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

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

5 [1] Graduate School of Life Sciences, Ritsumeikan University, Noji-Higashi 1-1-1, Kusatsu, Shiga,

6 525-8577, Japan

7 [2] College of Life Sciences, Ritsumeikan University, Noji-Higashi 1-1-1, Kusatsu, Shiga, 525-

8 8577, Japan

9 * To whom correspondence should be addressed:

10 Kota Kasahara, Tel: +81-77-566-1111; Fax: +81-77-561-3729; Email: ktkshr@fc.ritsumei.ac.jp.

11 † The first two authors are considered to be joint first authors.

Abstract

14   Elucidating the molecular mechanism of helix–coil transitions of short peptides is a long-

15   standing conundrum in physical chemistry. Although the helix–coil transitions of poly-glutamic

16   acid (PGA) have been extensively studied, the molecular details of its unfolding process still

17   remain unclear. We performed all-atom canonical molecular dynamics (MD) simulations for a

18   20-residue PGA, over a total of 19 μs, in order to investigate its helix-unfolding processes in

19   atomic resolution. Among the 28 simulations, starting with the α-helical conformation, all

20   showed an unfolding process triggered by the unwinding of terminal residues, rather than by

21   kinking and unwinding of the middle region of the chain. The helix-coil-helix conformation

22   which is speculated by the previous experiments was not observed. Upon comparison between

23   the N- and C-termini, the latter tended to be unstable and easily unfolded. While the probabilities

24   of helix elongation were almost the same among the N-terminal, middle, and C-terminal regions

25   of the chain, unwinding of the helix was enriched at the C-terminal region. The turn and 3_{10}-helix

26   conformations were kinetic intermediates in the formation and deformation of α-helix, consistent

27   with the previous computational studies for Ala-based peptides.

Introduction

31   Elucidation of the molecular mechanisms of protein folding is a central issue in physical

32   chemistry. Since protein folding involves formation of secondary structural elements as building

33   blocks of the tertiary structure (Richardson, 1981), understanding the dynamics of α-helical

34   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

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 enhance

44   helix formation. The reported helix content of short PGAs in acidic environments varied from

45   0.3 to 0.6, whereas it is below detectable limit in neutral pH (Clarke et al., 1999; Kimura et al.,

46   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 dichroism (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

58   the helix–coil transitions of Ala-based peptides, and the C-terminus has been reported to have a

59   higher denaturing tendency compared to the N-terminus (Young & Brooks, 1996; Takano et al.,

60   1999; Wu & Wang, 2001). In addition, the 3_{10}-helix and turn conformations were found to be

61   kinetic intermediates for the helix–coil transitions (Young & Brooks, 1996; Takano et al., 1999).

62   However, unlike that of the Ala-based peptides, helix-coil transitions of PGA peptides have not

63   been studied using the all-atom MD method.

64   Here, we utilized the all-atom canonical MD method to simulate unfolding dynamics of a

65   20-residue PGA with fully protonated side chains, mimicking an acidic environment. Using the

66   molecular model of a PGA with α-helical conformation as the initial structure, we repeated MD

67   simulations for unfolding processes with different initial conditions. In total, 19-μs dynamics,

68   consisting of 3 runs with 3.0 μs and 25 runs with 0.4 μs, were simulated. While various pathways

69   of unfolding were observed in these 28 time courses, PGA unfolding was mainly seen to be

70   triggered by denaturation of the termini, followed by propagation of the coil conformation

71   toward the opposite side. Multiple-helix conformations implied by the previous experiments did

72   not appear in the MD simulations.

73

74   Methods

75   Canonical MD Simulations

76   Dynamics of a 20-residue PGA, in an explicitly solvated periodic boundary cell, was

77   investigated by the canonical MD method. We prepared two α-helical PGA structures as the

78   initial structures for simulation. The first was an α-helical structure, sampled from an ensemble,

79   obtained by our replica-exchange MD (REMD) simulation, with an implicit solvent model. The

80   details of the REMD simulation will be described elsewhere (unpublished data by Iwai R *et al.*).

81   The second was an ideal α-helix, all the residues of which took the backbone dihedral angles $\varphi =$

82   -60° and ψ = -45°, built using *tLEaP* software attached to AMBER package. The N- and C-

83   termini of the PGA were capped with acetyl (Ace) and N-methyl (Nme) groups, respectively. All

84   the carboxyl groups of the side-chains were protonated and net charge of the PGA was zero.

85   Each molecular model of the PGA was placed in the truncated octahedral cell and solvated by

86   filling with TIP3P water molecules (Jorgensen et al., 1983). The number of atoms composing the

87   molecular system with the simulated structure of PGA was 10,592, and that with the ideal α-

88   helix was 11,081. After that, the energy minimizations were successively performed with the

89   steepest descent and conjugate gradient methods; the number of steps was 250 for each. The

90   systems were relaxed via a 200-ps *NPT* simulation using Berendsen barostat. For the system with

91   the ideal helix, the heavy atoms in the PGA were constrained during the relaxation run. The final

92   snapshots of these two systems, referred to as *Sim* and *Ide*, were used as the initial structures of

93   the production runs (Figure 1). Through the *NPT* relaxations, the cell dimensions shrank from

94   54.32 Å to 51.85 Å and from 55.10 Å to 52.66 Å for *Sim* and *Ide*, respectively. The convergence

95   of cell volumes was conformed in terms of the relative standard deviations in the last 100 ps of

96   the *NPT* simulations (*ca*. 0.23 %). As production runs, 8 and 20 runs of simulations were

97   performed with *Sim* and *Ide* systems, respectively. Accordingly, we termed these simulations as

98   *Sim1, Sim2, ..., Sim8*, and *Ide1, Ide2, ..., Ide20*. The initial atomic velocities were randomly

99   generated with different random seeds for each run. The simulation time of each run was 0.4 µs

100   except for *Sim1, Sim2*, and *Sim3* that lasted over 3.0 µs. These production runs were performed

101   with the *NVT* ensemble at 300 K using the Langevin thermostat. The integration time step was

102   2.0 fs; the covalent-bond lengths with hydrogen atoms were constrained with the SHAKE

103   algorithm (Ryckaert, Ciccotti & Berendsen, 1977). The non-bonded pairwise potentials were

104   truncated at 10 Å of the interatomic distance. For the potential energy calculations, AMBER

105   ff99SB force field (Hornak et al., 2006) was applied. All the simulations were carried out using

106   AMBER software.

107

108   Analyses

109   On the basis of the trajectories of the atomic coordinates, recorded every 20 ps in the

110   simulations, the helix–coil transitions of a PGA were analyzed using DSSP software (Kabsch &

111   Sander, 1983). DSSP recognizes the secondary structural elements in terms of hydrogen bonding

112   patterns of the main-chains and categorizes them into the following eight classes: α-helix, 3_{10}-

113   helix, π-helix, extended β-strand, isolated β-bridge, turn, bend, and others. Each class is

114   represented by an alphabetical symbol; *H*, *G*, *I*, *E*, *B*, *T*, *S*, and *O*, respectively. Note that the

115   symbol '*O*' is introduced in this paper for convenience, and it is denoted as ' ' (white or blank

116   space) in the output of the DSSP software. The secondary structure content in the *Ide* trajectories

117   was referred to as $P^{Ide}(x;i)$ for the contents of the secondary structure *x* (any of the eight classes)

118   at the *i*-th residue. The superscript "*Ide*" indicates that the ensemble was obtained from the 20

119   *Ide* runs with 0.4 μs each. The ensemble consisting of trajectories of 8 *Sim* runs with 0.4 μs each

120   is indicated as the superscript "*Sim*", and that of *Sim1–Sim3* with 3.0 μs each is indicated as the

121   superscript "*Sim1–3*". The secondary structure content for the entire chain is presented as $P^{Ide}(x)$.

