Structure Elucidation

DOI: 10.1002/ange.201105356

# Two Different Packing Arrangements of Antiparallel Polyalanine**

*Tetsuo Asakura,* Michi Okonogi, Kumiko Horiguchi, Akihiro Aoki, Hazime Saitô, David P. Knight, and Mike P. Williamson*

Polyalanine (polyA) sequences are the simplest polypeptide sequence to naturally form antiparallel β sheets, and are a key element in the structure of silk fibers.[1–4] Remarkably, there are no structures of antiparallel polyA longer than $Ala_3$ that have been determined at atomic resolution. Therefore, we performed a systematic analysis of antiparallel polyA by using X-ray crystallography and solid-state NMR spectroscopy. PolyA packs into two different arrangements, depending on the length of the sequence. Short sequences ($n = 6$ or less) pack into a rectangular arrangement, as shown by the crystal structure of $Ala_4$ (Figure 1). In contrast, longer sequences pack in a staggered arrangement, for which a structure was derived by using a combination of solid-state NMR spectroscopy and powder diffraction. Polymorphism is emerging as a common feature of amyloid fibers,[5,6] and the polymorphism that is identified here for polyA demonstrates that amyloid is not unique in this respect.

PolyA sequences of different lengths from $Ala_3$ through $Ala_8$ and $Ala_{12}$ were synthesized and crystallized. As has been shown previously, $Ala_3$ can be crystallized in both parallel and antiparallel β-sheet structures.[7] However, the longer peptides form only antiparallel structures, as indicated by FTIR and $^{13}C$ NMR spectra (Figure 2 and Figure S1 in the Supporting Information). Of these peptides, only $Ala_4$ formed single crystals that were large enough for X-ray diffraction experiments. In the structure of $Ala_4$, the molecules are aligned in head-to-tail rows with methyl groups arranged alternately above and below the plane of the sheets (Figure 1). Single water molecules bridge between adjacent N and C termini. The strands are packed into a rectangular lattice and form hydrogen bonds both side-to-side as well as end-to-end. The end-to-end interactions occur through the bridging water molecules. We note that this structure differs from an earlier crystal structure of antiparallel $Ala_3$[3] and from the standard and widely quoted model for polyalanine β sheets,[8] which was derived from fiber diffraction data. In our structure, all of the Ala residues are in equivalent positions, whereas in the other structures, there are two alternative locations.

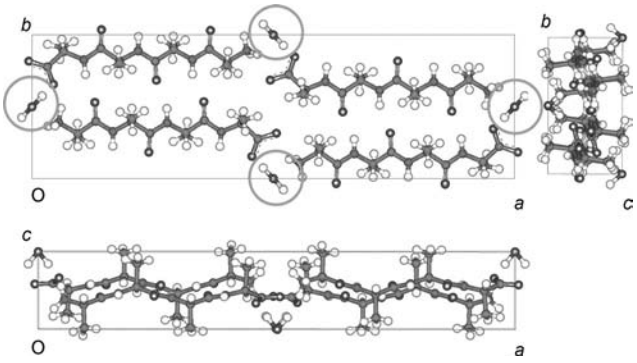


***Figure 1.*** Crystal structure of antiparallel $Ala_4$. Bridging water molecules are indicated by circles. In this structure the mean $\phi$ and $\psi$ dihedral angles are −152° and +152°, respectively. This is a low-energy region of the Ramachandran plot, but slightly above and to the left of the standard antiparallel β-sheet region in the conventional representation. This result is possibly a consequence of the straightness of the chain in this structure, relative to the twisted β sheet that is normally present in proteins.[23]

