Unfolding of α-Helical 20-residue poly-glutamic acid analyzed by multiple runs of canonical molecular dynamics simulations

Naoki Ogasawara [1], **Kota Kasahara** [Corresp., 2], **Ryosuke Iwai** [1], **Takuya Takahashi** [2]

[1] Graduate School of Life Sciences, Ritsumeikan University, Kusatsu, Shiga, Japan

[2] College of Life Sciences, Ritsumeikan University, Kusatsu, Shiga, Japan

Corresponding Author: Kota Kasahara
Email address: ktkshr@fc.ritsumei.ac.jp

Elucidating the molecular mechanism of helix–coil transitions of short peptides is a long-standing conundrum in physical chemistry. Although the helix–coil transitions of poly-glutamic acid (PGA) have been extensively studied, the molecular details of its unfolding process still remain unclear. We performed all-atom canonical molecular dynamics (MD) simulations for a 20-residue PGA , over a total of 19 μs, in order to investigate its helix-unfolding processes in atomic resolution. Among the 28 simulations, starting with the α-helical conformation, all showed an unfolding process triggered by the unwinding of terminal residues, rather than by kinking and unwinding of the middle region of the chain. The helix-turn-helix conformation which is speculated by the previous experiments was not observed. Upon comparison between the N- and C-termini, the latter tended to be unstable and easily unfolded. While the probabilities of helix elongation were almost the same among the N-terminal, middle, and C-terminal regions of the chain, unwinding of the helix was enriched at the C-terminal region. The turn and 3_{10}-helix conformations were kinetic intermediates in the formation and deformation of α-helix, consistent with the previous computational studies for Ala-based peptides.

Unfolding of α-Helical 20-residue poly-glutamic acid analyzed by multiple runs of canonical molecular dynamics simulations

Naoki Ogasawara[1, †], Kota Kasahara[2, †,], Ryosuke Iwai[1], Takuya Takahashi[2]*

[1] Graduate School of Life Sciences, Ritsumeikan University, Noji-Higashi 1-1-1, Kusatsu, Shiga, 525-8577, Japan

[2] College of Life Sciences, Ritsumeikan University, Noji-Higashi 1-1-1, Kusatsu, Shiga, 525-8577, Japan

* To whom correspondence should be addressed:
Kota Kasahara, Tel: +81-77-566-1111; Fax: +81-77-561-3729; Email: ktkshr@fc.ritsumei.ac.jp.
† The first two authors are considered to be joint first authors.

Abstract

Elucidating the molecular mechanism of helix–coil transitions of short peptides is a long-standing conundrum in physical chemistry. Although the helix–coil transitions of poly-glutamic acid (PGA) have been extensively studied, the molecular details of its unfolding process still remain unclear. We performed all-atom canonical molecular dynamics (MD) simulations for a 20-residue PGA, over a total of 19 μs, in order to investigate its helix-unfolding processes in atomic resolution. Among the 28 simulations, starting with the α-helical conformation, all showed an unfolding process triggered by the unwinding of terminal residues, rather than by kinking and unwinding of the middle region of the chain. The helix-coil-helix conformation which is speculated by the previous experiments was not observed. Upon comparison between the N- and C-termini, the latter tended to be unstable and easily unfolded. While the probabilities of helix elongation were almost the same among the N-terminal, middle, and C-terminal regions of the chain, unwinding of the helix was enriched at the C-terminal region. The turn and 3_{10}-helix conformations were kinetic intermediates in the formation and deformation of α-helix, consistent with the previous computational studies for Ala-based peptides.

Introduction

Elucidation of the molecular mechanisms of protein folding is a central issue in physical chemistry. Since protein folding involves formation of secondary structural elements as building blocks of the tertiary structure (Richardson, 1981), understanding the dynamics of α-helical folding and unfolding, or helix–coil transition, is essential. The helix–coil transition has been

35 extensively studied in both experimental and theoretical methods using mainly Ala-based
36 polypeptides (Baldwin, 1995; Chen, Zhou & Ding, 2007; Neumaier et al., 2013) due to the high
37 helix propensity of Ala residues (Spek et al., 1999). Another representative model peptide is the
38 poly-glutamic acid (PGA). Since the side-chain of Glu has a titratable group, the chemical nature
39 of PGA can be modulated by the solution pH, and its helix–coil equilibrium can be controlled by
40 pH adjustments (Nakamura & Wada, 1981; Clarke et al., 1999; Kimura et al., 2002; Inoue,
41 Baden & Terazima, 2005; Causgrove & Dyer, 2006; Finke et al., 2007; Stanley & Strey, 2008;
42 Donten & Hamm, 2013; Gooding et al., 2013). Previous experiments on the helix–coil
43 transitions of PGA reported that, compared to neutral environments, acidic environments
44 enhance helix formation. The reported helix content of short PGAs in acidic environments varied
45 from 0.3 to 0.6, whereas it is below detectable limit in neutral pH (Clarke et al., 1999; Kimura et
46 al., 2002; Finke et al., 2007). Detailed scenario of the dynamics of helix–coil transitions is still
47 controversial. The previous reports have presented two different types of PGA conformations in
48 acidic environments: (i) a single α-helix with denatured termini and (ii) multiple short α-helices
49 connected by coil regions. Kimura *et al.* proposed that the single α-helical conformation arises
50 via intermediate states with several short helices, based on Fourier-transform infra-red (FTIR)
51 spectroscopy and circular dichroic (CD) experiments (Kimura et al., 2002). Clarke *et al.*
52 implied, based on stopped-flow CD measurements, that the single long α-helical conformation
53 successively decomposes into multi-helical conformations (Clarke et al., 1999). Finke *et al.*
54 supported this scenario based on fluorescence resonance energy transfer (FRET) measurements
55 (Finke et al., 2007).

56 In order to shed light on peptide conformational transitions at the atomic level, molecular
57 dynamics (MD) simulation is a promising approach. This method has been applied to investigate

the helix–coil transitions of Ala-based peptides, and the C-terminus has been reported to have a higher denaturing tendency compared to the N-terminus (Young & Brooks, 1996; Takano et al., 1999; Wu & Wang, 2001). In addition, the 3_{10}-helix and turn conformations were found to be kinetic intermediates for the helix–coil transitions (Young & Brooks, 1996; Takano et al., 1999). However, unlike that of the Ala-based peptides, helix-coil transitions of PGA peptides have not been studied using the all-atom MD method.

Here, we utilized the all-atom canonical MD method to simulate unfolding dynamics of a 20-residue PGA with fully protonated side chains, mimicking an acidic environment. Using the molecular model of a PGA with α-helical conformation as the initial structure, we repeated MD simulations for unfolding processes with different initial conditions. In total, 19-μs dynamics, consisting of 3 runs with 3.0 μs and 25 runs with 0.4 μs, were simulated. While various pathways of unfolding were observed in these 28 time courses, PGA unfolding was mainly seen to be triggered by denaturation of the termini, followed by propagation of the coil conformation toward the opposite side. Multiple-helix conformations implied by the previous experiments did not appear in the MD simulations.