122   The transition probabilities of *i*-th residue, from the secondary structure *x* to *y* between the

123   successive snapshots (20 ps of the time interval), $P^{Ide}(y,x;i)$, were also evaluated. To measure the

124   time required for the complete unfolding of an α-helix, we defined the unfolding time, t_u, as the

125   time corresponding to the first snapshot without α-helical residues in a trajectory.

126

127   Results

128   Micro-second Dynamics of a PGA

129   In order to investigate long-term behavior of a PGA, we performed three runs of 3.0-μs MD

130   simulations (*Sim1*, *Sim2*, and *Sim3*) with the same initial atomic coordinates but different atomic

131   velocities (Figure 1). The initially formed α-helix was deformed immediately after beginning the

132   simulations in all the three runs (Figure 2). The unfolding times, t_u, defined as the time of the

133   first snapshot without an α-helical residue in PGA for each trajectory, were 31.06 ns, 100.52 ns,

134   and 7.38 ns in *Sim1*, *Sim2*, and *Sim3* simulations, respectively. In the simulation with the longest

135   unfolding time (*Sim2*), after unfolding of the initial α-helix, the helical conformation was

136   temporarily reformed at the N-terminal half of the chain at around 0.2 μs (Figure 2G). However,

137   the reformed helix was unfolded at 0.34 μs, and a helix longer than 13 residues was not formed

138   till the end. In the *Sim1* simulation, although the initial helix was immediately unfolded, a long

139   helix consisting of 17 residues was refolded and retained over sub-micro second time scale

140   (Figure 2E). This helix was nucleated between 12th to 16th residues at 0.62 μs (Figure 2D) and

141   propagated over the range from 2nd to 18th residues. While the N-terminal half of the helix was

142   deformed at 0.84 μs (Figure 2F), the latter half remained intact till 0.95 μs. On the other hand, re-

143   formation of stable helix did not occur in *Sim3*, although several helix-nucleation events were

144   observed. Overall, helix formation was a relatively rare event in this time scale. In addition,

145   while several helix-nucleation events were observed, the nucleated helices disappeared

146 immediately in most cases. Helix nucleation seemed to be coupled with the turn conformation

147 (Figures 2A, B, and C), discussion on which will be taken up later. Formation of a β-sheet was

148 also observed as a rare event. β-sheet formation in *Sim2* was exceptionally stable and was

149 retained during 0.63 μs (Figure 2H).

150   In the time course of the secondary structural elements at each residue (Figure 2A, B, and

151 C), some "bands" could be observed; for example, the turn conformation was almost always

152 formed at the 9th and 10th residues in *Sim2*. Since the tendency to form a turn at the 9th and 10th

153 residues was not observed in the other runs, it is considered to be due to an initial condition,

154 rather than an intrinsic propensity of the 9th and 10th residues. This indicates that there was the

155 strong time-correlation of secondary structure formation, and 3.0 μs was not enough to reach an

156 equilibrium state. The time course of the ensemble average of the helix content (summation over

157 the α- and 3_{10}-helix conformations; $P(H) + P(G)$), for *Sim1*, *Sim2*, and *Sim3* implies that the

158 trajectories were not well-converged (Figure 3). The gain of helix content in 0.5–1.0 μs of *Sim1*

159 corresponds to the refolding of the α-helix mentioned in the previous paragraph (Figure 2A, D, E,

160 F). While the helix content of the three trajectories became converged to similar values with the

161 evolution of time, they still acquired different values at the end of the simulations. The helix

162 content in the full-length trajectories of *Sim1*, *Sim2*, and *Sim3* were 0.14, 0.12, and 0.078,

163 respectively. In addition, the time courses of the end-to-end distance and radius of gyration also

164 showed slow equilibrations of the conformations (Figure S1). These results imply that

165 equilibration of the system requires longer time scales.

166

167 Unfolding Dynamics

168   Non-equilibrium processes involved in the transformation of an α-helix into denatured structures

169   were analyzed by scrutinizing the first part of each trajectory. We additionally performed 25

170   short (400 ns for each) simulations and analyzed the unfolding processes of the 28 simulations in

171   total. Note that 8 of them started from an α-helical conformation obtained from a simulation

172   (*Sim1–Sim8*; Figure 1), and the remaining 20 started from an artificially built ideal α-helix (*Ide1–*

173   *Ide20*; Figure 1). As a result, all the 28 runs showed corruption of the α-helical conformation

174   within 400 ns (Figures 4 and 5). The unfolding times (t_u) varied from 7.38 ns (*Sim3*) to 380.70 ns

175   (*Sim6*), and the average, median, and the standard deviation (SD) were 75.63 ns, 36.02 ns, and

176   92.18 ns, respectively (Table 1). There was no statistically significant difference between *Sim*

177   and *Ide* simulation results; the average (median; SD) of t_u were 72.65 ns (36.02 ns; 79.88 ns) and

178   83.09 ns (37.32 ns; 123.97 ns), for *Sim* and *Ide*, respectively.

179   The unfolding trajectories varied among the 28 trajectories. The fastest unfolding was

180   observed in *Sim3*. The helix deformed from both the N- and C-termini immediately after the

181   simulation began (Figure 5C). As described above, while a single-turn helix sometimes formed

182   at the N- and C-termini after unfolding, they did not grow into a longer helix. The bend

183   conformations were stably formed at the 5th, 6th, 7th, 10th, and 11th residues during 400 ns. On

184   the contrary, *Sim6* showed the slowest dynamics of unfolding. While three or four residues from

185   the N-terminus were immediately deformed, the remaining part of the helix was retained for a

186   long time (Figure 5F). As described above, strong time correlations were observed in all the

187   trajectories (Figures 4 and 5). After immediate unfolding of the α-helix, a denatured

188   conformation of the peptide was not randomized in this time scale.

189   For all the 28 trajectories, unfolding mechanisms were analyzed in terms of the order of

190   deformation for each region in the polypeptide chain. We classified the residues into three

191   regions; *i.e.*, the N-terminal region (2nd–7th residues), the middle region (8th–13th residues),

192   and the C-terminal region (14th–19th residues). The first and 20th residues were discarded

193   because of the following reasons: they would be highly influenced by the truncation of the chain;

194   the main-chain hydrogen bonding pattern of the first residue cannot be defined due to lack of the

195   N-terminal neighbor; all the regions should have the same number of residues. Next, the order of

196   unfolding, for these regions, was assessed based on the helix content of each region in the time

197   period ranging from the beginning of simulation to the unfolding time, t_u. As a result, the

198   unfolding process beginning with the deformation of the middle region was not observed, and all

199   the unfolding processes began with unwinding of one of the terminal regions (Table 1). In

200   addition, coil regions propagated toward both the directions in many cases. There are two

201   possible scenarios for completion of unfolding from any terminus: (i) the coil region appears in a

202   terminus and elongates toward the opposite terminus (“N, M, C” and “C, M, N” in Table 1), and

203   (ii) the opposite terminus is successively unfolded followed by elongation of coil regions from

204   both the termini to the middle (“N, C, M” and “C, N, M” in Table 1). The fact that the former

205   scenario was observed in only 3 and 2 runs among 20 *Ide* and 8 *Sim* runs, respectively, suggests

206   the latter being the major way of α-helix unfolding in this system.

