Modelling of reversible single chain polymer self-assembly: from the polymer towards the protein limit†

Denis Danilov,[a] Christopher Barner Kowollik[bc] and Wolfgang Wenzel*[a]

The thermodynamic properties of reversible single chain polymer self-assembly are characterized by all-atom simulations. The ensemble of closed chains collapses from multiple conformations for long chains to nearly unique conformations for shorter chains, suggesting that the engineered polymers can fold into stable unique conformations at moderate temperatures.

Single-chain nanoparticles, their synthesis, modelling and applications are a rapidly growing research topic in modern polymer science.[1] Modern polymerization processes such as atom transfer radical polymerization (ATRP), allow fine control over the molecular architecture, molecular weight and polydispersity of synthetic polymers.[2,3] Combined with novel modular ligation protocols, well-defined polymer chains can be constructed.[4,5] The exquisite level of control over the molecular architecture has enabled the synthesis of polymeric single chain nanoparticles where single polymer folding is driven by specially designed hydrogen donor–acceptor moieties within the lateral polymer chain.[6–10] Typically, two approaches are followed to induce single-chain folding of synthetic macromolecules: (i) the collapse of the chain through interactions of functionalities within the side groups (repeating unit folding)[11–17] or (ii) by placing recognition units at predetermined places along the polymer chain (selective point folding), which was experimentally pioneered by some of us.[6–10] *Via* selective point folding, the design of one or more orthogonal binding sites in the polymer chain permits the synthesis of simple biomimetic systems with the potential to surpass the stability of their biological

counterparts.[6–10] In addition, the study of these systems opens the exciting opportunity to shed light on the folding mechanism of proteins and DNA by comparing their characteristics with those of synthetic systems.

The availability of these emerging experimental techniques raises novel questions for polymer theory and modelling: what are the conditions on the polymer chain so that the polymer "folds" not just into a collapsed conformation, but into a single "unique" conformation? How does the folding process differ from that of proteins, where folding is driven by a multitude of interactions optimized by evolution,[18] while polymer folding is driven by a few engineered interactions? Which level of accuracy is required in terms of polydispersity and placement of the point folding units to achieve a dominant unique conformation in the "folded" ensemble?

Given the challenges in structure determination of single polymer chains, atomistic simulations, which have matured to fully characterize the folding process of small proteins,[19] may help to answer these questions. In addition, they may shorten the laborious experimental trial-and-error process to identify "well behaved" polymer architectures for single-chain folding and establish design principles. In the current study, we use all-atom simulations to fully characterize the thermodynamic equilibrium of one of the few experimental systems for which reversible single-chain point folding was demonstrated[6–10] and then change the system composition to predict at what chain length "folding" into a defined geometry of the single chain units is expected to occur.

Specifically, we investigate here the single chain self-assembling systems[9] shown in Scheme 1 carrying α,ω-donor–acceptor groups at the polymer chain ends, which may reversibly link by the formation of hydrogen bonds. For polymers of length L close to 50, reversible self-assembly was observed in the experiment as a function of the chain concentration in solution.[9] The change of the hydrodynamic diameter in the experiment was interpreted as circular self-assembly of the single chains at low concentrations, while inter-chain linkage is observed at higher concentrations. The reversible transition from single chain self-assembly to inter-chain linkage is only possible if each of the two states is in thermodynamic equilibrium with open configurations. Thus, it is believed that

[a] *Institute of Nanotechnology (INT), Karlsruhe Institute of Technology (KIT), Hermann von Helmholtz Platz 1, 76344 Eggenstein Leopoldshafen, Germany. E mail: wolfgang.wenzel@kit.edu*
[b] *Preparative Macromolecular Chemistry, Institut fur Technische Chemie und Polymerchemie, Karlsruhe Institute of Technology (KIT), Engesserstr. 18, 76131 Karlsruhe, Germany*
[c] *Institut fur Biologische Grenzflachen (IBG), Karlsruhe Institute of Technology (KIT), Hermann von Helmholtz Platz 1, 76344 Eggenstein Leopoldshafen, Germany*
† Electronic supplementary information (ESI) available: Computational methods, shape measures of polymer chain, representative closed conformations, video of transitions between closed and open states for L = 50. See
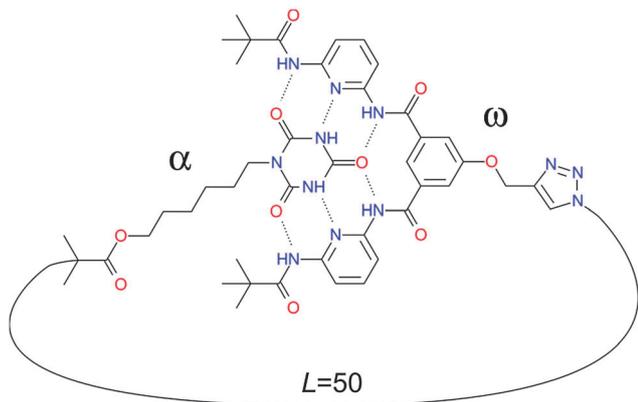
Scheme 1 Model system of a polystyrene precision designed polymer with α,ω donor acceptor for single chain self assembly.[9]

