the E-L contact number 24 located at the very end of the major helix (Figure 1), which is not even identified as purely  $\alpha$ -helical in ref 27 because of somewhat nonstandard backbone torsional angles adopted by the NMR structure of  $T\beta_9$ , in that area, is characterized by remarkably persistent  $\alpha$ -helical hydrogen bonds at longer times, especially upon the lowest ( $\Delta T = 300$  K) temperature jump (Figure 9a). Additional helicity islands found, e.g., at positions 11–13, 16, 19, 23 along the major helix, as well as at positions 32–33 and 36 along the minor helix,

may suggest, naively, that a number of internal bubbles nucleate along the helices, e.g., at positions 18 or 35. However, we note that such features are partly a result of ensemble averaging of single helical domains at different locations. Therefore, neither a smooth pattern of increasing/decreasing  $\alpha$ -helix persistence nor a “wavy” residual-helicity pattern that involves multiple  $\alpha$ -helical islands (cf. Figure 9a,c,e) have any global physical meaning in the gas phase. Indeed, because the extended  $\alpha$ -helical moieties of  $T\beta_9$ , found in solution give rise to shorter, randomly positioned  $\alpha$ -helices separated by random loops upon evaporation of the protein (Figure 3), the results of Figure 9a,c,e can only be used to judge the ensemble-averaged residual helicity characteristic of a particular  $i \rightarrow i + 4$  hydrogen-bonding contact  $l$ ,  $1 \leq l \leq 37$ , in the backbone of  $T\beta_9$  at a given time.

In regard to the actual (ensemble-averaged) time scales involved in the helix-to-coil transition, we note that they can be readily estimated from the MD ensemble under study for each of the 37 amino-acid hydrogen-bonding contacts in the backbone

chain of  $T\beta_9$  (Figure 1), as well as for the entire macromolecule. Because the temporal evolutions of the individual  $i \rightarrow i + 4$  amino-acid contacts ( $\langle h_i(t) \rangle_n$ ; Figure 9a,c,e) cannot be compared directly with the results of UED simulations reported here, in what follows we limit the discussion of temporal features to the profiles of the ensemble-averaged residual helicity of the entire macromolecule ( $\langle h(t) \rangle_n = \sum_l \langle h_l(t) \rangle_n$ ; Figure 4b,d,f).

The actual time constants thus obtained are in good agreement with those assessed using UED methodology (the biexponential fitting procedure similar to the one reported above yields  $\tau_{300} = 0.8 \pm 0.3$  ns,  $\tau_{600} = 6 \pm 2$  ps,  $\tau_{900} = 4.8 \pm 0.6$  ps and  $\tau_{300} = 90 \pm 20$  ns,  $\tau_{600} = 270 \pm 20$  ps,  $\tau_{900} = 85 \pm 3$  ps for the faster and slower temporal components, respectively). However, it is worth re-emphasizing here that assessing the state of the hydrogen bond  $l$  connecting two ( $i$  and  $i + 4$ ) amino-acid residues in the backbone of an  $\alpha$ -helix at each particular time point, which constitutes the basis for calculating  $\langle h_l(t) \rangle_n$  and, consequently,  $\langle h(t) \rangle_n$ , is not sufficient to estimate the residual helicity across the ensemble. As we have demonstrated at the beginning of the present Section, the (initially)  $\alpha$ -helical moieties of  $T\beta_9$  tend to adopt quasi-helical torsional angles across their backbone chains long after the coherent rupture of  $\alpha$ -helical hydrogen bonds. Similarly, the hydrogen bonding can also occur despite fluctuations away from the canonical values of  $\alpha$ -helical torsional angles taking place in the backbone. Consequently, the molecular-structure-dependent  $\Delta \langle f_B(r, t) \rangle_n$  metric we used to assess the residual-helicity fraction of  $T\beta_9$  in the gas phase during the course of UED simulations provides a more accurate quantification of the temporal change in helicity across the studied macromolecular ensemble.