207   When comparing the N- and C-termini of the peptide chain, unfolding from the C-

208   terminus was preferred over that from the N-terminus; 13 out of the 20 *Ide* runs and 7 out of the

209   8 *Sim* runs showed unfolding from the C-terminus. Difference between the two termini was

210   clearer in *Sim* runs than in *Ide* runs, probably because of the slightly distorted initial structure of

211   *Sim* (Figure 1). The ensemble averages of residue-wise α-helix contents in *Ide1–20* with 0.4 μs

212   each ($P^{Ide}(H;i)$), *Sim1–8* with 0.4 μs each ($P^{Sim}(H;i)$), and *Sim1–3* with 3.0 μs each ($P^{Sim1-3}(H;i)$)

213   also showed a lower helical tendency at the C-terminus than at the N-terminus (Figure 6). The

214   previous simulation studies (Young & Brooks, 1996; Wu & Wang, 2001; Finke et al., 2007) had

215   also reported that helix formation of the C-terminal residues was unstable compared to that of the

216   N-terminal ones.

217

218   Secondary Structural Transitions

219   To analyze the detailed mechanisms of conformational transitions in shorter time scales, we

220   assessed the probability of the event that the i-th residue in the secondary structure A at time t is

221   transformed into B at time $t + 20$ ps; the averaged probability over the 20 Ide runs is referred to

222   as $P^{Ide}(B,A;i)$. For simplicity, we focused on the four classes of secondary structure elements; H,

223   G, T, and $\overline{HGT}$, which means any of the other five structural elements (I, E, S, B, and O). The

224   cases $i = 2$, 11, and 19 were analyzed as representatives of the N-terminal, middle, and C-

225   terminal residues, respectively (Table 2). The C-terminal residues showed a weaker tendency to

226   retain the α-helical conformation compared to the other residues ($P^{Ide}(H,H;2) = 0.94$, $P^{Ide}(H,H;11)$

227    $= 0.96$, and $P^{Ide}(H,H;19) = 0.58$). The weaker tendency to retain the same conformation in the C-

228   terminal region was also observed in the other secondary structures. The results of Sim runs were

229   qualitatively consistent with that of Ide runs (Table S1).

230   The helix–coil transitions mainly occurred via the turn conformation. More than half of

231   the conformational transitions from the α-helix directed to the turn conformation; $P^{Ide}(T,H;i)/P^{Ide}($

232    $\overline{H},H;i$) for $i = 2$, 11, and 19 were 0.52, 0.73, and 0.80, respectively, where $\overline{H}$ denotes the

233   secondary structure other than H. In addition, formation of the α-helix via turn was enriched in

234   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,

235   respectively. Thus, the turn conformation can be considered as an intermediate state in the helix–

236   coil transition, especially at the C-terminus. Another intermediate in the α-helix formation is the

237   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 $i =$

238   2, 11, and 19 were 0.55, 0.62, and 0.64, respectively), it also transformed into an α-helix,

239   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

240   0.076, respectively (Supplementary Figure S2). This result agreed with the previous theoretical

241   studies, which reported that the 3_{10}-helix is not a thermodynamic intermediate but could be a

242   kinetic intermediate (Young & Brooks, 1996; Wu & Wang, 2001).

243   In addition to the position of amino acids in the polypeptide chain, effect of the α-helical

244   ends was analyzed. We focused on segments consisting of three consecutive residues in the chain,

245   and the state of the segment was defined as the combination of secondary structures of the three

246   residues, grouped into the two classes, *i.e.*, α-helix ("*H*") and others ("*-*"; it has the same

247   meaning as "$\overline{H}$"). The state of a segment was divided into the following eight classes: "*HHH*", "

248   *HH-*", "*-HH*", "*H-H*", "*H--* ", "*--H*", and "*---*". The state "*-H-*" is impossible, becauseα-helical

249   conformation coincides with at least four consecutive residues. The probability of the event that

250   the central residue of a segment forms an α-helix at the next snapshot (20 ps later) was analyzed

251   for each class. For instance, probability for the class "*HH-*", denoted as $P^{Ide}(H, HH\text{-})$, means the

252   probability to retain α-helical conformation for the residue at the C-terminal end of an α-helix,

253   regardless of the position in the chain (*i*). The probability of deformation of the C-terminal end

254   of an α-helix can be shown as $P^{Ide}(\text{-}, HH\text{-}) = 1 - P^{Ide}(H, HH\text{-})$. The probabilities are summarized in

255   Table 3; the case of *Sim* runs is shown in Table S2. We found that a residue at the interior of an

256   α-helix was more stable to maintain the α-helical conformation, compared to the terminal

257   residues; $P^{Ide}(H, HHH) = 0.97$. It is noteworthy that the C-terminal end of an α-helix is more

258   frequently deformed than the N-terminal one; $P^{Ide}(H, HH\text{-}) = 0.74$ and $P^{Ide}(H, \text{-}HH) = 0.92$. In

259   addition, α-helix elongation toward the C-terminus was enriched compared to that toward the

260   opposite direction; $P^{Ide}(H, H--) = 0.23$ and $P^{Ide}(H, --H) = 0.04$. The C-terminal end of an α-helix

261   unstably changed its conformation while the N-terminal end tended to retain its conformation.

262   We also evaluated the relationship between the two definitions of position; position in an

263   α-helix (the N-terminal end, internal, and the C-terminal end) and position in the polypeptide

264   chain (the N-terminal region $[2 \leq i \leq 7]$, middle region $[8 \leq i \leq 13]$, and C-terminal region $[14 \leq i$

265    $\leq 19]$). The probability of helix–coil transitions in the center of a three-residue segment x was

266   assessed for each of the three regions y: $P^{Ide}(-,x;y) = 0.04$, where x is "$HH-$" or "$-HH$" for the C-

267   and N-terminal ends of an α-helix, respectively, and y is any of "N", "M", and "C", for the N-

268   terminal, middle, and C-terminal regions, respectively. The probabilities to unfold the N- and C-

269   terminal ends of an α-helix varied with respect to the position of the ends in the entire chain;

270   namely, higher probabilities were observed in the C-terminal region of the peptide chain (P^{Ide}

271    $(-, HH-;C) > P^{Ide}(-, HH-;N)$ and $P^{Ide}(-, -HH;C) > P^{Ide}(-, -HH;N)$ in Table 3). While residue-wise α-

272   helical content (Figure 6) and α-helix retention probability (Table 2) indicate the highest α-

273   helical propensity for the middle region, the lowest probabilities to unfold the ends of α-helix

274   were found in the N-terminal region. In contrast, probabilities for elongation of an α-helix were

275   almost the same for all the three regions (see $P^{Ide}(H, H--)$ and $P^{Ide}(H, --H)$ in Table 3). Therefore,

276   an α-helical PGA tended to unfold from the C-terminus.

277   On the other hand, the α-helix nucleation was observed in low probabilities regardless of

278   positions in the chain; $P^{Ide}(H, ---) = 0.02$ for all three regions.

279

280   Discussion

281   In this study, we examined the dynamics of a 20-residue PGA with 28 runs of all-atom canonical

282   MD simulations. While three of them simulated 3.0-μs time courses, the systems were not well-

283   equilibrated (Figure 3) and complete refolding of the α-helix was not observed (Figures 4 and 5).

284   The time scale required for α-helix formation by PGA, still remains controversial. The suggested

285   time-scale varies from sub-micro to milliseconds (Clarke et al., 1999; Kimura et al., 2002;

286   Causgrove & Dyer, 2006; Qin et al., 2014). Our simulation results imply that a time range of few

287   micro-seconds is too short to refold PGA in acidic environments.