Evidence from solid-state cross-polarization/magic-angle spinning (CP/MAS) $^{13}C$ NMR spectra and X-ray powder diffraction patterns shows that the short, antiparallel polyA oligomers $Ala_3$, $Ala_5$, and $Ala_6$ have similar crystal structures to that of $Ala_4$ (Figure 2). In the $^{13}C$ NMR spectra of $Ala_3$, $Ala_5$, and $Ala_6$ there is a single central β-carbon signal at $\delta =$ 20.4 ppm that is surrounded by smaller resonances. The smaller signals are generated by the carbon atoms at the two termini. In the spectrum of $Ala_6$ there is also a broad signal at $\delta = 17$ ppm, which is assigned to disordered residues.[9] The X-ray powder diffraction spectra (Figure 2b) have two prominent peaks: one at $2\theta = 17.2°$, which is a result of the 5.16 Å spacing between the Ala planes, and one at $2\theta = 19.2°$, which is a result of the 4.62 Å spacing between adjacent chains, parallel to dimension $b$ of the unit cell. These data demonstrate that the structure of the $Ala_4$ crystal is also maintained in microcrystalline samples. This is an important finding, because it is now clear that amyloid crystals and fibrils can have different morphologies, even for identical sequences, which leads to problems in structural analysis.[10,11]

For $Ala_7$ and higher, both the $^{13}C$ NMR spectra and X-ray diffraction data are markedly different from those for short polyA sequences (Figure 2). The $^{13}C$ NMR spectra have two

[*] Prof. Dr. T. Asakura, M. Okonogi, K. Horiguchi, Dr. A. Aoki, Dr. H. Saitô
Department of Biotechnology
Tokyo University of Agriculture and Technology
2-24-16, Nakacho, Koganei, Tokyo 184-8588 (Japan)
E-mail: asakura@cc.tuat.ac.jp
Dr. D. P. Knight
Oxford Biomaterials Ltd.
Magdalen Centre, Oxford, OX4 4GA (UK)
Prof. Dr. M. P. Williamson
Department of Molecular Biology and Biotechnology
University of Sheffield, Firth Court, Western Bank
Sheffield S10 2TN (UK)

[**] This work was supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science, Culture and Sports of Japan (23245045).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201105356.

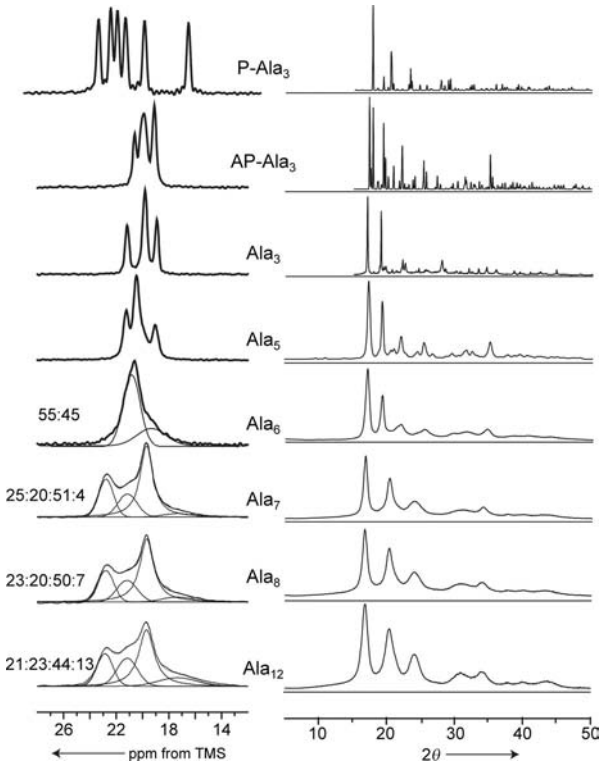


***Figure 2.*** $^{13}$C CP/MAS NMR spectra and X-ray powder diffraction patterns of $Ala_n$ oligomers (parallel $Ala_3$ (P) and antiparallel $Ala_3$ (AP) to $Ala_{12}$). a) $^{13}$C CP/MAS NMR spectra (β-carbon region). The spectra for $Ala_6$ and longer chains are fitted to two signals ($Ala_6$) or four signals (all others). The percentages of the different components are shown. TMS = tetramethylsilane. b) X-ray powder diffraction patterns. Vertical lines indicate the peak that is sensitive to the packing arrangement, as discussed in the text.