the self-assembled conformational ensemble is not static and the polymers fold and unfold reversibly, very much like peptides and proteins. Presently, little is known about the conformational ensembles of the closed and open conformations as a function of chain length.

The time-scale on which reversible polymer folding is believed to occur is far larger than those reachable with present day explicit solvent all-atom molecular dynamics simulations.[19] In order to develop a detailed thermodynamic description, we have performed very long ($O(10^9)$ steps) all-atom implicit-solvent Monte Carlo simulations[20] of single α,ω-donor–acceptor chains with different linker length and temperature. Self-assembly is driven by the complementarity of the α,ω-donor–acceptor end groups, which can form six hydrogen bonds, connected by L units of polystyrene, as illustrated in Scheme 1. For the system realized in the experiment (L close to 50), we performed long simulations, each started at a fully extended state, at constant temperatures in the temperature range from 220 K to 340 K. Fig. 1 shows a sample trajectory: we find two recurring values of the radius of gyration R_g and energy values (see inset) corresponding to fully folded and fully open conformations (refer to the movie clip in the ESI†) and many transitions between these two conformations. The closed and open conformations can be clearly distinguished by the radius of gyration, R_g, and the α,ω-distance (refer to Fig. S10, ESI†). In the closed state, all six hydrogen bonds are formed, leading to a gain in potential energy. For the chain with $L = 50$, we computed the equilibrium transition temperature between closed and open states as $T = 276.9$ K which compares with an experimental transition at ambient temperature (see Table 1). Observation of a sufficient number of transitions is a prerequisite for calculating the probability density surface (Fig. 1 bottom, see Fig. S3, ESI† for the curves at $T = 270$ K and $T = 280$ K), from which we can deduce ΔH and ΔS (refer to the ESI† and Table 1).

In order to describe the geometric shape of the closed configurations (see Fig. S7, ESI† for representative examples), we use the radius of gyration and three shape measures[21] (see ESI,† Section S4) to characterize the asphericity, acylindricity and relative anisotropy of the polymer chains (see Table 1, Fig. S6, ESI†). For example, the asphericity and acylindricity shape measures have been used in previous simulations of polyelectrolytes, where the
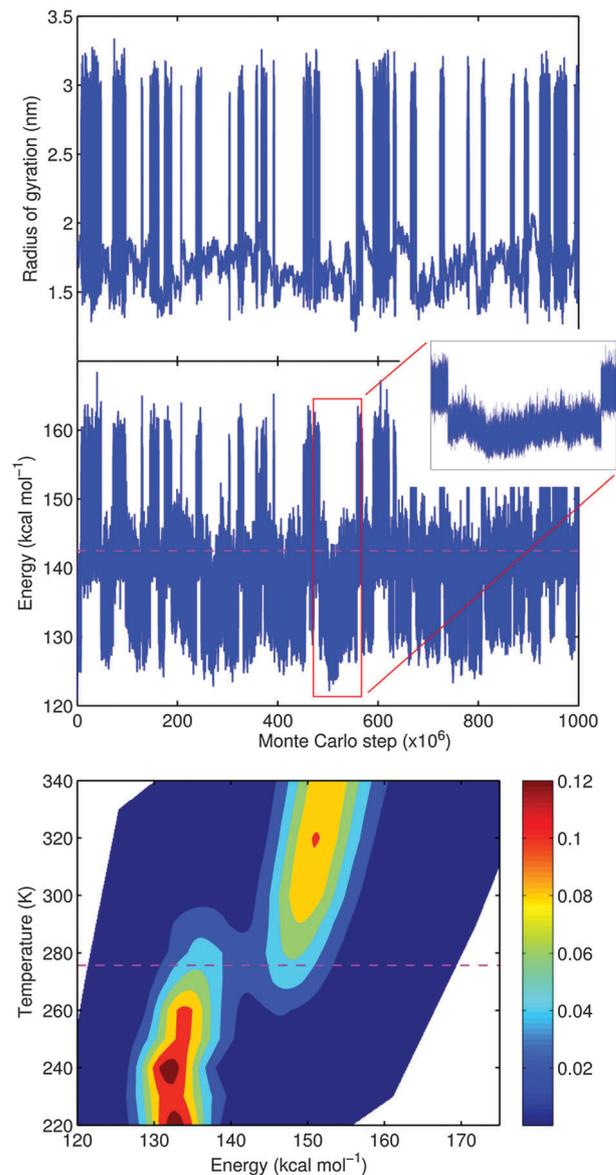
Fig. 1 Top: radius of gyration and energy of the conformations for $L = 50$ as function of the number of steps (in millions). Large values of the energy correspond to open, small values closed conformations (dashed line: separator of the open/closed conformations in energy), inset: sharp transitions between two clearly separated ensembles of open and closed conformations. Bottom: probability density profile computed from the trajectory (red regions are more occupied than blue regions), indicating two state folding behaviour. At the transition temperature (dashed line) the occupation of both states is equally probable.