288   We mainly focused on the non-equilibrium dynamics of unfolding processes and repeated

289   28 runs of simulations with the two different initial α-helical structures. The results from these

290   two initial structures were qualitatively similar. Higher stability of the α-helical conformation

291   was shown to be in the middle of the polypeptide chain than at the termini. All the unfolding

292   processes of the α-helix began from a terminus, but a helix-coil-helix conformation was not

293   stably observed. In many cases, the unfolding proceeded toward both directions, rather than

294   starting from a terminus and ending at the opposite. In addition, unfolding from the C-terminal

295   side was preferred over that from the N-terminal side (Table 1). The probability of retention of α-

296   helix at each residue was lower in the C-terminus than in the N-terminus (Table 2). While the

297   probabilities of α-helix elongation were almost the same irrespective of whether the end was

298   located at the N-terminus, middle, or C-terminus of the polypeptide chain, the probabilities of

299   unwinding of the α-helix tended to be higher at the C-terminus of the chain (Table 3). The

300   instability of α-helix at the C-terminus was due to the enhancement of unfolding, rather than

301   reduction of folding. In the process of folding and unfolding of the α-helices, the turn and 3_{10}-

302   helix conformations can be kinetic intermediates as consistent to the precedent studies (Young &

303   Brooks, 1996; Wu & Wang, 2001; Pal, Chakrabarti & Basu, 2003).

304   Despite the wide acceptance of the all-atom MD method, there are still some issues under

305   consideration. First, treatment of denatured proteins has not been fully validated in current force

306   fields. Underestimation of the radius of gyration of denatured proteins by standard force fields

307   and water models has been previously reported (Piana, Klepeis & Shaw, 2014). While there is no

308   gold standard yet, some improved force fields and water models have been proposed to simulate

309   denatured proteins (Piana et al., 2015; Henriques & Skepö, 2016; Huang et al., 2016). Second,

310   although the force field applied here, AMBER ff99SB, is one of the standard force fields, there

311   are some reports about its weakness; *e.g.*, underestimation of helix stability (Sorin & Pande,

312   2005) and discrepancy with the quantum mechanical calculations (Takano, Kusaka & Nakamura,

313   2016). Third, finite-size effects have been reported for the helix-stability of a model polypeptide

314   (Weber, Hünenberger & McCammon, 2000; Kastenholz & Hünenberger, 2004; Reif et al., 2009;

315   Kasahara, Sakuraba & Fukuda, 2018). To avoid this problem, we used the large periodic

316   boundary cells, which have at least a 10-Å margin between the solute termini and the cell

317   boundaries, and the cell size was well equilibrated via the *NPT* simulations.

318   In fact, helix content in the simulated ensembles (Figures 3 and 6) were lower than the

319   experimentally reported values, which is in the range of 0.3 to 0.6. The ensemble averages [and

320   SD] of end-to-end distances (19.11 [8.17], 19.57 [7.61], and 15.71 [5.85] Å for *Ide*, *Sim*, and

321   *Sim1–3*, respectively) were inconsistent with the FRET measurements by Finke *et al.*, which

322   were 23–24 Å at pH 4 (Finke et al., 2007). However, differences in the experimental method and

323   conditions may cause differences in the helix content (Kimura et al., 2002), since precise

324   measurement of the latter for short peptides is not straight forward (Kelly, Jess & Price, 2005;

325   Greenfield, 2007). Discussion on the quantitative aspects of the results, *e.g.*, helix contents and

326   folding kinetics, provided by both the experimental and theoretical methods in this study, should

327   be carefully considered. From qualitative aspects, our results were consistent with the reported

328   theoretical studies, in spite of several differences in the materials and methods, *e.g.*, peptide

329   sequence, parameters, and sampling methods. For example, the weaker helix formation

330   propensity at the C-terminus and the kinetic intermediates of helix–coil transitions were

331   consistently concluded from this study in agreement with the previous theoretical studies. They

332   are robust conclusions, regardless of adjustable settings and simulation methods. In addition to

333   that, our simulation results provide statistics of kinetic details of helix–coil transition by multiple

334   runs of canonical MD. The weaker helix formation propensity at the C-terminus is due to high

335   frequency of unwinding rather than disfavoring of folding. Helix-coil-helix conformations

336   speculated by previous experiments were not observed.

337   Note that effects of peptide length, which is one of the most important determinants of

338   the helix–coil transitions of polypeptide, is not analyzed in this study. In general, the microscopic

339   behavior of peptides depends on peptide length (Gómez-Sicilia et al., 2015). While this study

340   focused only on the behavior of 20-residue PGA by following the previous study by Finke *et al*

341   (Finke et al., 2007), some other previous experiments reported the effects of the length of PGA;

342   For example, Clarke *et al.* examined 34-, 57-, and 163-residue PGAs (Clarke et al., 1999),

343   Kimura *et al*. used 34- and 190-residue PGAs (Kimura et al., 2002), and Donten and Hamm used

344   20-, 50-, and 440-residue PGAs (Donten & Hamm, 2013). They demonstrated that longer PGAs

345   tend to have slower folding kinetics and higher helix contents. For future works, simulating

346   systems with longer PGAs would be useful for understanding the molecular mechanisms of

347   effects of peptide length.

348

Conclusion

350 In this study, the unfolding mechanism of α-helix in 20-residue PGA was investigated using all-

351 atom canonical MD simulations. Our results suggested that the unfolding was triggered by

352 unwinding of a terminus, whereas the multiple short-helical conformations, implied in the

353 previous experiments (Clarke et al., 1999; Kimura et al., 2002; Finke et al., 2007), were not

354 stably observed in the simulated trajectories within the micro-second time-scale. The instability

355 of C-terminus is consistent with the previously reported result from generalized ensemble

356 simulations of the poly-Ala peptides (Young & Brooks, 1996; Takano et al., 1999; Wu & Wang,

357 2001). The mechanism of helix–coil transitions, shown here, might reflect the nature of the

358 peptide backbone, and provide insight into the helix–coil transitions for general cases of

359 polypeptides.

360

Acknowledgements

362 The supercomputer resources were provided by the HPCI System Research Projects (Project IDs:

363 hp170020 and hp170025) and the National Institute of Genetics, Research Organization of

364 Information and Systems, Japan. We thank Tomoya Hirano for help with data analyses.

365

366 REFERENCES

367 Baldwin RL 1995. α-Helix formation by peptides of defined sequence. *Biophysical chemistry*
368   55:127–135. DOI: 10.1016/0301-4622(94)00146-B.
369 Causgrove TP, Dyer RB 2006. Nonequilibrium protein folding dynamics: laser-induced pH-jump
370   studies of the helix–coil transition. *Chemical Physics* 323:2–10. DOI:
371   10.1016/j.chemphys.2005.08.032.