Methods

Canonical MD Simulations

Dynamics of a 20-residue PGA, in an explicitly solvated periodic boundary cell, was investigated by the canonical MD method. We prepared two α-helical PGA structures as the initial structures for simulation. The first was an α-helical structure, sampled from an ensemble,

obtained by our replica-exchange MD (REMD) simulation, with an implicit solvent model. The details of the REMD simulation will be described elsewhere (unpublished data by Iwai R *et al*.). The second was an ideal α-helix, all the residues of which took the backbone dihedral angles φ = -60° and ψ = -45°, built using *tLEaP* software attached to AMBER package. The N- and C-termini of the PGA were capped with acetyl (Ace) and N-methyl (Nme) groups, respectively. All the carboxyl groups of the side-chains were protonated and net charge of the PGA was zero. Each molecular model of the PGA was placed in the truncated octahedral cell and solvated by filling with TIP3P water molecules (Jorgensen et al., 1983). The number of atoms composing the molecular system with the simulated structure of PGA was 10,592, and that with the ideal α-helix was 11,081. After that, the energy minimizations were successively performed with the steepest descent and conjugate gradient methods; the number of steps was 250 for each. The systems were relaxed via a 200-ps *NPT* simulation using Berendsen barostat. For the system with the ideal helix, the heavy atoms in the PGA were constrained during the relaxation run. The final snapshots of these two systems, referred to as *Sim* and *Ide*, were used as the initial structures of the production runs (Figure 1). The cell dimensions were 54.32 Å and 55.10 Å for *Sim* and *Ide*, respectively. As production runs, 8 and 20 runs of simulations were performed with *Sim* and *Ide* systems, respectively. Accordingly, we termed these simulations as *Sim1*, *Sim2*, ..., *Sim8*, and *Ide1*, *Ide2*, ..., *Ide20*. The initial atomic velocities were randomly generated with different random seeds for each run. The simulation time of each run was 0.4 μs except for *Sim1*, *Sim2*, and *Sim3* that lasted over 3.0 μs. These production runs were performed with the *NVT* ensemble at 300 K using the Langevin thermostat. The integration time step was 2.0 fs; the covalent-bond lengths with hydrogen atoms were constrained with the SHAKE algorithm (Ryckaert, Ciccotti & Berendsen, 1977). The non-bonded pairwise potentials were truncated at 10 Å of the interatomic

distance. For the potential energy calculations, AMBER ff99SB force field (Hornak et al., 2006) was applied. All the simulations were carried out using AMBER software.

Analyses

On the basis of trajectories of the atomic coordinates, recorded every 20 ps in the simulations, the helix–coil transitions of a PGA were analyzed using DSSP software (Kabsch & Sander, 1983). DSSP recognizes the secondary structural elements in terms of hydrogen bonding patterns of the main-chains, and categorizes them into the following eight classes: α-helix, 3_{10}-helix, π-helix, extended β-strand, isolated β-bridge, turn, bend, and others. Each class is represented by an alphabetical symbol; *H*, *G*, *I*, *E*, *B*, *T*, *S*, and *O*, respectively. Note that the symbol '*O*' is introduced in this paper for convenience, and it is denoted as ' ' (white or blank space) in the output of the DSSP software. The secondary structure content in the *Ide* trajectories were referred to as $P^{Ide}(x;i)$ for the contents of the secondary structure *x* (any of the eight classes) at the *i*-th residue. The superscript "*Ide"* indicates that the ensemble was obtained from the 20 *Ide* runs with 0.4 μs each. The ensemble consisting of trajectories of 8 *Sim* runs with 0.4 μs each is indicated as the superscript "*Sim"*, and that of *Sim1–Sim3* with 3.0 μs each is indicated as the superscript "*Sim1–3"*. The secondary structure content for the entire chain is presented as $P^{Ide}(x)$. The transition probabilities of *i*-th residue, from the secondary structure *x* to *y* between the successive snapshots (20 ps of the time interval), $P^{Ide}(y,x;i)$, were also evaluated. To measure the time required for complete unfolding of an α-helix, we defined the unfolding time, t_u, as the time corresponding to the first snapshot without α-helical residues in a trajectory.

124

125 Results

126 Micro-second Dynamics of a PGA

127 In order to investigate long-term behavior of a PGA, we performed three runs of 3.0-μs MD
128 simulations (*Sim1*, *Sim2*, and *Sim3*) with the same initial atomic coordinates but different atomic
129 velocities (Figure 1). The initially formed α-helix was deformed immediately after beginning the
130 simulations in all the three runs (Figure 2). The unfolding times, t_u, defined as the time of the
131 first snapshot without an α-helical residue in PGA for each trajectory, were 31.06 ns, 100.52 ns,
132 and 7.38 ns in *Sim1*, *Sim2*, and *Sim3* simulations, respectively. In the simulation with the longest
133 unfolding time (*Sim2*), after unfolding of the initial α-helix, the helical conformation was
134 temporarily reformed at the N-terminal half of the chain at around 0.2 μs (Figure 2G). However,
135 the reformed helix was unfolded at 0.34 μs, and a helix longer than 13 residues was not formed
136 till the end. In the *Sim1* simulation, although the initial helix was immediately unfolded, a long
137 helix consisting of 17 residues was refolded and retained over sub-micro second time scale
138 (Figure 2E). This helix nucleated between 12th to 16th residues at 0.62 μs (Figure 2D) and
139 propagated over the range from 2nd to 18th residues. While the N-terminal half of the helix was
140 deformed at 0.84 μs (Figure 2F), the latter half remained intact till 0.95 μs. On the other hand, re-
141 formation of stable helix did not occur in *Sim3*, although several helix-nucleation events were
142 observed. Overall, helix formation was a relatively rare event in this time scale. In addition,
143 while several helix-nucleation events were observed, the nucleated helices disappeared
144 immediately in most cases. Helix nucleation seemed to be coupled with the turn conformation
145 (Figures 2A, B, and C), discussion on which will be taken up later. Formation of a β-sheet was

146 also observed as a rare event. β-sheet formation in *Sim2* was exceptionally stable and was
147 retained during 0.63 μs (Figure 2H).

148 In the time course of the secondary structural elements at each residue (Figure 2A, B, and
149 C), some "bands" could be observed; for example, the turn conformation was almost always
150 formed at 9th and 10th residues in *Sim2*. Since the tendency to form a turn at the 9th and 10th
151 residues was not observed in the other runs, it is considered to be due to an initial condition,
152 rather than an intrinsic propensity of the 9th and 10th residues. This indicates that there was the
153 strong time-correlation of secondary structure formation, and 3.0 μs was not enough to reach an
154 equilibrium state. The time course of the ensemble average of the helix content (summation over
155 the α- and 3_{10}-helix conformations; $P(H) + P(G)$), for *Sim1*, *Sim2*, and *Sim3* implies that the
156 trajectories were not well-converged (Figure 3). While the helix content of the three trajectories
157 became converged to similar values with the evolution of time, they still aquired different values
158 at the end of the simulations. The helix content in the full-length trajectories of *Sim1*, *Sim2*, and
159 *Sim3* were 0.14, 0.12, and 0.078, respectively. This implies that equilibration of the system
160 requires longer time scales.

161

162 Unfolding Dynamics

163 Non-equilibrium processes involved in the transformation of an α-helix into denatured structures
164 were analyzed by scrutinizing the first part of each trajectory. We additionally performed 25
165 short (400 ns for each) simulations and analyzed the unfolding processes of the 28 simulations in
166 total. Note that 8 of them started from an α-helical conformation obtained from a simulation
167 (*Sim1–Sim8*; Figure 1), and remaining 20 started from an artificially built ideal α-helix (*Ide1–*

168 *Ide20*; Figure 1). As a result, all the 28 runs showed corruption of the α-helical conformation
169 within 400 ns (Figures 4 and 5). The unfolding times (t_u) varied from 7.38 ns (*Sim3*) to 380.70 ns
170 (*Sim6*), and the average, median, and the standard deviation (SD) were 75.63 ns, 36.02 ns, and
171 92.18 ns, respectively (Table 1). There was no statistically significant difference between *Sim*
172 and *Ide* simulation results; the average (median; SD) of t_u were 72.65 ns (36.02 ns; 79.88 ns) and
173 83.09 ns (37.32 ns; 123.97 ns), for *Sim* and *Ide*, respectively.

174 The unfolding trajectories were varied among the 28 trajectories. The fastest unfolding
175 was observed in *Sim3*. The helix deformed from both the N- and C-termini immediately after the
176 simulation began (Figure 5C). As described above, while a single-turn helix sometimes formed
177 at the N- and C-termini after unfolding, they did not grow into a longer helix. The bend
178 conformations were stably formed at 5, 6, 7, 10, and 11th residues during 400 ns. On the
179 contrary, *Sim6* showed the slowest dynamics of unfolding. While three or four residues from the
180 N-terminus were immediately deformed, the remaining part of the helix was retained for a long
181 time (Figure 5F). As described above, strong time correlations were observed in all the
182 trajectories (Figures 4 and 5). After immediate unfolding of the α-helix, a denatured
183 conformation of the peptide was not randomized in this time scale.