**4.6. Global Structures: Coherent vs Incoherent Dynamics Evidenced in the Radii of Gyration.** Another important characteristic of the ensemble that is readily obtained from the UED data is the temporal evolution of the root-mean-square radius of gyration,  $\langle R_g^2(t) \rangle_n^{1/2}$ , which measures the overall *compactness* of macromolecules averaged over the ensemble. In ref 14 we noted that from the accurate internuclear distance density distribution  $P(r)$  of a macromolecule,  $P(r)dr$  being the probability of finding an internuclear distance  $r_{ij} \in [r, r + dr]$ , one can, in principle, obtain its radius of gyration  $R_g = (I/M)^{1/2}$ , where  $M$  is the molecular mass and  $I$  is the moment of inertia. In fact, according to ref 33:

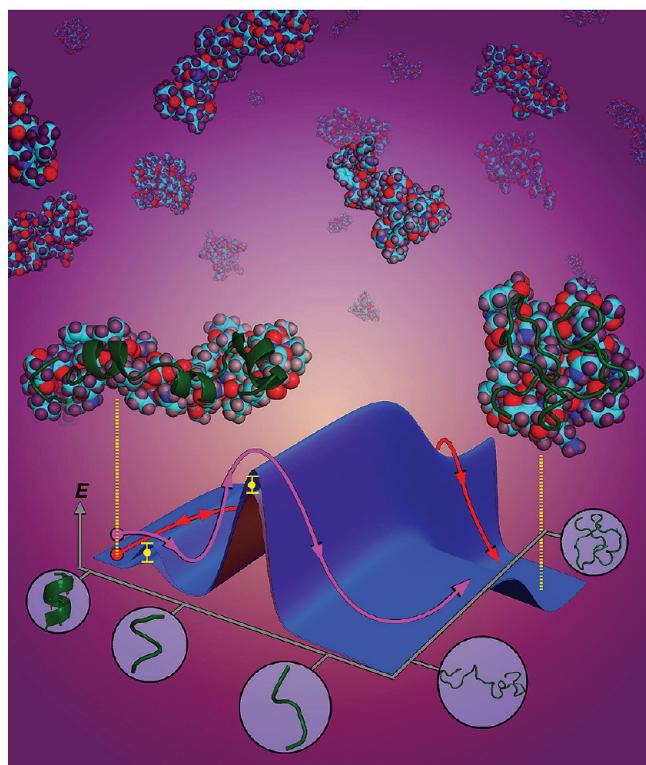
$$R_g^2 = 1/2 \int_0^\infty P(r) r^2 dr$$

and

$$\langle R_g^2(t) \rangle_n = 1/2 \int_0^\infty \langle D(r, t) \rangle_n r^3 dr$$

where normalized  $\langle D(r, t) \rangle_n = \langle P(r, t) \rangle_n / r$  is the ensemble-averaged radial distribution function calculated at time point  $t$  on the interval  $s \in [0, \infty]$  without imposing any artificial damping ( $k \equiv 0$ ) during the course of Fourier transformation.<sup>19a</sup> The empirical background correction implicit in the data refinement procedure, the finite active  $s$ -range of UED diffractometer, and the spatial resolution of the CCD detector are factors that affect the experimental determination of  $\langle R_g^2(t) \rangle_n^{1/2}$ . However, for a temporally evolving macromolecular ensemble as modeled by ensemble-convergent MD simulations,  $\langle R_g^2(t) \rangle_n^{1/2}$  of an ensemble of peptides (or their backbones) can be obtained *directly* by integrating over the radial distribution function  $\langle f(r, t) \rangle_n$  (or  $\langle f_B(r, t) \rangle_n$ ), respectively.

### Energy landscape: Unfolding dynamics of thymosin *in vacuo*



**Figure 11.** Energy landscape of unfolding dynamics. The scheme combines the local and global behavior. For lower ( $\Delta T = 300$  K) temperature jumps, the diffusion-limited cooling process during which the global structure becomes more globular precedes the loss of local helical structure (red arrow). For higher ( $\Delta T = 600, 900$  K) temperature jumps, the helicity is lost first (magenta arrow). For all temperatures, the loss of structural helicity happens after the loss of the backbone hydrogen bonds stabilizing the helix; thus, the helix-to-coil transition is a three state mechanism.