372 Chen Y, Zhou Y, Ding J 2007. The helix–coil transition revisited. *Proteins: Structure, Function,*
373   *and Bioinformatics* 69:58–68. DOI: 10.1002/prot.21492.
374 Clarke DT, Doig AJ, Stapley BJ, Jones GR 1999. The alpha-helix folds on the millisecond time
375   scale. *Proceedings of the National Academy of Sciences of the United States of America*
376   96:7232–7237. DOI: 10.1073/pnas.96.13.7232.
377 Donten ML, Hamm P 2013. pH-jump induced α-helix folding of poly-l-glutamic acid. *Chemical*
378   *Physics* 422:124–130. DOI: 10.1016/j.chemphys.2012.11.023.
379 Finke JM, Jennings PA, Lee JC, Onuchic JN, Winkler JR 2007. Equilibrium unfolding of the
380   poly(glutamic acid)20 helix. *Biopolymers* 86:193–211. DOI: 10.1002/bip.20719.
381 Gooding EA, Sharma S, Petty SA, Fouts EA, Palmer CJ, Nolan BE, Volk M 2013. pH-dependent
382   helix folding dynamics of poly-glutamic acid. *Chemical Physics* 422:115–123. DOI:
383   10.1016/j.chemphys.2012.11.009.
384 Gómez-Sicilia À, Sikora M, Cieplak M, Carrión-Vázquez M 2015. An Exploration of the
385   Universe of Polyglutamine Structures. *PLoS computational biology* 11:e1004541. DOI:
386   10.1371/journal.pcbi.1004541.
387 Greenfield NJ 2007. Using circular dichroism spectra to estimate protein secondary structure.
388   *Nature Protocols* 1:2876–2890. DOI: 10.1038/nprot.2006.202.
389 Henriques J, Skepö M 2016. Molecular Dynamics Simulations of Intrinsically Disordered
390   Proteins: On the Accuracy of the TIP4P-D Water Model and the Representativeness of
391   Protein Disorder Models. *Journal of Chemical Theory and Computation* 12:3407–3415.
392   DOI: 10.1021/acs.jctc.6b00429.
393 Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C 2006. Comparison of
394   multiple Amber force fields and development of improved protein backbone parameters.
395   *Proteins: Structure, Function, and Bioinformatics* 65:712–725. DOI: 10.1002/prot.21123.
396 Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, Grubmüller H, Mackerell AD
397   2016. CHARMM36m: an improved force field for folded and intrinsically disordered
398   proteins. *Nature Methods* 14:71–73. DOI: 10.1038/nmeth.4067.
399 Inoue K, Baden N, Terazima M 2005. Diffusion Coefficient and the Secondary Structure of
400   Poly-l-glutamic Acid in Aqueous Solution. *The Journal of Physical Chemistry B* 109:22623–
401   22628. DOI: 10.1021/jp052897y.
402 Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML 1983. Comparison of simple
403   potential functions for simulating liquid water. *The Journal of chemical physics* 79:926–935.
404   DOI: 10.1063/1.445869.
405 Kabsch W, Sander C 1983. Dictionary of protein secondary structure: Pattern recognition of
406   hydrogen-bonded and geometrical features. *Biopolymers* 22:2577–2637. DOI:
407   10.1002/bip.360221211.
408 Kasahara K, Sakuraba S, Fukuda I 2018. Enhanced Sampling of Molecular Dynamics
409   Simulations of a Polyalanine Octapeptide: Effects of the Periodic Boundary Conditions on
410   Peptide Conformation. *The Journal of Physical Chemistry B* 122:2495–2503. DOI:
411   10.1021/acs.jpcb.7b10830.
412 Kastenholz MA, Hünenberger PH 2004. Influence of Artificial Periodicity and Ionic Strength in
413   Molecular Dynamics Simulations of Charged Biomolecules Employing Lattice-Sum
414   Methods. *The Journal of Physical Chemistry B* 108:774–788. DOI: 10.1021/jp0350924.
415 Kelly SM, Jess TJ, Price NC 2005. How to study proteins by circular dichroism. *Biochimica et*
416   *Biophysica Acta (BBA) - Proteins and Proteomics* 1751:119–139. DOI:
417   10.1016/j.bbapap.2005.06.005.

418 Kimura T, Takahashi S, Akiyama S, Uzawa T, Ishimori K, Morishima I 2002. Direct
419   Observation of the Multistep Helix Formation of Poly- l-glutamic Acids. *Journal of the*
420   *American Chemical Society* 124:11596–11597. DOI: 10.1021/ja026639f.
421 Nakamura H, Wada A 1981. Dielectric studies of aqueous solutions of poly(L-glutamic acid).
422   *Biopolymers* 20:2567–2582. DOI: 10.1002/bip.1981.360201207.
423 Neumaier S, Reiner A, Büttner M, Fierz B, Kiefhaber T 2013. Testing the diffusing boundary
424   model for the helix-coil transition in peptides. *Proceedings of the National Academy of*
425   *Sciences* 110:12905–12910. DOI: 10.1073/pnas.1303515110.
426 Pal L, Chakrabarti P, Basu G 2003. Sequence and Structure Patterns in Proteins from an Analysis
427   of the Shortest Helices: Implications for Helix Nucleation. *Journal of Molecular Biology*
428   326:273–291. DOI: 10.1016/S0022-2836(02)01338-4.
429 Piana S, Donchev AG, Robustelli P, Shaw DE 2015. Water Dispersion Interactions Strongly
430   Influence Simulated Structural Properties of Disordered Protein States. *The Journal of*
431   *Physical Chemistry B* 119:5113–5123. DOI: 10.1021/jp508971m.
432 Piana S, Klepeis JL, Shaw DE 2014. Assessing the accuracy of physical models used in protein-
433   folding simulations: quantitative evidence from long molecular dynamics simulations.
434   *Current opinion in structural biology* 24:98–105. DOI: 10.1016/j.sbi.2013.12.006.
435 Qin Z-J, Shimizu A, Li J, Ikeguchi M, Shinjo M, Kihara H 2014. α-helix formation rate of
436   oligopeptides at subzero temperatures. *BIOPHYSICS* 10:9–13. DOI:
437   10.2142/biophysics.10.9.
438 Reif MM, Kräutler V, Kastenholz MA, Daura X, Hünenberger PH 2009. Molecular Dynamics
439   Simulations of a Reversibly Folding β-Heptapeptide in Methanol: Influence of the Treatment
440   of Long-Range Electrostatic Interactions. *The Journal of Physical Chemistry B* 113:3112–
441   3128. DOI: 10.1021/jp807421a.
442 Richardson JS 1981. The anatomy and taxonomy of protein structure. *Advances in protein*
443   *chemistry* 34:167–339.
444 Ryckaert JP, Ciccotti G, Berendsen HJC 1977. Numerical integration of the cartesian equations
445   of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of*
446   *Computational Physics* 23:327–341. DOI: 10.1016/0021-9991(77)90098-5.
447 Sorin EJ, Pande VS 2005. Exploring the Helix-Coil Transition via All-Atom Equilibrium
448   Ensemble Simulations. *Biophysical Journal* 88:2472–2493. DOI:
449   10.1529/biophysj.104.051938.
450 Spek EJ, Olson CA, Shi Z, Kallenbach NR 1999. Alanine Is an Intrinsic α-Helix Stabilizing
451   Amino Acid. *Journal of the American Chemical Society* 121:5571–5572. DOI:
452   10.1021/ja990056x.
453 Stanley CB, Strey HH 2008. Osmotically Induced Helix-Coil Transition in Poly(Glutamic Acid).
454   *Biophysical Journal* 94:4427–4434. DOI: 10.1529/biophysj.107.122705.
455 Takano M, Yamato T, Higo J, Suyama A, Nagayama K 1999. Molecular Dynamics of a 15-
456   Residue Poly(l-alanine) in Water: Helix Formation and Energetics. *Journal of the American*
457   *Chemical Society* 121:605–612. DOI: 10.1021/ja982919c.
458 Takano Y, Kusaka A, Nakamura H 2016. Density functional study of molecular interactions in
459   secondary structures of proteins. *BIOPHYSICS* 13:27–35. DOI: 10.2142/biophysico.13.0_27.
460 Weber W, Hünenberger PH, McCammon JA 2000. Molecular Dynamics Simulations of a
461   Polyalanine Octapeptide under Ewald Boundary Conditions: Influence of Artificial
462   Periodicity on Peptide Conformation. *The Journal of Physical Chemistry B* 104:3668–3675.
463   DOI: 10.1021/jp9937757.