authors found non-monotonic dependence of the asphericity on the salt concentration.[22,23] Fig. 2 (pink data) shows the distribution of the radius of gyration for the closed ensemble; the value for the open ensemble is $R_g = 2.23$ nm. All sampled open conformations show very narrow distributions for all the shape measures (refer to Fig. S6, ESI†). On average, the open configuration has a higher asphericity and a low value of the acylindricity. Thus, in the open state the polymer fluctuates around an elongated shape with the end groups pointing in opposite directions. The ensemble of the closed configuration shows a different behaviour. For $L = 50$ all

Table 1 Simulation results: single chain transition temperature, changes in entropy and enthalpy at this temperature between the open and closed ensemble, mean radius of gyration of the closed ensemble and the width of its distribution as function of polymer length L

L	T_M K	ΔH kcal mol^{-1}	ΔS kcal mol^{-1} K^{-1}	R_g nm (closed)	σ nm (closed)	R_g nm (open)
10	299.0	12.21	0.0408	0.66	0.032	0.70
20	292.0	12.58	0.0431	0.98	0.067	1.14
30	290.5	12.58	0.0433	1.25	0.086	1.52
50	276.9	12.91	0.0466	1.74	0.118	2.23

closed states have a well-defined mean radius of gyration of 1.74 nm with a mean square deviation of 0.118 nm.

To establish the validity of our model we have computed the hydrodynamic radius[24] as 2.26 nm in the open state and 1.93 nm in the closed configurations in quantitative agreement with the experiment, where open and closed chains have a hydrodynamic radius of 2.50 and 2.05 nm respectively.[9]

We thus conclude that the experimental system with $L = 50$ folds reversibly, however not into a unique native conformation. We have therefore investigated how the behaviour of the closed conformations changes as a function of the length of the polymer, as a shorter chain should confine the closed conformation at the cost of an increased entropy loss at the transition. In order to determine whether unique closed conformations are accessible with this system, we have generated trajectories for polymers with $L = 30$, 20, 10 and computed the folding probability density landscapes (refer to Fig. S2 in the ESI†). The thermodynamic data for all systems is summarized in Table 1. We observe a rather weak dependence of the folding temperature on L, which follows from the logarithmic dependence of the entropy loss ΔS on the chain length (see ESI,† Section S3), but a progressive sharpening of the width of the distribution of R_g (see Fig. 2).

As the inset of Fig. 2 clearly demonstrates, there is a striking transition where the reduced fluctuation of radius of gyration, *i.e.* the ratio of the standard deviation to the mean, no longer scales with the number of units, but is much smaller. To illustrate the transition,
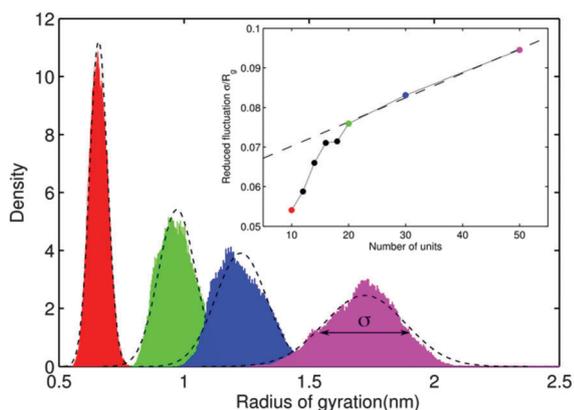
additional simulation data for chain lengths between 10 and 20 are shown in the inset of Fig. 2. As a single polymer unit has a dimension of approximately $5.4 \times 4.3 \times 2.5$ Å, the typical fluctuation of the chain in the closed conformation is of the order of the size of the unit for $L < 15$. Thus, as the system size decreases we predict the collapse of the "closed" conformation with local shape fluctuations that correspond to those also present in folded proteins. Fig. 3 shows alignment of ten closed conformations obtained from uncorrelated closed intervals in the Monte Carlo trajectory, which were separated by open transitions. Fig. 3a shows that the polymer with $L = 10$ in Fig. 3a folds into a unique configuration with root-mean-square deviation (RMSD) of backbone atoms equal to just 0.95 Å between completely uncorrelated closed conformations. Such precision in the alignment of the backbone is typically reached only for well struc-tured proteins. For $L = 20$ there are some fluctuations in the backbone, while the overall conformation is still well preserved (RMSD = 2.4 Å), reminiscent of an NMR ensemble of a small structured protein, while for $L = 50$, the conformation is still closed, but no longer unique. The data thus show that the limit of a unique conformation is accessible at experimentally feasible temperatures.