184 For all the 28 trajectories, unfolding mechanisms were analyzed in terms of the order of
185 deformation for each region in the polypeptide chain. We classified the residues into three
186 regions; *i.e.*, the N-terminal region (2nd–7th residues), the middle region (8th–13th residues),
187 and the C-terminal region (14th–19th residues). The first and 20th residues were discarded
188 because of the following reasons: they would be highly influenced by the truncation of the chain;
189 the main-chain hydrogen bonding pattern of the first residue cannot be defined due to lack of the
190 N-terminal neighbor; all the regions should have the same number of residues. Next, the order of

191 unfolding, for these regions, was assessed based on the helix content of each region in the time
192 period ranging from the beginning of simulation to the unfolding time, t_u. As a result, the
193 unfolding process beginning with the deformation of the middle region was not observed, and all
194 the unfolding processes began with unwinding of one of the terminal regions (Table 1). In
195 addition, coil regions propagated toward both the directions in many cases. There are two
196 possible scenarios for completion of unfolding from any terminus: (i) the coil region appears in a
197 terminus and elongates toward the opposite terminus ("N, M, C" and "C, M, N" in Table 1), and
198 (ii) the opposite terminus is successively unfolded followed by elongation of coil regions from
199 both the termini to the middle ("N, C, M" and "C, N, M" in Table 1). The fact that the former
200 scenario was observed in only 3 and 2 runs among 20 *Ide* and 8 *Sim* runs, respectively, suggests
201 the latter being the major way of α-helix unfolding in this system.

202 Comparing between the N- and C-termini of the peptide chain, unfolding from the C-
203 terminus was preferred over that from the N-terminus; 13 runs out of the 20 *Ide* runs and 7 runs
204 out of the 8 *Sim* runs showed unfolding from the C-terminus. Difference between the two termini
205 was clearer in *Sim* runs than in *Ide* runs, probably because of the slightly distorted initial
206 structure of *Sim* (Figure 1). The ensemble averages of residue-wise α-helix content in *Ide1–20*
207 with 0.4 μs each ($P^{Ide}(H;i)$), *Sim1–8* with 0.4 μs each ($P^{Sim}(H;i)$), and *Sim1–3* with 3.0 μs each (
208 $P^{Sim1-3}(H;i)$) also showed a lower helical tendency at the C-terminus than at the N-terminus
209 (Figure 6). The previous simulation studies (Young & Brooks, 1996; Wu & Wang, 2001; Finke
210 et al., 2007) had also reported that helix formation of the C-terminal residues was unstable
211 compared to that of the N-terminal ones.

212

Secondary Structural Transitions

To analyze the detailed mechanisms of conformational transitions in shorter time scales, we assessed the probability of the event that the *i*-th residue in the secondary structure *A* at time *t* is transformed into *B* at time *t* + 20 ps; the averaged probability over the 20 *Ide* runs is referred to as $P^{Ide}(B,A;i)$. For simplicity, we focused on the four classes of secondary structure elements; *H*, *G*, *T*, and $\overline{HGT}$, which means any of the rest five (*I*, *E*, *S*, *B*, and *O*). The cases i = 2, 11, and 19 were analyzed as representatives of the N-terminal, middle, and C-terminal residues, respectively (Table 2). The C-terminal residues showed a weaker tendency to retain the α-helical conformation compared to the other residues ($P^{Ide}(H,H;2) = 0.94$, $P^{Ide}(H,H;11) = 0.96$, and $P^{Ide}(H,H;19) = 0.58$). The weaker tendency to retain the same conformation in the C-terminal region was also observed in the other secondary structures. The results of *Sim* runs were qualitatively consistent with that of *Ide* runs (Table S1).

The helix–coil transitions mainly occurred via the turn conformation. More than half of the conformational transitions from α-helix directed to the turn conformation; $P^{Ide}(T,H;i)/P^{Ide}(\overline{H},H;i)$ for i = 2, 11, and 19 were 0.52, 0.73, and 0.80, respectively. Where $\overline{H}$ denotes the secondary structure other than H. In addition, formation of the α-helix via turn was enriched in the C-terminal residue; $P^{Ide}(H,T;i)/P^{Ide}(\overline{T},T;i)$ for i = 2, 11, and 19 were 0.19, 0.62, and 0.42, respectively. Thus, the turn conformation can be considered as an intermediate state in the helix–coil transition, especially at the C-terminus. Another intermediate in the α-helix formation is the 3_{10}-helix. While a major destination state of a 3_{10}-helix was the turn ($P^{Ide}(T,G;i)/P^{Ide}(\overline{G},G;i)$ for i = 2, 11, and 19 were 0.55, 0.62, and 0.64, respectively), it also transformed into an α-helix,

235 especially at the middle position; $P^{Ide}(H,G;i)/P^{Ide}(\overline{G},G;i)$ for i = 2, 11, and 19 were 0.20, 0.34, and
236 0.076, respectively. This result agreed with the previous theoretical studies, which reported that
237 the 3_{10}-helix is not a thermodynamic intermediate but could be a kinetic intermediate (Young &
238 Brooks, 1996; Wu & Wang, 2001).

239 In addition to the position of amino acids in the polypeptide chain, effect of the α-helical
240 ends was analyzed. We focused on segments consisting of three consecutive residues in the chain,
241 and state of the segment was defined as the combination of secondary structures of the three
242 residues, grouped into the two classes, *i.e.*, α-helix ("H") and others ("–"; it has the same
243 meaning as "$\overline{H}$"). The state of a segment was divided into the following eight classes: "HHH", "
244 HH–", "–HH", "H–H", "H–– ", "–––HH", and "–––". The state "–H–" is impossible, because α-
245 helical conformation coincides with at least four consecutive residues. The probability of the
246 event that the central residue of a segment forms an α-helix at the next snapshot (20 ps later) was
247 analyzed for each class. For instance, probability for the class "HH–", denoted as $P^{Ide}(H, HH–)$,
248 means the probability to retain α-helical conformation for the residue at the C-terminal end of an
249 α-helix, regardless of the position in the chain (i). The probability of deformation of the C-
250 terminal end of an α-helix can be shown as $P^{Ide}(–, HH–) = 1 – P^{Ide}(H, HH–)$. The probabilities are
251 summarized in Table 3; the case of *Sim* runs is shown in Table S2. We found that, a residue at
252 the interior of an α-helix was stabler to maintain the α-helical conformation, compared to the
253 terminal residues; $P^{Ide}(H, HHH) = 0.97$. It is noteworthy that the C-terminal end of an α-helix is
254 more frequently deformed than the N-terminal one; $P^{Ide}(H, HH–) = 0.74$ and $P^{Ide}(H, –HH) = 0.92$.
255 In addition, α-helix elongation toward the C-terminus was enriched compared to that toward the
256 opposite direction; $P^{Ide}(H, H––) = 0.23$ and $P^{Ide}(H, ––H) = 0.04$. The C-terminal end of an α-helix
257 unstably changed its conformation while the N-terminal end tended to retain its conformation.

We also evaluated the relationship between the two definitions of position; position in an α-helix (the N-terminal end, internal, and the C-terminal end) and position in the polypeptide chain (the N-terminal region [$2 \leq i \leq 7$], middle region [$8 \leq i \leq 13$], and C-terminal region [$14 \leq i \leq 19$]). The probability of helix–coil transitions in the center of a three-residue segment x was assessed for each of the three regions y: $P^{Ide}(-,x;y) = 0.04$, where x is "$HH-$" or "$-HH$" for the C- and N-terminal ends of an α-helix, respectively, and y is any of "N", "M", and "C", for the N-terminal, middle, and C-terminal regions, respectively. The probabilities to unfold the N- and C-terminal ends of an α-helix varied with respect to the position of the ends in the entire chain; namely, higher probabilities were observed in the C-terminal region of the peptide chain ($P^{Ide}(-, HH-;C) > P^{Ide}(-, HH-;N)$ and $P^{Ide}(-, -HH;C) > P^{Ide}(-, -HH;N)$ in Table 3). While residue-wise α-helical content (Figure 6) and α-helix retention probability (Table 2) indicate the highest α-helical propensity for the middle region, the lowest probabilities to unfold the ends of α-helix was found in the N-terminal region. In contrast, probabilities for elongation of an α-helix were almost the same for all the three regions (see $P^{Ide}(H, H--)$ and $P^{Ide}(H, --H)$ in Table 3). Therefore, an α-helical PGA tended to unfold from the C-terminus.