β-carbon resonances at $\delta = 22.7$ ppm and 19.6 ppm. Line fitting indicates that there is also a third broad signal at approximately $\delta = 21$ ppm, as well as the broad signal at $\delta =$ 17 ppm, which was detected for $Ala_6$. In the X-ray diffraction spectra, the second peak shifts to $2\theta = 20.4°$, which corresponds to a spacing of 4.35 Å between adjacent chains. Therefore, it is clear that there is also a "long form" packing arrangement, which is different to the "short form" packing that is typified by $Ala_4$.

The structure of the long form was investigated by using a sample of $Ala_7$ in which the β-carbon atom of the central Ala residue ($Ala^4$) was labeled with $^{13}$C. The $^{13}$C NMR spectrum (Figure S2 in the Supporting Information) was identical to that of the nonlabeled sample, except for the absence of the broad signal at $\delta = 17$ ppm, which confirmed that this signal arises from disorder in the terminal residue. Treatment of this sample with trifluoroacetic acid (TFA) resulted in the loss of the central broad resonance at $\delta = 21$ ppm (Figure S2 in the Supporting Information), which suggests that this signal can be assigned to a carbon atom in a separate, TFA-labile region.

Further details about the long-form structure were obtained from a 2D dipolar-assisted rotational resonance (DARR) spectrum[12] of uniformly $^{13}$C-labelled $Ala_7$ (Figure 3 in the Supporting Information). A correlation was detected between the two signals at $\delta = 22.7$ ppm and 19.6 ppm, which implies that these two $^{13}$C nuclei are close together. There was no correlation to the central broad signal at $\delta = 21$ ppm, which indicates that this carbon atom is not close to the other two. Therefore, the two resonances at $\delta = 22.7$ ppm and 19.6 ppm can be assigned to ordered Ala residues within the same crystalline region, rather than a heterogeneous distribution of conformations along a single chain. The long form must, therefore, have Ala residues that are located in two different environments. In contrast, the broad signal at $\delta = 21$ ppm is assigned to a fraction of carbon atoms (ca. 20% by integration) that are in a spatially distinct region. We note that this signal has a very similar chemical shift to the main signal in the short form. This chemical shift is consistent with an assignment to carbon atoms that occupy regions of the short form that are spatially separate from the majority of the sample. Thus, the NMR spectroscopy data suggest that polyA sequences longer than $Ala_6$ consist of a major fraction in which there are two equally populated environments for Ala residues, together with a minor fraction of the short-form structure, in which there is only a single Ala environment.

Further information on the long-form structure came from a rotational echo double resonance (REDOR) spectrum (Figure 3).[13] This analysis showed that the intermolecular

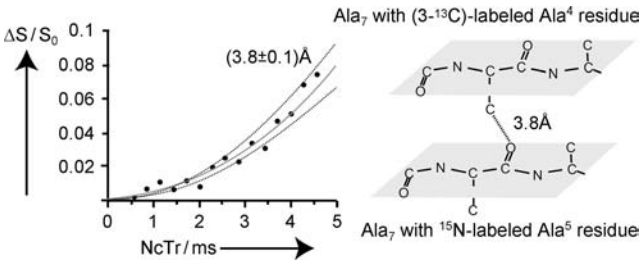


***Figure 3.*** REDOR plot for determination of the intermolecular distance between the $^{13}$C β-carbon label in the $Ala^4$ residue in one $Ala_7$ molecule and the $^{15}$N label in the $Ala^5$ residue in the neighboring $Ala_7$ molecule. $Ala_7$ molecules with a $^{13}$C-labeled $Ala^4$ residue were surrounded by other $Ala_7$ molecules with $^{15}$N labels in the $Ala^5$ residues (molar ratio 1:3). The reduction in the intensity of the β-carbon signal (●) is shown as a function of rotor cycle (NcTr). Dashed lines indicate the error margin.

distance between a $^{13}$C β-labeled $Ala^4$ residue in one $Ala_7$ chain and a $^{15}$N-labeled $Ala^5$ residue in an adjacent $Ala_7$ chain is $(3.8 \pm 0.1)$ Å. These details allow us to propose a model for the structure of the long form (Figure 4b), which is similar to the standard Arnott model.[8] The main difference in the structure of the long form relative to the short form is that the packing of the chains in adjacent planes is staggered rather than rectangular.