It is interesting to note that the width of the distribution of R_g is reduced as a function of L even when measured at the transition temperature, which is increasing with decreasing system size. One might expect that due to the increase in

Fig. 2 Histogram of the radius of gyration of the closed chain ensembles for $L = 10$ (red), $L = 20$ (green), $L = 30$ (blue) and $L = 50$ (purple). Inset: reduced width of the R_g distributions as a function of the length of the polymer chain.
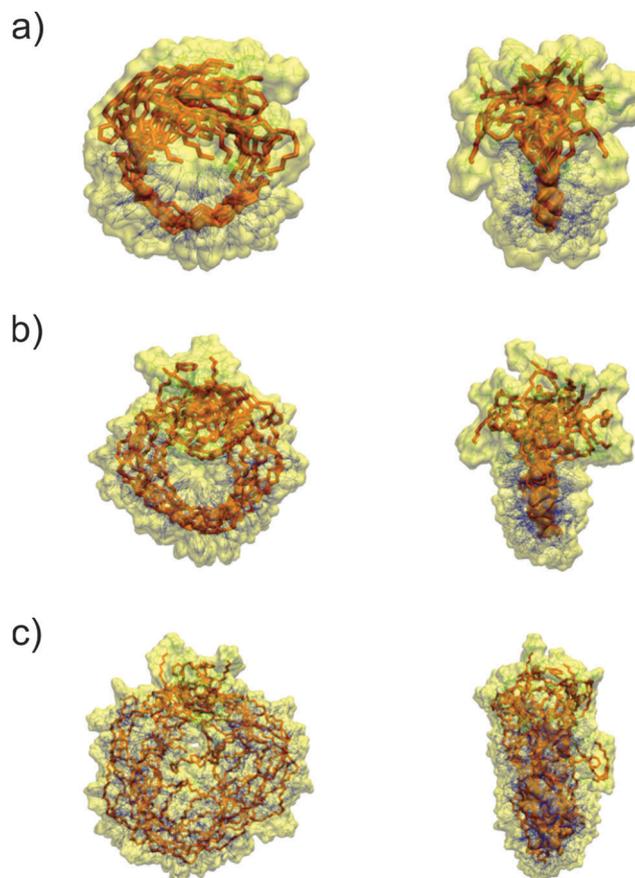
Fig. 3 Front and side views of ten closed conformations obtained from uncorrelated closed intervals for polymers with (a) $L = 10$ (RMSD = 0.95 Å), (b) $L = 20$ (RMSD = 2.44 Å) and (c) $L = 50$ (RMSD = 6.40 Å).

temperature the fluctuations in the closed ensemble increase at the transition temperature with decreasing length, yet the opposite is the case as the rigidity enforced by formation of the hydrogen bonds reduces the flexibility of the closed conformation.

In the present investigation we have thus succeeded to fully sample the opening and closing transitions of self-folding single polymer chains and evidence that collapse into a unique closed conformation is possible. This is only possible through very extensive atomistically resolved molecular simulations for a range of temperature and polymer length, resulting in a full thermodynamic characterization of the folding behaviour and a detailed view of the conformational distribution of these polymers, which is presently difficult to obtain experimentally. For the experimentally realized system we find good agreement with respect to the transition temperature and the change in the hydrodynamic radius. In contrast to proteins and peptides, where the native ensemble is stabilized by a large number of native interactions, selective point folding needs to succeed with only a few orthogonal interactions sites,[6–10] as these are difficult to install in larger numbers into precision polymer chains. Thus, our current results are highly encouraging: the near-unique closed conformation in the simple folding system studied here is stabilized by a single strong pair of interaction sites below a certain chain length limit. Observation of single chain collapse of the conformational ensemble to a near-unique conformation thus opens the prospect that polymers with unique native conformations can be engineered with relatively few interaction sites at temperatures that are experimentally attainable.[7,10]