On the other hand, the α-helix nucleation was observed in low probabilities regardless of positions in the chain; $P^{Ide}(H, ---) = 0.02$ for all three regions.

Discussion

In this study, we examined the dynamics of a 20-residue PGA with 28 runs of all-atom canonical MD simulations. While three of them simulated 3.0-μs time courses, the systems were not well-equilibrated (Figure 3) and complete refolding of the α-helix was not observed (Figures 4 and 5).

The time scale required for α-helix formation by PGA, still remains controversial. The suggested time-scale varies from sub-micro to milli seconds (Clarke et al., 1999; Kimura et al., 2002; Causgrove & Dyer, 2006; Qin et al., 2014). Our simulation results imply that a time range of few micro-seconds is too short to refold PGA in acidic environments.

We mainly focused on the non-equilibrium dynamics of unfolding processes and repeated 28 runs of simulations with the two different initial α-helical structures. The results from these two initial structures were qualitatively similar. Higher stability of the α-helical conformation was shown to be in the middle of the polypeptide chain than at the termini. All the unfolding processes of the α-helix began from a terminus, but a helix-coil-helix conformation was not stably observed. In many cases, the unfolding proceeded toward both directions, rather than starting from a terminus and ending at the opposite. In addition, unfolding from the C-terminal side was preferred over that from the N-terminal side (Table 1). The probability of retention of α-helix at each residue was lower in C-terminus than in N-terminus (Table 2). While the probabilities of α-helix elongation were almost the same irrespective of whether the end was located at the N-terminus, middle, or C-terminus of the polypeptide chain, the probabilities of unwinding of the α-helix tended to be higher at the C-terminus of the chain (Table 3). The instability of α-helix at the C-terminus was due to the enhancement of unfolding, rather than reduction of folding. In the process of folding and unfolding of the α-helices, the turn and 3_{10}-helix conformations can be kinetic intermediates as consistent to the precedent studies (Young & Brooks, 1996; Wu & Wang, 2001; Pal, Chakrabarti & Basu, 2003).

Despite the wide acceptance of the all-atom MD method, there are still some issues under consideration. First, treatment of denatured proteins has not been fully validated in current force fields. Underestimation of the radius of gyration of denatured proteins by standard force fields

303 and water models has been previously reported (Piana, Klepeis & Shaw, 2014). Improved force
304 fields and water models have been proposed to simulate denatured proteins (Piana et al., 2015;
305 Henriques & Skepö, 2016). Second, although the force field applied here, AMBER ff99SB, is
306 one of the standard force fields, there are some reports about its weakness; *e.g.*, underestimation
307 of helix stability (Sorin & Pande, 2005) and discrepancy with the quantum mechanical
308 calculations (Takano, Kusaka & Nakamura, 2016). Third, finite-size effects have been reported
309 for the helix-stability of a model polypeptide (Weber, Hünenberger & McCammon, 2000;
310 Kastenholz & Hünenberger, 2004; Reif et al., 2009; Kasahara, Sakuraba & Fukuda, 2018). In
311 fact, helix content in the simulated ensembles (Figures 3 and 6) were lower than the
312 experimentally reported values, which is in the range of 0.3 to 0.6. However, differences in the
313 experimental method and conditions may cause differences in the helix content (Kimura et al.,
314 2002), since precise measurement of the latter for short peptides is not straight forward (Kelly,
315 Jess & Price, 2005; Greenfield, 2007).

316 Finally, discussion on the quantitative aspects of the results, *e.g.*, helix contents and
317 folding kinetics, provided by both the experimental and theoretical methods in this study, should
318 be carefully considered. From qualitative aspects, our results were consistent with the reported
319 theoretical studies, in spite of several differences in the materials and methods, *e.g.*, peptide
320 sequence, parameters, and sampling methods. For example, the weaker helix formation
321 propensity at the C-terminus and the kinetic intermediates of helix–coil transitions were
322 consistently concluded from this study in agreement with the previous theoretical studies. They
323 are robust conclusions, regardless of adjustable settings and simulation methods. In addition to
324 that, our simulation results provide statistics of kinetic details of helix–coil transition by multiple
325 runs of canonical MD. The weaker helix formation propensity at the C-terminus is due to high

326 frequency of unwinding rather than disfavoring of folding. Helix-coil-helix conformations
327 speculated by previous experiments were not observed.

328

329 Conclusion

330 In this study, the unfolding mechanism of α-helix in 20-residue PGA was investigated using all-
331 atom canonical MD simulations. Our results suggested that the unfolding was triggered by
332 unwinding of a terminus, whereas the multiple short-helical conformations, implied in the
333 previous experiments (Clarke et al., 1999; Kimura et al., 2002; Finke et al., 2007), were not
334 stably observed in the simulated trajectories within the micro-second time-scale. The instability
335 of C-terminus is consistent with the previously reported result from generalized ensemble
336 simulation of the poly-Ala peptides (Young & Brooks, 1996; Takano et al., 1999; Wu & Wang,
337 2001). The mechanism of helix–coil transitions, shown here, might reflect the nature of the
338 peptide backbone, and provide insight into the helix–coil transitions for general cases of
339 polypeptides.

340

341 Acknowledgements

342 The supercomputer resources were provided by the HPCI System Research Projects (Project IDs:
343 hp170020 and hp170025) and the National Institute of Genetics, Research Organization of
344 Information and Systems, Japan. We thank Tomoya Hirano for help with data analyses.

345

346 REFERENCES

347 Baldwin RL 1995. α-Helix formation by peptides of defined sequence. *Biophysical chemistry*
348 55:127–135. DOI: 10.1016/0301-4622(94)00146-B.
349 Causgrove TP, Dyer RB 2006. Nonequilibrium protein folding dynamics: laser-induced pH-jump
350 studies of the helix–coil transition. *Chemical Physics* 323:2–10. DOI:
351 10.1016/j.chemphys.2005.08.032.
352 Chen Y, Zhou Y, Ding J 2007. The helix–coil transition revisited. *Proteins: Structure, Function,*
353 *and Bioinformatics* 69:58–68. DOI: 10.1002/prot.21492.
354 Clarke DT, Doig AJ, Stapley BJ, Jones GR 1999. The α-helix folds on the millisecond time
355 scale. *Proceedings of the National Academy of Sciences of the United States of America*
356 96:7232–7237. DOI: 10.1073/pnas.96.13.7232.
357 Donten ML, Hamm P 2013. pH-jump induced α-helix folding of poly-l-glutamic acid. *Chemical*
358 *Physics* 422:124–130. DOI: 10.1016/j.chemphys.2012.11.023.
359 Finke JM, Jennings PA, Lee JC, Onuchic JN, Winkler JR 2007. Equilibrium unfolding of the
360 poly(glutamic acid)$_{20}$ helix. *Biopolymers* 86:193–211. DOI: 10.1002/bip.20719.
361 Gooding EA, Sharma S, Petty SA, Fouts EA, Palmer CJ, Nolan BE, Volk M 2013. pH-dependent
362 helix folding dynamics of poly-glutamic acid. *Chemical Physics* 422:115–123. DOI:
363 10.1016/j.chemphys.2012.11.009.
364 Greenfield NJ 2007. Using circular dichroism spectra to estimate protein secondary structure.
365 *Nature Protocols* 1:2876–2890. DOI: 10.1038/nprot.2006.202.
366 Henriques J, Skepö M 2016. Molecular Dynamics Simulations of Intrinsically Disordered
367 Proteins: On the Accuracy of the TIP4P-D Water Model and the Representativeness of
368 Protein Disorder Models. *Journal of chemical theory and computation* 12:3407–3415. DOI:
369 10.1021/acs.jctc.6b00429.
370 Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C 2006. Comparison of
371 multiple Amber force fields and development of improved protein backbone parameters.
372 *Proteins: Structure, Function, and Bioinformatics* 65:712–725. DOI: 10.1002/prot.21123.
373 Inoue K, Baden N, Terazima M 2005. Diffusion Coefficient and the Secondary Structure of
374 Poly-l-glutamic Acid in Aqueous Solution. *The Journal of Physical Chemistry B* 109:22623–
375 22628. DOI: 10.1021/jp052897y.
376 Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML 1983. Comparison of simple
377 potential functions for simulating liquid water. *The Journal of Chemical Physics* 79:926–
378 935. DOI: 10.1063/1.445869.
379 Kabsch W, Sander C 1983. Dictionary of protein secondary structure: Pattern recognition of
380 hydrogen-bonded and geometrical features. *Biopolymers* 22:2577–2637. DOI:
381 10.1002/bip.360221211.
382 Kasahara K, Sakuraba S, Fukuda I 2018. Enhanced Sampling of Molecular Dynamics
383 Simulations of a Polyalanine Octapeptide: Effects of the Periodic Boundary Conditions on
384 Peptide Conformation. *The Journal of Physical Chemistry B*, Accepted.
385 Kastenholz MA, Hünenberger PH 2004. Influence of Artificial Periodicity and Ionic Strength in
386 Molecular Dynamics Simulations of Charged Biomolecules Employing Lattice-Sum
387 Methods. *The Journal of Physical Chemistry B* 108:774–788. DOI: 10.1021/jp0350924.
388 Kelly SM, Jess TJ, Price NC 2005. How to study proteins by circular dichroism. *Biochimica et*
389 *Biophysica Acta (BBA) - Proteins and Proteomics* 1751:119–139. DOI:
390 10.1016/j.bbapap.2005.06.005.