The results presented here show that polyA can adopt two different crystal forms, depending on the length of the sequence. The study has three implications. First, the structure of polyA is crucial for understanding the properties of silk. Many silks, such as the very strong dragline spider silk, are thought to derive their strength from the crystalline polyA regions, which are set within a more elastic, glycine-rich matrix.[14–16] Different spiders have different lengths of polyA in their silks, some regions of polyA have six or fewer residues

  

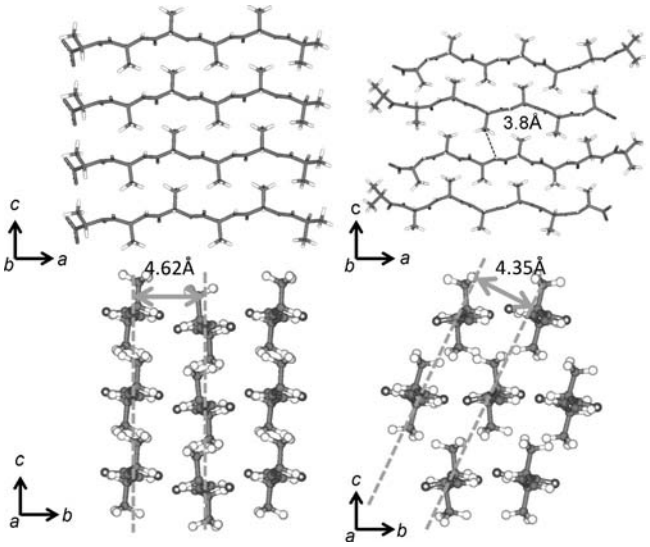


***Figure 4.*** Structures of the antiparallel short form and long form of polyA. a) A model of $Ala_6$, based on the $Ala_4$ crystal structure that was determined in this study. b) Long-form model of $Ala_7$. Crystal directions are as in Figure 1 and the critical packing separation is indicated, together with the REDOR-derived distance for $Ala_7$.

and some have more than six residues.[2] The data presented here imply that shorter polyA stretches may consist mainly of the short-form structure, whereas the structure of polyA stretches that are longer than six residues will be mainly the long form, but will contain roughly 20% of the short-form structure (compare the intensities of the signal at $\delta = 21$ ppm in the $^{13}C$ NMR spectra in Figure 2a, which is assigned to the short-form structure). It is of interest that one of the strongest spider silks known, the dragline silk from the golden orb-web spider *Nephila clavipes*, contains unusually short polyA repeats and has been shown to contain two different polyA crystal forms within the same fiber, which may correspond to the two packing arrangements seen here.[17,18] Our preliminary findings suggest that this may also be true for other dragline silks. Therefore, it is possible that the remarkable strength of these silks is derived from dissipative stick-slip deformation. This deformation is aided by the differing conformations and mechanical properties of the two crystalline forms, which may act synergistically.[15]