In summary, folding into these conformations occurs through a reversible two-state process similar to that observed for small peptides and proteins. Simulations, such as those reported here, may in the future aid experimentalists in the design of polymer sequences that fold into unique states and to explore the degree of uncertainty in sequence control – which is presently unavoidable in all experimental methods – that the desired folded state can tolerate. In addition, studying single-polymer folding probes interaction mechanisms that radically differ from those present in proteins and DNA, which may help us test the generality of paradigms developed for biomolecular self-assembly

Notes and references

1 C. K. Lyon, A. Prasher, A. M. Hanlon, B. T. Tuten, C. A. Tooley, P. G. Frank and E. B. Berda, *Polym. Chem.*, 2015, **6**, 181 197.
2 K. Matyjaszewski and J. Xia, *Chem. Rev.*, 2001, **101**, 2921 2990.
3 M. Ouchi, T. Terashima and M. Sawamoto, *Chem. Rev.*, 2009, **109**, 4963 5050.
4 C. Barner Kowollik and A. J. Inglis, *Macromol. Chem. Phys.*, 2009, **210**, 987 992.
5 C. J. Hawker and K. L. Wooley, *Science*, 2005, **309**, 1200 1205.
6 O. Altintas and C. Barner Kowollik, *Macromol. Rapid Commun.*, 2012, **33**, 958 971.
7 O. Altintas, E. Lejeune, P. Gerstel and C. Barner Kowollik, *Polym. Chem.*, 2012, **3**, 640 651.
8 O. Altintas, T. Rudolph and C. Barner Kowollik, *J. Polym. Sci., Part A: Polym. Chem.*, 2011, **49**, 2566 2576.
9 O. Altintas, P. Gerstel, N. Dingenouts and C. Barner Kowollik, *Chem. Commun.*, 2010, **46**, 6291 6293.
10 O. Altintas, P. Krolla Sidenstein, H. Gliemann and C. Barner Kowollik, *Macromolecules*, 2014, **47**, 5877 5888.
11 T. Terashima, T. Mes, T. F. A. De Greef, M. A. J. Gillissen, P. Besenius, A. R. A. Palmans and E. W. Meijer, *J. Am. Chem. Soc.*, 2011, **133**, 4742 4745.
12 B. V. K. J. Schmidt, N. Fechler, J. Falkenhagen and J. F. Lutz, *Nat. Chem.*, 2011, **3**, 234 238.
13 M. A. J. Gillissen, T. Terashima, E. W. Meijer, A. R. A. Palmans and I. K. Voets, *Macromolecules*, 2013, **46**, 4120 4125.
14 N. Hosono, M. A. J. Gillissen, Y. Li, S. S. Sheiko, A. R. A. Palmans and E. W. Meijer, *J. Am. Chem. Soc.*, 2013, **135**, 501 510.
15 R. K. Roy and J. F. Lutz, *J. Am. Chem. Soc.*, 2014, **136**, 12888 12891.
16 O. Shishkan, M. Zamfir, M. A. Gauthier, H. G. Börner and J. F. Lutz, *Chem. Commun.*, 2014, **50**, 1570 1572.
17 A. Sanchez Sanchez, D. A. Fulton and J. A. Pomposo, *Chem. Commun.*, 2014, **50**, 1871 1874.
18 J. N. Onuchic, Z. Luthey Schulten and P. G. Wolynes, *Annu. Rev. Phys. Chem.*, 1997, **48**, 545 600.
19 K. Lindorff Larsen, S. Piana, R. O. Dror and D. E. Shaw, *Science*, 2011, **334**, 517 520.
20 T. Strunk, M. Wolf, M. Brieg, K. Klenin, A. Biewer, F. Tristram, M. Ernst, P. J. Kleine, N. Heilmann, I. Kondov and W. Wenzel, *J. Comput. Chem.*, 2012, **33**, 2602 2613.
21 D. N. Theodorou and U. W. Suter, *Macromolecules*, 1985, **18**, 1206 1214.
22 P. Y. Hsiao, *Macromolecules*, 2006, **39**, 7125 7137.
23 A. K. Das and P. Y. Hsiao, *J. Phys. Chem. B*, 2014, **118**, 6265 6276.
24 J. G. de la Torre and V. A. Bloomfield, *Q. Rev. Biophys.*, 1981, **14**, 81 139.