391 Kimura T, Takahashi S, Akiyama S, Uzawa T, Ishimori K, Morishima I 2002. Direct
392 Observation of the Multistep Helix Formation of Poly-l-glutamic Acids. *Journal of the*
393 *American Chemical Society* 124:11596–11597. DOI: 10.1021/ja026639f.
394 Nakamura H, Wada A 1981. Dielectric studies of aqueous solutions of poly(L-glutamic acid).
395 *Biopolymers* 20:2567–2582. DOI: 10.1002/bip.1981.360201207.
396 Neumaier S, Reiner A, Büttner M, Fierz B, Kiefhaber T 2013. Testing the diffusing boundary
397 model for the helix-coil transition in peptides. *Proceedings of the National Academy of*
398 *Sciences* 110:12905–12910. DOI: 10.1073/pnas.1303515110.
399 Pal L, Chakrabarti P, Basu G 2003. Sequence and Structure Patterns in Proteins from an Analysis
400 of the Shortest Helices: Implications for Helix Nucleation. *Journal of Molecular Biology*
401 326:273–291. DOI: 10.1016/S0022-2836(02)01338-4.
402 Piana S, Donchev AG, Robustelli P, Shaw DE 2015. Water Dispersion Interactions Strongly
403 Influence Simulated Structural Properties of Disordered Protein States. *The Journal of*
404 *Physical Chemistry B* 119:5113–5123. DOI: 10.1021/jp508971m.
405 Piana S, Klepeis JL, Shaw DE 2014. Assessing the accuracy of physical models used in protein-
406 folding simulations: quantitative evidence from long molecular dynamics simulations.
407 *Current opinion in structural biology* 24:98–105. DOI: 10.1016/j.sbi.2013.12.006.
408 Qin Z-J, Shimizu A, Li J, Ikeguchi M, Shinjo M, Kihara H 2014. α-helix formation rate of
409 oligopeptides at subzero temperatures. *BIOPHYSICS* 10:9–13. DOI:
410 10.2142/biophysics.10.9.
411 Reif MM, Kräutler V, Kastenholz MA, Daura X, Hünenberger PH 2009. Molecular Dynamics
412 Simulations of a Reversibly Folding β-Heptapeptide in Methanol: Influence of the Treatment
413 of Long-Range Electrostatic Interactions. *The Journal of Physical Chemistry B* 113:3112–
414 3128. DOI: 10.1021/jp807421a.
415 Richardson JS 1981. The anatomy and taxonomy of protein structure. *Advances in protein*
416 *chemistry* 34:167–339.
417 Ryckaert JP, Ciccotti G, Berendsen HJC 1977. Numerical integration of the cartesian equations
418 of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of*
419 *Computational Physics* 23:327–341. DOI: 10.1016/0021-9991(77)90098-5.
420 Sorin EJ, Pande VS 2005. Exploring the Helix-Coil Transition via All-Atom Equilibrium
421 Ensemble Simulations. *Biophysical Journal* 88:2472–2493. DOI:
422 10.1529/biophysj.104.051938.
423 Spek EJ, Olson CA, Shi Z, Kallenbach NR 1999. Alanine Is an Intrinsic α-Helix Stabilizing
424 Amino Acid. *Journal of the American Chemical Society* 121:5571–5572. DOI:
425 10.1021/ja990056x.
426 Stanley CB, Strey HH 2008. Osmotically Induced Helix-Coil Transition in Poly(Glutamic Acid).
427 *Biophysical Journal* 94:4427–4434. DOI: 10.1529/biophysj.107.122705.
428 Takano M, Yamato T, Higo J, Suyama A, Nagayama K 1999. Molecular Dynamics of a 15-
429 Residue Poly(l-alanine) in Water: Helix Formation and Energetics. *Journal of the American*
430 *Chemical Society* 121:605–612. DOI: 10.1021/ja982919c.
431 Takano Y, Kusaka A, Nakamura H 2016. Density functional study of molecular interactions in
432 secondary structures of proteins. *BIOPHYSICS* 13:27–35. DOI: 10.2142/biophysico.13.0_27.
433 Weber W, Hünenberger PH, McCammon JA 2000. Molecular Dynamics Simulations of a
434 Polyalanine Octapeptide under Ewald Boundary Conditions: Influence of Artificial
435 Periodicity on Peptide Conformation. *The Journal of Physical Chemistry B* 104:3668–3675.
436 DOI: 10.1021/jp9937757.

Wu X, Wang S 2001. Helix Folding of an Alanine-Based Peptide in Explicit Water. *The Journal of Physical Chemistry B* 105:2227–2235. DOI: 10.1021/jp004048a.

Young WS, Brooks CL 1996. A microscopic view of helix propagation: N and C-terminal helix growth in alanine helices. *Journal of Molecular Biology* 259:560–572. DOI: 10.1006/jmbi.1996.0339.

Figure Legends

Figure 1. The initial structures of MD simulations. (A) The structure built by a REMD simulation, termed *Sim*. (B) The structure based on the ideal α-helix, termed *Ide*. (C) φ-ψ angles of 2nd–20th residues in *Sim* (triangles; the open triangles indicate the 2nd and 20th) and *Ide* (circle; all residues have the same values).

Figure 2. The 3.0-μs time courses of the secondary structure elements and examples of snapshots for *Sim1*, *Sim2*, and *Sim3* simulations. (A, B, and C) The time courses for *Sim1*, *Sim2*, and *Sim3*, respectively. The horizontal axis is the simulation time, and the vertical axis indicates the amino acid position in the peptide chain. Each block is filled by one of the eight types of colors regarding the secondary structure elements *H*, *G*, *I*, *E*, *B*, *T*, *S*, and *O*, presented as red, maroon, dark-red, gray, black, dark-cyan, cyan, and white, respectively. (D, E, F, G, and H) Snapshots at (D) 0.625 μs in *Sim1*, (E) 0.804 μs in *Sim1*, (F) 0.842 μs in *Sim1*, (G) 0.199 μs in *Sim2*, and (H) 1.000 μs in *Sim2*.

Figure 3. Time course of helix content averaged over accumulated time duration of each trajectory in *Sim1*, *Sim2*, and *Sim3*.

Figure 4. The 400-ns time courses of the secondary structure elements of *Ide1–20* for the panels (A)–(T), respectively. See the legend of Figures 2A, B, and C.

Figure 5. The 400-ns time courses of the secondary structure elements of *Sim1–8* for the panels (A)–(H), respectively. See the legend of Figures 2A, B, and C.

464 **Figure 6.** Residue-wise secondary structure content of α-helix (*H*; solid), 3_{10}-helix (*G*; dashed),
465 and turn conformations (*T*; dotted). (A) The average over 20 *Ide* runs ($P^{Ide}(x;i)$). (B) The average
466 over the 400-ns trajectories of 8 *Sim* runs ($P^{Sim}(x;i)$). (C) The average over 3.0-μs trajectories of
467 *Sim1*, *Sim2*, and *Sim3* ($P^{Sim1-3}(x;i)$).