464 Wu X, Wang S 2001. Helix Folding of an Alanine-Based Peptide in Explicit Water. *The Journal*
465   *of Physical Chemistry B* 105:2227–2235. DOI: 10.1021/jp004048a.
466 Young WS, Brooks CL 1996. A microscopic view of helix propagation: N and C-terminal helix
467   growth in alanine helices. *Journal of Molecular Biology* 259:560–572. DOI:
468   10.1006/jmbi.1996.0339.

469
470
471 Figure Legends
472

473 **Figure 1.** The initial structures of MD simulations. (A) The structure built by a REMD

474 simulation, termed *Sim*. (B) The structure based on the ideal α-helix, termed *Ide*. (C) φ-ψ angles

475 of 2nd–20th residues in *Sim* (triangles; the open triangles indicate the 2nd and 20th) and *Ide*

476 (circle; all residues have the same values).

477 **Figure 2.** The 3.0-μs time courses of the secondary structure elements and examples of

478 snapshots for *Sim1*, *Sim2*, and *Sim3* simulations. (A, B, and C) The time courses for *Sim1*, *Sim2*,

479 and *Sim3*, respectively. The horizontal axis is the simulation time, and the vertical axis indicates

480 the amino acid position in the peptide chain. Each block is filled by one of the eight types of

481 colors regarding the secondary structure elements *H, G, I, E, B, T, S*, and *O*, are indicated as red,

482 maroon, dark-red, gray, black, dark-cyan, cyan, and white, respectively. (D, E, F, G, and H)

483 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

484 *Sim2*, and (H) 1.000 μs in *Sim2*.

485 **Figure 3.** The time course of helix content averaged over accumulated time duration of each

486 trajectory in *Sim1*, *Sim2*, and *Sim3*.

487 **Figure 4.** The 400-ns time courses of the secondary structure elements of *Ide1–20* for the panels

488 (A)–(T), respectively. See the legend of Figures 2A, B, and C.

489 **Figure 5.** The 400-ns time courses of the secondary structure elements of *Sim1–8* for the panels

490 (A)–(H), respectively. See the legend of Figures 2A, B, and C.

491   **Figure 6.** Residue-wise secondary structure content of α-helix (H; solid), 3_{10}-helix (G; dashed),

492   and turn conformations (T; dotted). (A) The average over 20 *Ide* runs ($P^{Ide}(x;i)$). (B) The average

493   over the 400-ns trajectories of 8 *Sim* runs ($P^{Sim}(x;i)$). (C) The average over 3.0-μs trajectories of

494   *Sim1*, *Sim2*, and *Sim3* ($P^{Sim1-3}(x;i)$).

495   **Supplementary Figure S1.** Time course of (A) end-to-end distance and (B) radius of gyration of

496   PGA in *Sim1* (red), *Sim2* (green), and *Sim3* (blue) runs.

497   **Supplementary Figure S2.** Probabilities of transition between different secondary structures in

498   *Ide* runs. The horizontal axis indicates the position of each residue, i. Transitions from α-helix

499   (H), 3_{10}-helix (G), and turn (T) are shown in solid, dashed, and dotted lines, respectively.

500   Destination states are indicated as red, maroon, dark-cyan, and green, for H, G, T, and structural

501   elements other than these, respectively.

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).

1447

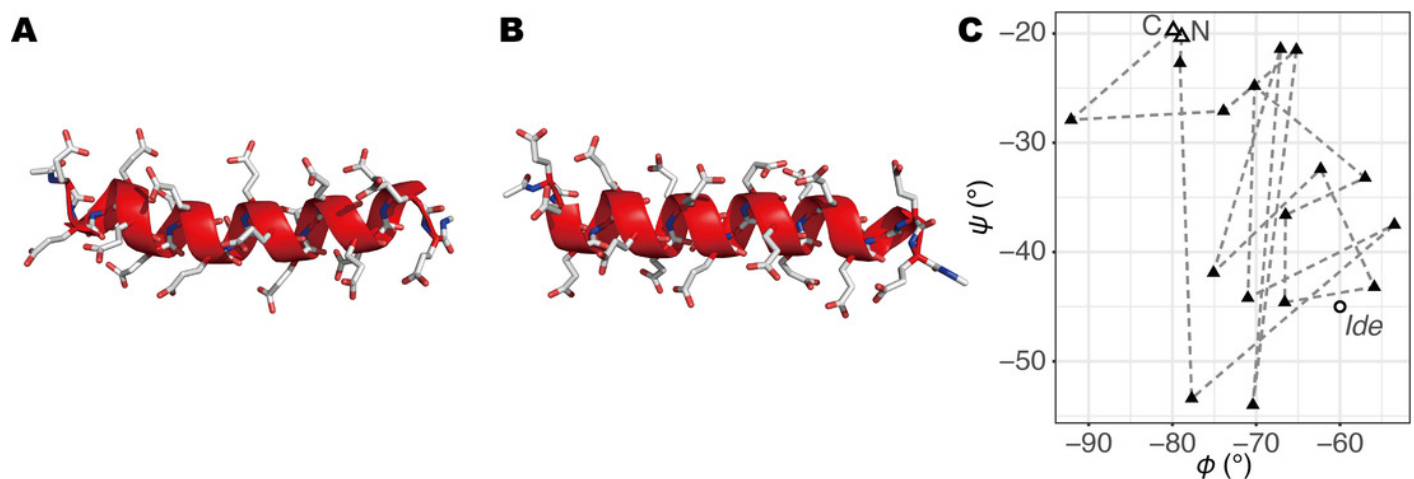



Figure 2

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*, are indicated 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*.