Shown in Figure 10 are the temporal-evolution profiles of  $\langle R_g^2(t) \rangle_n^{1/2}$ , as obtained from MD/UED data averaged over  $n = 200$ . A remarkable feature evident in the results of Figure 10c–f ( $\Delta T = 600, 900$  K) is the initial *expansion* of macromolecular structures in the ensemble taking place on ultrashort time scales, which is followed by a much slower macromolecular-contraction processes at longer time scales. Upon the higher temperature jumps ( $\Delta T = 600, 900$  K), the formerly  $\alpha$ -helical moieties of  $T\beta_9$  expand *coherently* in the gas phase because of the ultrafast (coherent) rupture of the  $\alpha$ -helical hydrogen bonds, which is followed by the diffusion-driven conformational smearing and the subsequent formation of compact globular structures. Notably, the ensemble-averaged behavior characteristic of  $\Delta T = 300$  K is totally different (*j.f.* Figures 10a,b): the (faster) contraction process that is accomplished within a few nanoseconds leads to the (slower) long-lived contraction dynamics at longer times. The faster contraction process is due to non-native long-range hydrogen bond formation, whereas the slower one is associated with the diffusion-driven conformational dynamics leading to a more compact globular structure. The breakage of native hydrogen bonds occurs over a time period that is long compared with the diffusive (contractive) motion, and the expansion due to bond breakage is obscured by diffusive contraction.

Interestingly, the above behavior resembles that modeled theoretically and observed experimentally in this laboratory for

the two-component (substrate–adsorbate) assemblies subjected to a laser-pulse irradiation.<sup>34</sup> The transient anisotropic change in  $c_0$  of the adsorbates (fatty-acid and phospholipid layers) described in such studies is vastly different from that observed in steady state. At equilibrium, the observed changes are in  $a_0$  and  $b_0$  (not in  $c_0$ ), and the diffraction intensity monotonically decreases, reflecting the thermal, incoherent motions (Debye–Waller effect) and phase transitions. On the ultrashort time scale, the expansion is along  $c_0$ , unlike in the thermal case, and the amplitude of change is much larger than that predicted by incoherent thermal expansion. The changes in Bragg-spot intensity and width are very different from those observed by equilibrium heating. Following the ultrafast  $T$ -jump, the structure first expands within  $\sim 10$  ps (because of atomic displacements) along the  $c$ -direction. On the nanosecond and longer time scale, the structure shrinks as it reaches the equilibrium state (incoherent movement of atoms), and the original configuration is recovered by heat diffusion on the millisecond time scale between pulses. This behavior is in contrast with that observed at steady state, as mentioned above.

## 5. CONCLUSION

There are two fundamental challenges pertinent to understanding macromolecular dynamics. The first challenge, which is *structural complexity*, is inherent to the study of any inhomogeneous structure with many degrees of freedom. The second challenge, which is *cooperative dynamics*, is a hurdle characteristic of biomolecular systems. Therefore, in order to understand macromolecular dynamics, novel coarse-graining approaches must be employed to elucidate the essential *ensemble-wide* aspects of structural change, while preserving the *mechanistic nature of the dynamics* (e.g., cooperative motion or resonance phenomena). The need for appropriate coarse-graining becomes apparent as we ascend the complexity ladder. The overwhelming complexity associated with the ensemble-wide macromolecular (un)folding behavior makes computational simulation an essential tool for understanding proteins at the molecular level. A number of coarse-graining techniques have been developed to investigate the complex energy landscapes of protein (un)folding.<sup>35</sup> For example, in the case of helix–coil transitions, disconnectivity graphs have been successfully applied to energetically sort the discrete states of native, part-native, and non-native populations that dominate the folding dynamics.<sup>36</sup>

In contrast, the approach presented here (“4D computational microscopy”), which is based on the adequate degree of coarse-graining and the ensemble-convergent MD simulations, enables studies of structural dynamics of complex systems in real time with atomic-scale spatial resolution, and with adequate statistical certainty and signal-to-noise ratio that are comparable to those of experimental observations, in this case electron diffraction. In applying the analysis tools of ultrafast electron diffraction (UED), we have found that the time-dependent ensemble-averaged radial distribution function,  $\langle f(r, t) \rangle_n$ , is a powerful method for monitoring the macromolecular dynamics across all rungs of the complexity ladder while preserving the relevant structural information. In addition to the standard metrics of, for example, percent native contact and root-mean-square deviation (rmsd) from the native structure, we have calculated  $\langle f(r, 0) \rangle_n$ , the signature for the fundamental  $\alpha$ -helical motif which possesses a unique structural resonance centered at a characteristic distance scale in diffraction. Such resonances become sharp signatures