468

Figure 1

The initial structures of MD simulations.

(A) The structure built by a REMD simulation, termed *Sim*. (B) The structure based on the ideal α-helix, termed *Ide*. (C) *φ*-*ψ* angles of 2nd–20th residues in *Sim* (triangles; the open triangles indicate the 2nd and 20th) and *Ide* (circle; all residues have the same values).

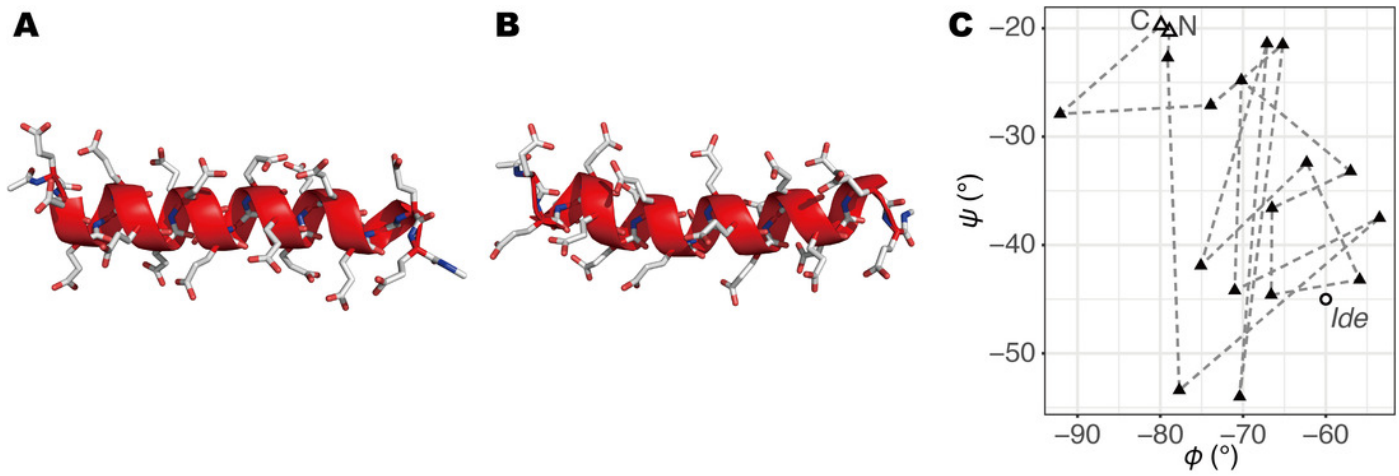

Figure 2

The 3.0-μs time courses of the secondary structure elements and examples of snapshots for *Sim1*, *Sim2*, and *Sim3* simulations.

(A, B, and C) The time courses for *Sim1*, *Sim2*, and *Sim3*, respectively. The horizontal axis is the simulation time, and the vertical axis indicates the amino acid position in the peptide chain. Each block is filled by one of the eight types of colors regarding the secondary structure elements *H*, *G*, *I*, *E*, *B*, *T*, *S*, and *O*, presented as red, maroon, dark-red, gray, black, dark-cyan, cyan, and white, respectively. (D, E, F, G, and H) Snapshots at (D) 0.625 μs in *Sim1*, (E) 0.804 μs in *Sim1*, (F) 0.842 μs in *Sim1*, (G) 0.199 μs in *Sim2*, and (H) 1.000 μs in *Sim2*.

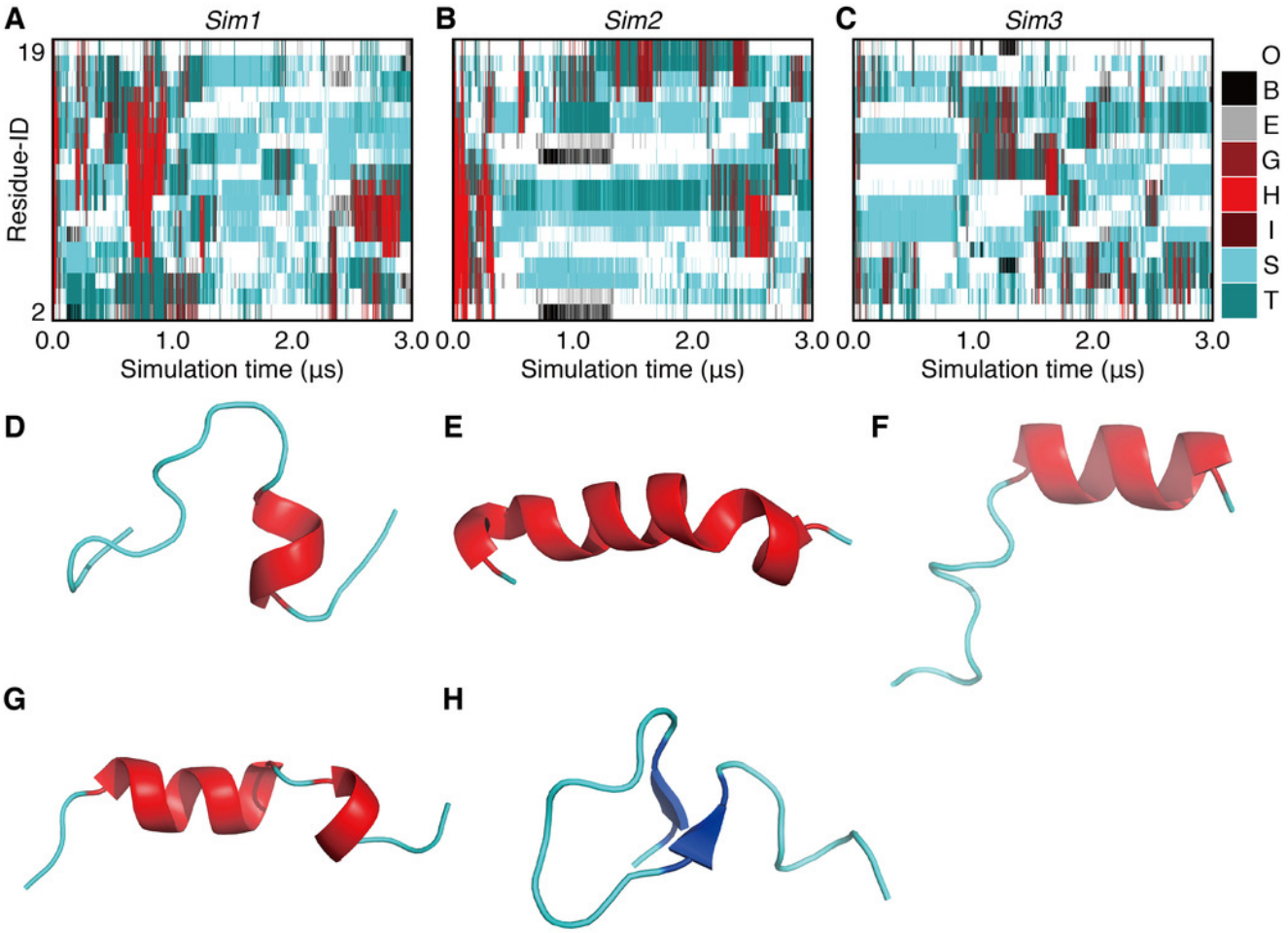
A
Sim1
B
Sim2
C
Sim3
19
Residue-ID
2
0.0
1.0
2.0
3.0
Simulation time (μs)
O
B
E
G
H
I
S
T
D
E
F
G
H

Figure 3

Time course of helix content averaged over accumulated time duration of each trajectory in *Sim1*, *Sim2*, and *Sim3*.