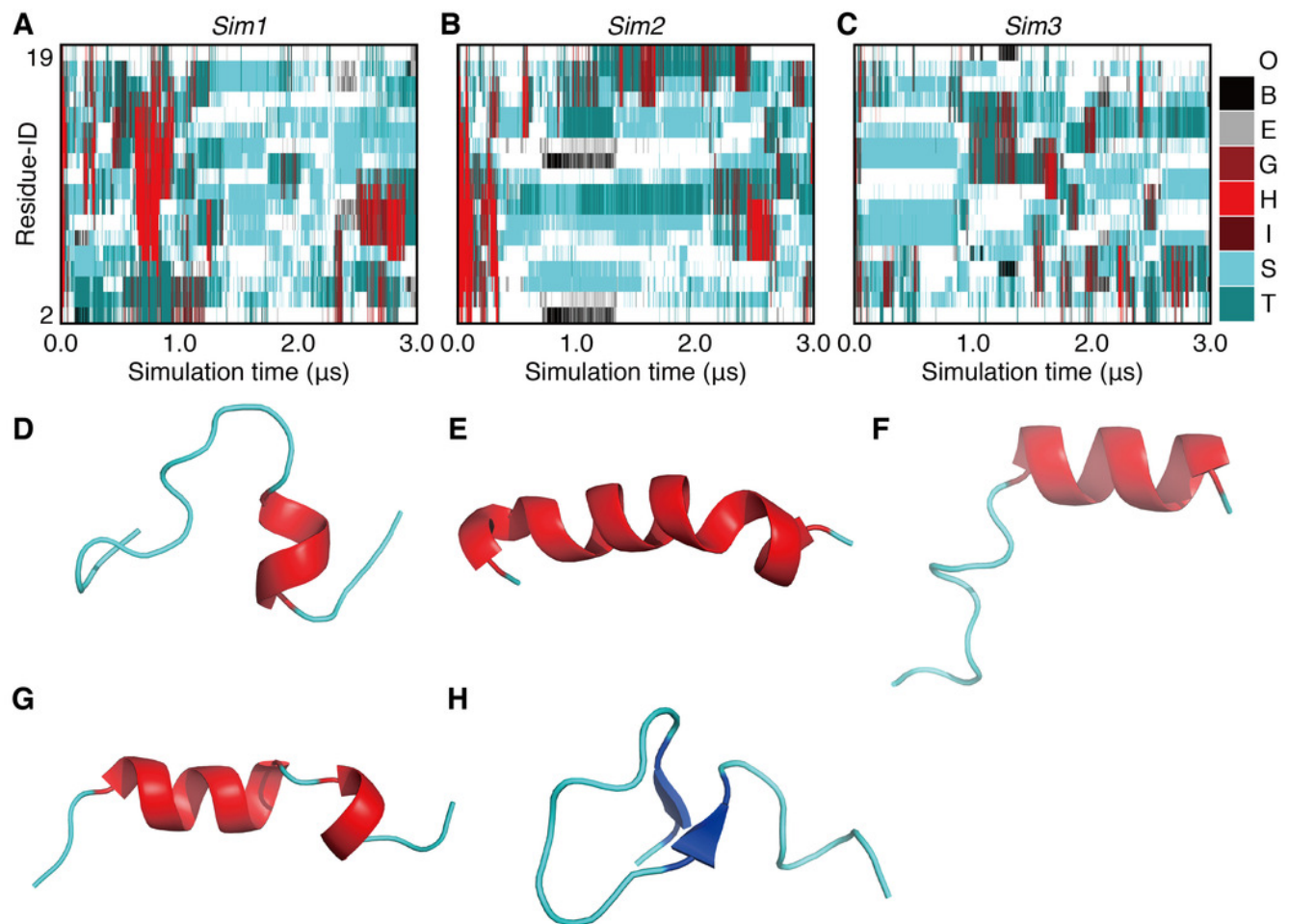



Figure 3

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

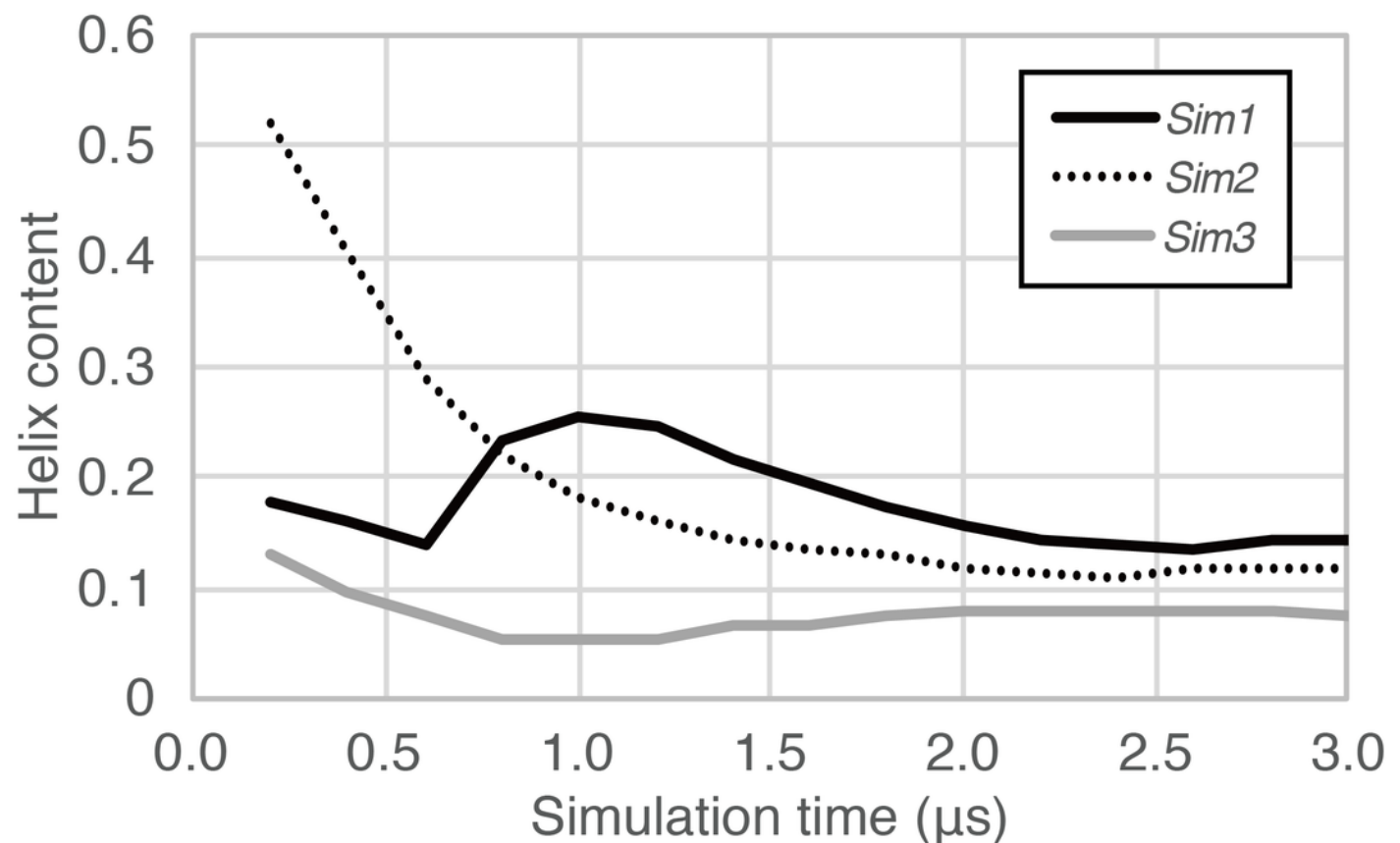



Figure 4

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.