when considering only the backbone atoms of the structures,  $\langle f_B(r, t) \rangle_n$ . Earlier we illustrated the sensitivity of the resonance to helix content by constructing ensembles of partially unwound helices of  $T\beta_9$  in the assumption of a random coil distribution.<sup>14</sup> In the present study, we confirmed the helical signature for the helix–coil transition from ensemble-averaged MD simulations of the same protein in vacuo.

In addition, we provided the actual temporal evolution of the fractional helical content as well as the time scale of helicity loss for a variety of temperature jumps. Such local dynamics were obtained simultaneously with global structural evolution by examining the large- $r$  regimes of  $\langle f(r, t) \rangle_n$  and  $\langle f_B(r, t) \rangle_n$  and calculating corresponding radii of gyration. In doing so we demonstrated that the above rudiments of UED data analysis constitute a universal coarse-graining approach that simultaneously captures *local and global structural fingerprints* in real time. Following a temperature jump high enough to preclude the formation of nonnative intramolecular hydrogen bonds, the radii of gyration were found to first increase on the ultrafast time scale of coherent hydrogen bond cleavage as the structures were expanding along the  $\alpha$ -helix axis during the initial loss of native hydrogen bonds. This was followed by gradual decrease of the radii of gyration across the ensemble as the global structure was diffusing into a compact globule stabilized by non native contacts and chain entropy.

The *interplay of enthalpic and entropic forces* could be seen in the different dependences of these two processes on temperature. For lower  $T$ -jumps ( $\Delta T = 300$  K), conformational diffusion was shown to precede any significant native hydrogen bond cleavage. However, with the increasing  $\Delta T$ , the Arrhenius bond breaking process speeds up exponentially while the diffusive motion is roughly proportional to the square root of the final temperature of the ensemble. At a certain critical final temperature the time scales for the two processes cross, and, for higher  $T$ -jumps ( $\Delta T = 600, 900$  K), the native contacts appear to break before a significant global conformational change can occur. The above behavior is evident in both the ensemble-averaged radial distribution functions and the radii of gyration, but calculation of their backbone-related temporal profiles,  $\langle f_B(r, t) \rangle_n$  and  $(\langle R_g^2(t) \rangle_n)_B^{1/2}$ , is required to obtain a clear-cut picture of the structural change. We note that a similar phenomenon was reported recently for a DNA double helix.<sup>15a</sup> In addition, unfolding of the helix was found to follow a three-state process in which the intermediate ensemble consists of structures lacking the canonical helical structure and hydrogen-bonding, but nevertheless possessing residual helicity. The barriers of this three-state mechanism were calculated from the ensemble kinetic data and the transition from the intermediate state ensemble to the coil state was found to be slower than the global structural relaxation for low temperature jumps.

The temperature dependence of the global and local dynamics is shown for representative structures in Figure 8a. The elucidation of the three-state unfolding kinetics of the local structure and the persistence of the intermediate ensemble state for longer time scales than global structural reorganization implies that the use of global order parameters such as the radius of gyration to monitor the progress of (un)folding may underestimate the unfolding time. In thinking about protein dynamics, this work challenges the notion that the largest length scales necessarily correspond to the longest time scales. These conclusions are synthesized in Figure 11. Our forthcoming paper will deal with the same protein in aqueous solution with focus on folding (unfolding) structural dynamics and the role of water in the (un)folding events.

## ■ ASSOCIATED CONTENT

**S Supporting Information.** This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

zewail@caltech.edu

## ■ ACKNOWLEDGMENT

We are grateful to the National Science Foundation for funding of this research. M.M.L. acknowledges financial support from the Krell Institute and the US Department of Energy (DoE Grant No. DE-FG02-97ER25308) for a graduate fellowship at Caltech.