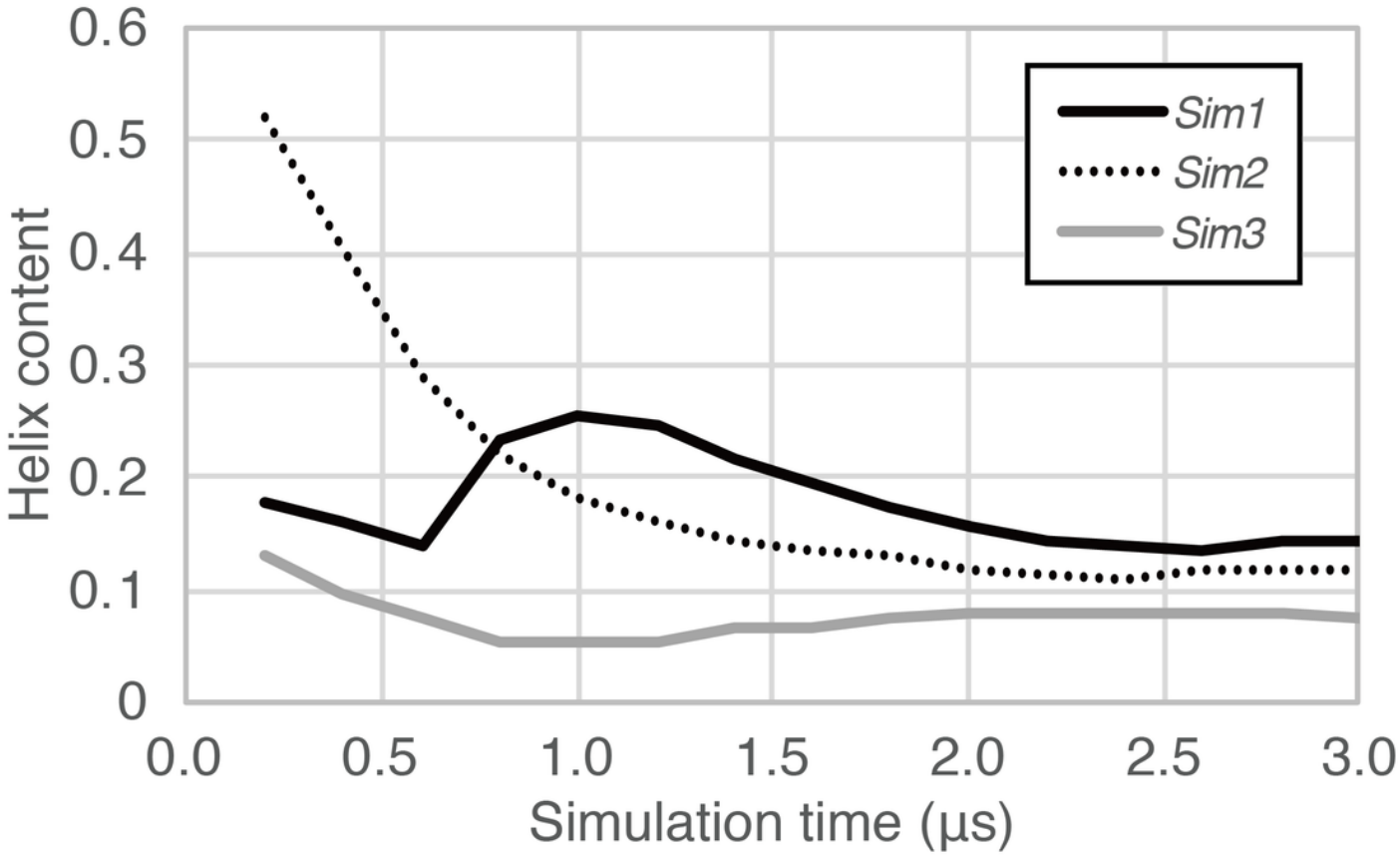

Figure 4

The 400-ns time courses of the secondary structure elements of *Ide1–20* for the panels (A)–(T), respectively.

See the legend of Figures 2A, B, and C.

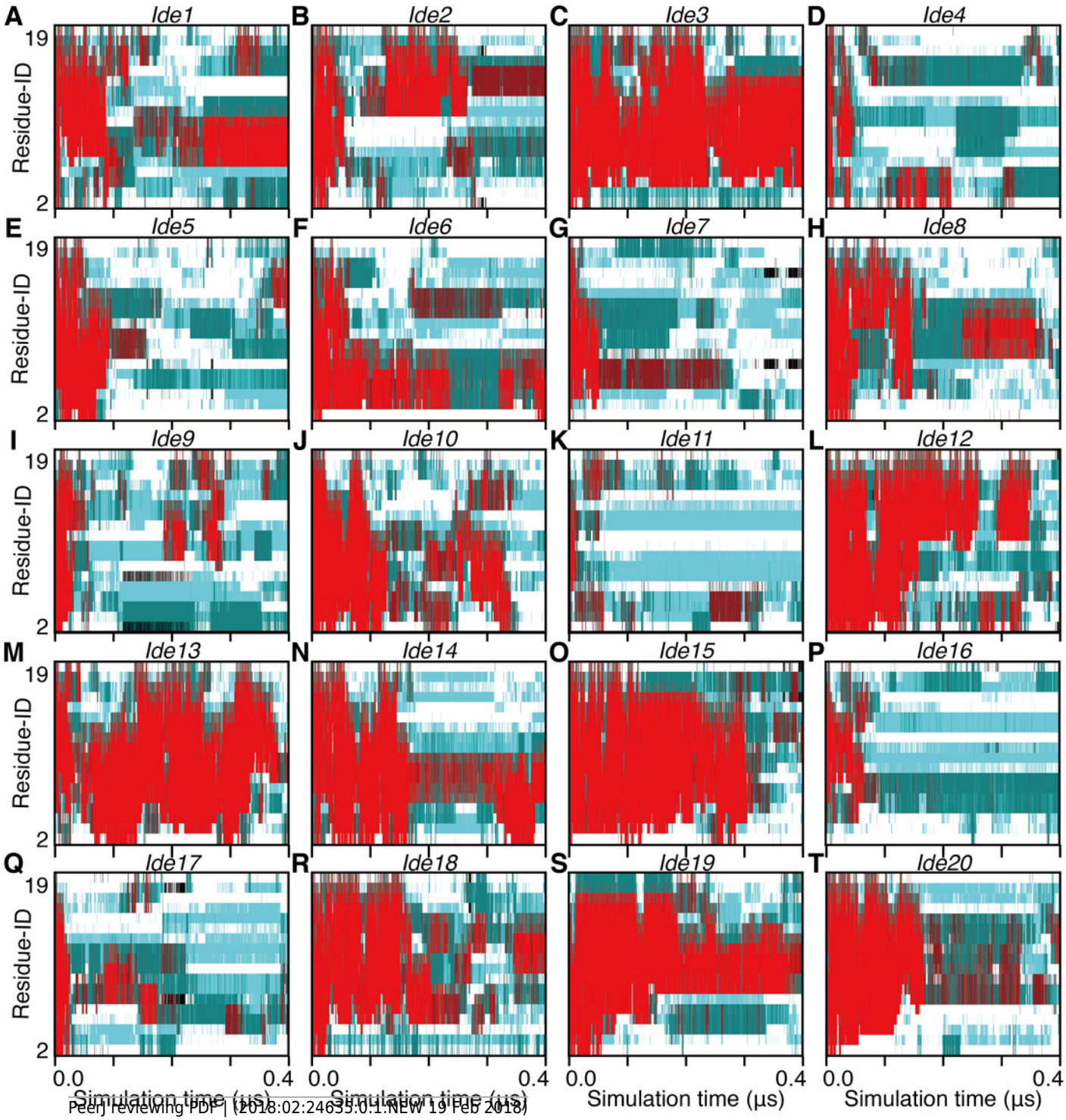

Figure 5

The 400-ns time courses of the secondary structure elements of *Sim1–8* for the panels (A)–(H), respectively.

See the legend of Figures 2A, B, and C.