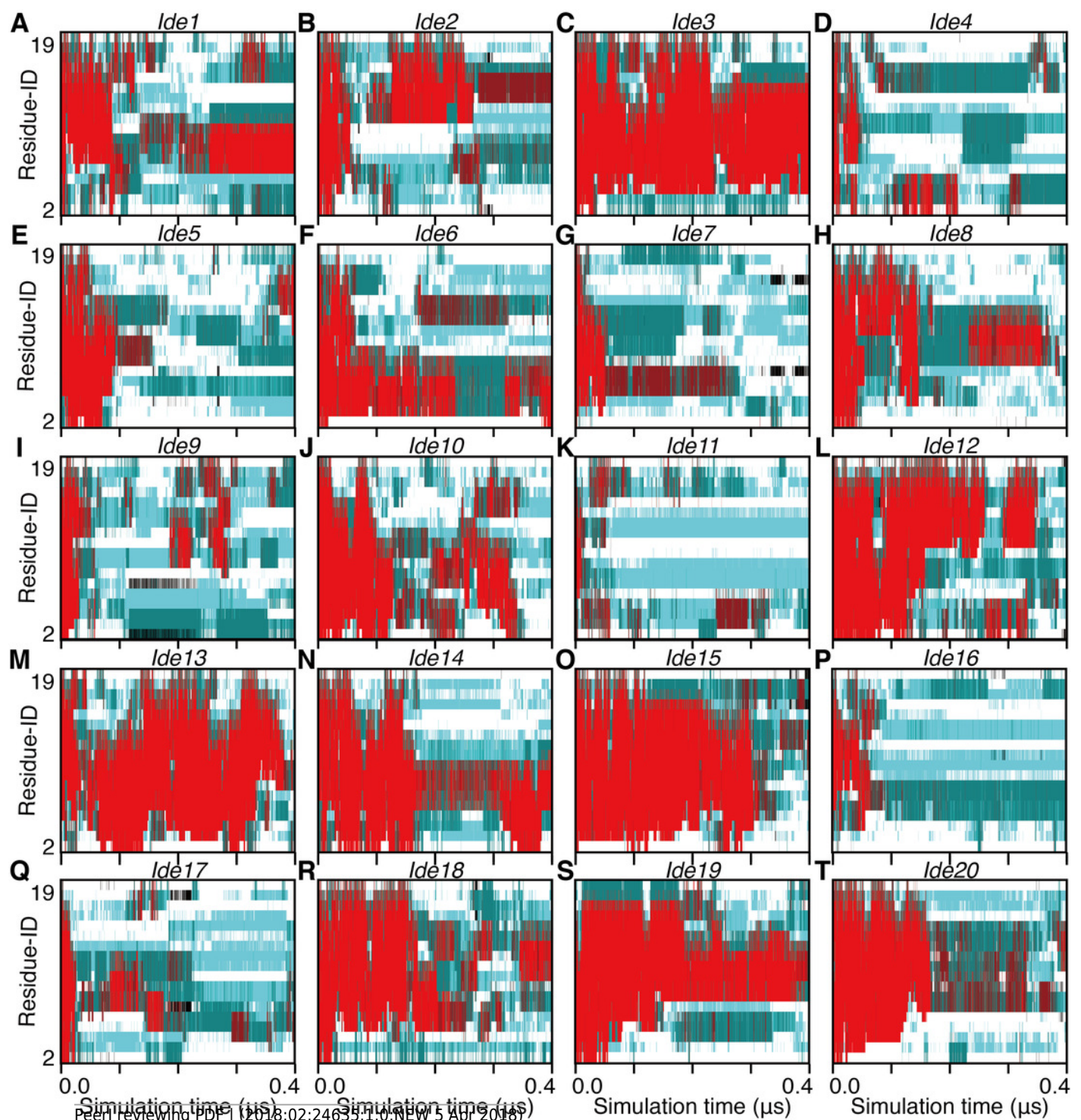



Figure 5

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.

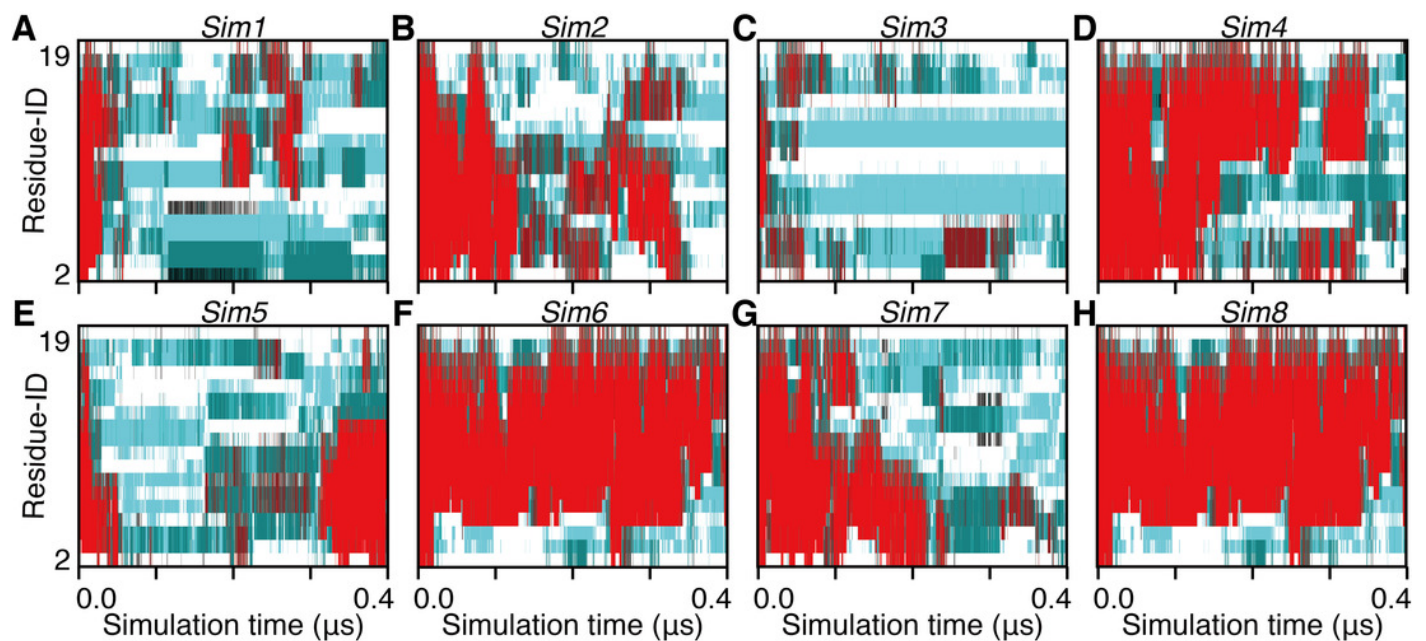
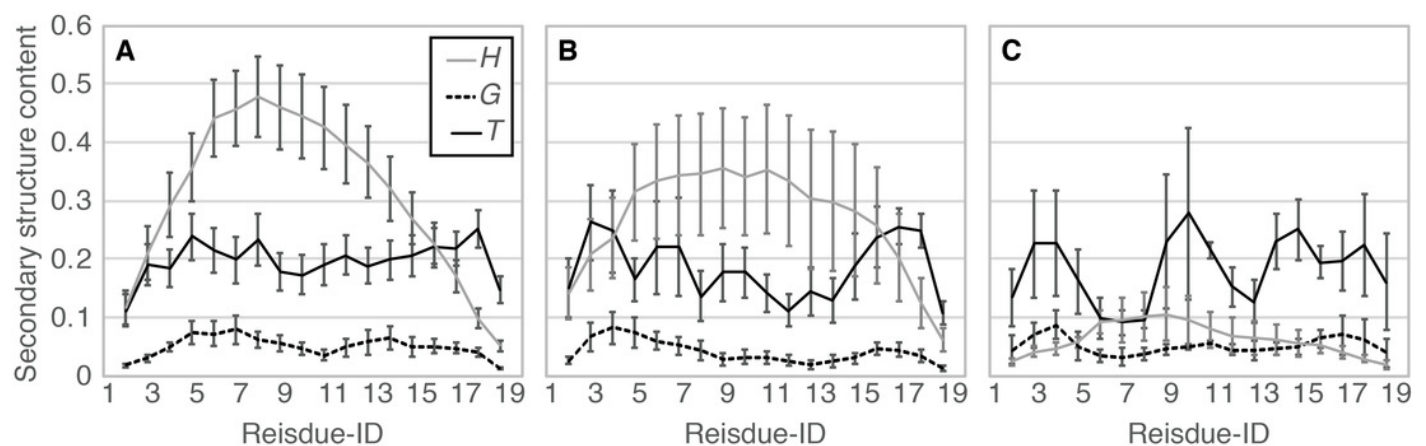



Figure 6

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* ($P^{Sim1-3}(x; i)$).

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

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

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.

1   **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

2   [*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.