## ■ REFERENCES

- (1) See, for example:(a) Lodish, H.; Berk, A.; Zipurksy, S. L.; Matsudaira, P.; Baltimore, D.; Darnell, J. *Molecular Cell Biology*, 4th ed.; Freeman: New York, 2000. (b) Lin, M. M. *Proc. DOE CSGF Annual Conf.*; Washington Court Hotel: Washington, DC, June 22–24, 2010, pp 28–29.
- (2) See, for example:(a) Dobson, C. M.; Wolynes, P. G.; Parrinello, M., in *Physical Biology: From Atoms to Medicine*; Zewail, A. H., Ed. Imperial College Press: London, 2008, pp 247–335. (b) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins* **1995**, *21*, 167–195. (c) Liu, F.; Gruebele, M. *Chem. Phys. Lett.* **2008**, *461*, 1–8.
- (d) Clementi, C. *Curr. Opin. Struct. Biol.* **2008**, *18*, 10–15.
- (3) Dobson, C. M. *Nature* **2003**, *426*, 884–890.
- (4) See, for example:Horton, H. R.; Moran, L. A.; Ochs, R. S.; Rawn, D. J.; Scrimgeour, K. G. *Principles of Biochemistry*, 3rd ed.; Prentice-Hall: Upper Saddle River, 2002.
- (5) Pauling, L. *The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry*, 3rd ed.; Cornell University Press: Ithaca, 1960.
- (6) (a) Pauling, L.; Corey, R. B. *Proc. Natl. Acad. Sci. U.S.A.* **1951**, *37*, 235–240 and references therein. (b) Pauling, L.; Corey, R. B. *Proc. Natl. Acad. Sci. U.S.A.* **1951**, *37*, 729–740.
- (7) See, for example:B. G. Malmström, Andersson, B. In *The Nobel Prize: The First 100 Years*; Wallin Levinovitz, A., Ringertz, N., Eds.; World Scientific: Singapore, 2001, pp. 73–108.
- (8) For recent reviews, see:(a) Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*; Freeman: New York, 1999. (b) Doshi, U. In *Protein Folding, Misfolding and Aggregation: Classical Themes and Novel Approaches*; Muñoz, V., Ed. RSC: Cambridge, 2008, pp. 28–48. (c) Gruebele, M. In *Protein Folding, Misfolding and Aggregation: Classical Themes and Novel Approaches*; Muñoz, V., Ed.; RSC: Cambridge, 2008; pp. 106–138.
- (d) Hamm, P.; Helbing, J.; Bredenbeck, J. *Annu. Rev. Phys. Chem.* **2008**, *59*, 291–317.
- (9) (a) Thompson, P. A.; Eaton, W. A.; Hofrichter, J. *Biochemistry* **1997**, *36*, 9200–9210. (b) Thompson, P. A.; Muñoz, V.; Jas, G. S.; Henry, E. R.; Eaton, W. A.; Hofrichter, J. *J. Phys. Chem. B* **2000**, *104*, 378–389. (c) Jas, G. S.; Eaton, W. A.; Hofrichter, J. *J. Phys. Chem. B* **2001**, *105*, 261–272.
- (10) Mohammed, O. F.; Jas, G. S.; Lin, M. M.; Zewail, A. H. *Angew. Chem., Int. Ed. Engl.* **2009**, *48*, 5628–5632.
- (11) Ma, H.; Wan, C.; Wu, A.; Zewail, A. H. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 712–716.
- (12) Zewail, A. H. *Angew. Chem., Int. Ed. Engl.* **2000**, *39*, 2586–2631.
- (13) Lin, M. M.; Mohammed, O. F.; Jas, G. S.; Zewail, A. H., *Proc. Natl. Acad. Sci. U.S.A.* doi: 10.1073/pnas.1113649108.
- (14) Lin, M. M.; Shorokhov, D.; Zewail, A. H. *Chem. Phys. Lett.* **2006**, *420*, 1–7.
- (15) See, for example:(a) Lin, M. M.; Shorokhov, D.; Zewail, A. H. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10619–10632. (b) Lin, M. M.; Meinholt, L.; Shorokhov, D.; Zewail, A. H. *Phys. Chem. Chem. Phys.* **2008**, *10*, 4227–4239.
- (16) Lin, M. M.; Shorokhov, D.; Zewail, A. H. *J. Phys. Chem. A* **2009**, *113*, 4075–4093.