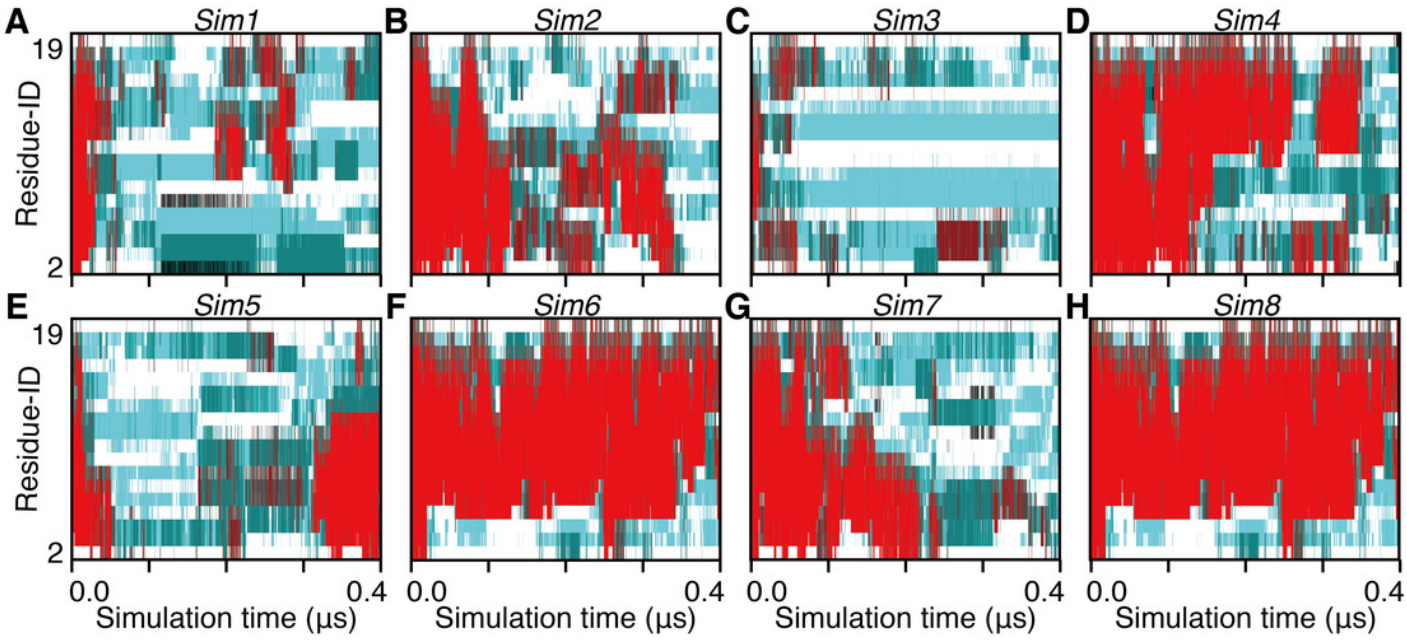

Figure 6

Residue-wise secondary structure content of α-helix (*H*; solid), 3_{10}-helix (*G*; dashed), and turn conformations (*T*; dotted).

(A) The average over 20 *Ide* runs ($P^{Ide}(x; i)$). (B) The average over the 400-ns trajectories of 8 *Sim* runs ($P^{Sim}(x; i)$). (C) The average over 3.0-μs trajectories of *Sim1*, *Sim2*, and *Sim3* ($P^{Sim1-3}(x; i)$).

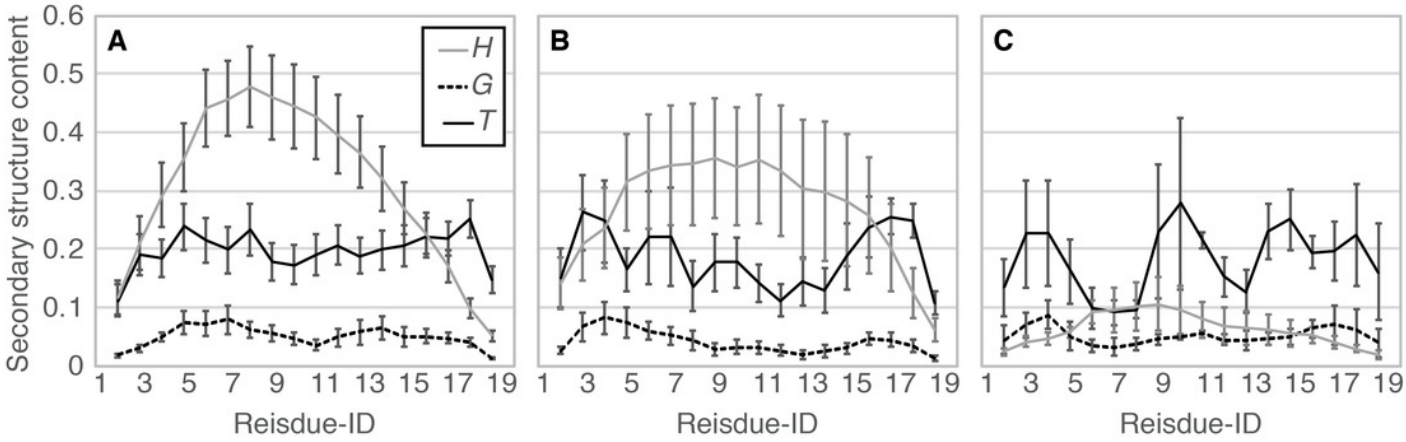

Table 1(on next page)

Unfolding properties of each run.

1 **Table 1.** Unfolding properties of each run.

Run-ID	t_u	Unfolding order	P(H)+P(G)
Ide1	8.52	N,C,M	0.34
Ide2	36.98	C,N,M	0.34
Ide3	88.26	C,N,M	0.64
Ide4	10.30	C,M,N	0.13
Ide5	74.82	C,M,N	0.20
Ide6	47.62	C,N,M	0.32
Ide7	18.10	C,N,M	0.15
Ide8	40.42	C,N,M	0.30
Ide9	23.88	C,N,M	0.14
Ide10	101.34	N,C,M	0.47
Ide11	257.92	C,N,M	0.60
Ide12	16.32	C,M,N	0.30
Ide13	29.52	N,C,M	0.62
Ide14	19.40	N,C,M	0.45
Ide15	249.24	C,N,M	0.60
Ide16	13.02	N,C,M	0.09
Ide17	23.62	C,N,M	0.14
Ide18	35.06	C,N,M	0.44
Ide19	192.74	N,C,M	0.50
Ide20	165.86	N,C,M	0.41
Sim1	31.06	C,N,M	0.15
Sim2	100.52	C,N,M	0.12
Sim3	7.38	C,N,M	0.08
Sim4	79.66	C,M,N	0.51
Sim5	24.38	C,M,N	0.22
Sim6	380.74	N,C,M	0.65
Sim7	87.06	C,N,M	0.35
Sim8	23.22	C,N,M	0.08

2

Table 2(on next page)

Probabilities of secondary structure transitions.

1 **Table 2.** Probabilities of secondary structure transitions, $P^{Ide}(B,A;i)$.

i	2				11				19			
$x \setminus y$	H	G	T	$\overline{HGT}$	H	G	T	$\overline{HGT}$	H	G	T	$\overline{HGT}$
H	0.94	0.01	0.03	0.02	0.96	0.01	0.03	0.00	0.58	0.01	0.34	0.07
G	0.08	0.60	0.22	0.10	0.15	0.55	0.28	0.02	0.04	0.50	0.32	0.14
T	0.04	0.04	0.79	0.13	0.07	0.04	0.88	0.01	0.13	0.03	0.69	0.15
$\overline{HGT}$	0.00	0.00	0.02	0.98	0.00	0.00	0.04	0.95	0.00	0.00	0.03	0.97

2

3

Table 3(on next page)

Probabilities of helix folding and unfolding in *Ide* runs.

[*1] The N-terminal region consisting of the 2nd–7th residues. [*2] The middle region consisting of the 3 8th–13th residues. [*3] The C-terminal region consisting of the 14th–19th residues.

Table 3. Probabilities of helix folding and unfolding in *Ide* runs.

	All	N[*1]	M[*2]	C[*3]
$P^{Ide}(H, HHH)$	0.96	0.97	0.97	0.91
$P^{Ide}(-, HH-)$	0.26	0.16	0.25	0.30
$P^{Ide}(-, -HH)$	0.08	0.06	0.13	0.34
$P^{Ide}(H, H--)$	0.23	0.24	0.22	0.24
$P^{Ide}(H, --H)$	0.04	0.03	0.07	0.05
$P^{Ide}(H, H-H)$	0.09	0.03	0.10	0.08
$P^{Ide}(H, ---)$	0.02	0.02	0.02	0.02

[*1] The N-terminal region consisting of the 2nd–7th residues. [*2] The middle region consisting of the 8th–13th residues. [*3] The C-terminal region consisting of the 14th–19th residues.