- (17) See, for example:Eaton, W. A.; Muñoz, V.; Thompson, P. A.; Henry, E. R.; Hofrichter, J. *Acc. Chem. Res.* **1998**, *31*, 745–753 and references therein.
- (18) Schellman, J. A. *J. Phys. Chem.* **1958**, *62*, 1485–1494.
- (19) (a) Shorokhov, D.; Park, S. T.; Zewail, A. H. *Chem. Phys. Chem.* **2005**, *6*, 2228–2250. (b) Srinivasan, R.; Lobastov, V. A.; Ruan, C.-Y.; Zewail, A. H. *Helv. Chim. Acta* **2003**, *86*, 1763–1838.
- (20) See, for example:Yang, D.-S.; Gedik, N.; Zewail, A. H. *J. Phys. Chem. C* **2007**, *111*, 4889–4919.
- (21) See, for example:Glaeser, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 1779–1780.
- (22) See, for example:(a) Barran, P. E. *Angew. Chem., Int. Ed. Engl.* **2011**, *50*, 3120–3122. (b) Florance, H. V.; Stopford, A. P.; Kalapothakis, J. M.; McCullough, B. J.; Bretherick, A.; Barran, P. E. *Analyst* **2011**, *136*, 3446–3452. (c) Benesch, J. L. P.; Robinson, C. V. *Nature* **2009**, *462*, 576–577.
- (23) See, for example:(a) *Landolt-Börnstein Numerical Data and Functional Relationships in Science and Technology (Group II, Molecules and Radicals)*, Vol. 25: *Structure Data of Free Polyatomic Molecules*; Kuchitsu, K., Ed.; Springer: Berlin–Heidelberg, 1998. (b) *Stereochemical Applications of Gas-Phase Electron Diffraction*; Hargittai, I.; Hargittai, M., Eds.; VCH: New York, 1988.
- (24) For a comprehensive up-to-date overview, see:Shorokhov, D.; Zewail, A. H. *J. Am. Chem. Soc.* **2009**, *131*, 17998–18015, and references therein.
- (25) Harris, K. D. M.; Habershon, S.; Cheung, E. Y.; Johnston, R. L. *Z. Kristallogr.* **2004**, *219*, 838–846.
- (26) Habershon, S.; Zewail, A. H. *Chem. Phys. Chem.* **2006**, *7*, 353–362.
- (27) Stoll, R.; Voelter, W.; Holak, T. A. *Biopolymers* **1997**, *41*, 623–634.
- (28) See, for example:Idris, A.; Alam, M. T.; Ikai, A. *Protein Eng.* **2000**, *13*, 763–770.
- (29) Rueda, M.; Kalko, S. G.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2003**, *125*, 8007–8014.
- (30) Lifson, S.; Roig, A. *J. Chem. Phys.* **1961**, *34*, 1963–1974.
- (31) Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1959**, *31*, 526–535.
- (32) Poland, D.; Scheraga, H. A. *J. Chem. Phys.* **1966**, *45*, 2071–2090.
- (33) Woenckhaus, J.; Köhling, R.; Thiagarajan, P.; Littrell, K. C.; Seifert, S.; Royer, C. A.; Winter, R. *Biophys. J.* **2001**, *80*, 1518–1523.
- (34) (a) Tang, J.; Yang, D.-S.; Zewail, A. H. *J. Phys. Chem. C* **2007**, *111*, 8957–8970. (b) Chen, S.; Seidel, M. T.; Zewail, A. H. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 8854–8859. (c) Seidel, M. T.; Chen, S.; Zewail, A. H. *J. Phys. Chem. C* **2007**, *111*, 4920–4938. (d) Chen, S.; Seidel, M. T.; Zewail, A. H. *Angew. Chem., Int. Ed. Engl.* **2006**, *45*, 5154–5158.
- (35) Wales, D. J. *Energy Landscapes with Applications to Clusters, Biomolecules and Glasses*; Cambridge University Press: Cambridge, 2003.
- (36) See, for example:(a) Mortenson, P. N.; Evans, D. A.; Wales, D. J. *J. Chem. Phys.* **2002**, *117*, 1363–1376. (b) Mortenson, P. N.; Wales, D. J. *J. Chem. Phys.* **2001**, *114*, 6443–6454.