The second implication of our results is for modeling of peptide conformations in solution and in aggregated forms. PolyA has been widely used in studies of peptide conformation because of its simplicity and small side chain. For example, it will be of considerable interest whether the different structures that were determined in this study can be reproduced in silico. There has been a series of theoretical and experimental studies on short polyA sequences that suggest that polyA prefers to be in the polyproline II conformation in solution (centered at approximately $\phi = -66°$, $\psi = +137°$), although there is also a significant fraction in the β-sheet region (approximately $\phi = -121°$, $\psi = +128°$). The results of these studies imply that the energy difference between these conformations is small.[19–21] In this study we have shown that crystal packing forces and hydrogen bonding are sufficient to shift the conformation to a third conformation with $\phi = -152°$, $\psi = +152°$, which confirms the suggestion that the β-sheet conformation is readily entropically disfavored.[22] We note that in the polyproline II conformation, the peptide chain is twisted by 120° per residue, whereas in our crystal structure it is completely linear (zero twist), and in β sheets in proteins it is essentially always twisted but to a lesser extent than in the polyproline II conformation.[23] Therefore, there is apparently a correlation between the degree of twist in the strand and the backbone $\phi$ angle, with only a small energy difference across a wide range of angles. It is, therefore, unsurprising to find considerable plasticity in structure in polyA, which includes context-dependent structural propensity.[19] The polymorphism that we have characterized has some similarities to the polymorphisms of amyloid fibrils, as both are different intermolecular packing arrangements of extended β strands. However, there is an important difference in that amyloid fibrils are overwhelmingly composed of parallel strands, whereas the polyA sequences examined in this study are antiparallel.

Third, we note that an increasing number of human diseases have been shown to arise from an expansion in polyA sequences.[24,25] The disease etiology is suggested to arise from the formation of β sheets that aggregate into well-ordered fibrils,[26] although the exact mechanism of cellular toxicity is not clear.[25] It has been suggested that the fibrils consist of antiparallel sheets,[27] in common with the fibrils that are found in polyglutamine expansion diseases.[28,29] Therefore, the structures described in this study may be useful in modeling polyA expansion fibrils.

Received: July 29, 2011
Revised: November 25, 2011
Published online: December 23, 2011

**Keywords:** alanine · peptides · polymorphism · structure elucidation · X-ray diffraction

[1] T. Asakura, D. L. Kaplan in *Encyclopedia of Agricultural Science*, *Vol. 4* (Ed.: C. J. Arutzen), Academic Press, London, **1994**, pp. 1–11.
[2] J. M. Gosline, P. A. Guerette, C. S. Ortlepp, K. N. Savage, *J. Exp. Biol.* **1999**, *202*, 3295–3303.
[3] J. D. van Beek, L. Beaulieu, H. Schäfer, M. Demura, T. Asakura, B. H. Meier, *Nature* **2000**, *405*, 1077–1079.
[4] J. D. van Beek, S. Hess, F. Vollrath, B. H. Meier, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 10266–10271.
[5] J.-P. Colletier, A. Laganowsky, M. Landau, M. Zhao, A. B. Soriaga, L. Goldschmidt, D. Flot, D. Cascio, M. R. Sawaya, D. Eisenberg, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 16938–16943.
[6] R. Nelson, M. R. Sawaya, M. Balbirnie, A. O. Madsden, C. Riekel, R. Grothe, D. Eisenberg, *Nature* **2005**, *435*, 773–778.
[7] J. K. Fawcett, N. Camerman, A. Camerman, *Acta Crystallogr. Sect. B* **1975**, *31*, 658–665.
[8] S. Arnott, S. D. Dover, A. Elliott, *J. Mol. Biol.* **1967**, *30*, 201–208.
[9] T. Asakura, J. M. Yao, *Protein Sci.* **2002**, *11*, 2706–2713.
[10] P. C. A. van der Wel, J. R. Lewandowski, R. G. Griffin, *J. Am. Chem. Soc.* **2007**, *129*, 5117–5130.
[11] P. C. A. van der Wel, J. R. Lewandowski, R. G. Griffin, *Biochemistry* **2010**, *49*, 9457–9469.

  

[12] K. Takegoshi, S. Nakamura, T. Terao, *J. Chem. Phys.* **2003**, *118*, 2325 – 2341.
[13] T. Gullion, J. Schaefer, *J. Magn. Reson.* **1989**, *81*, 196 – 200.
[14] M. Cetinkaya, S. B. Xiao, B. Markert, W. Stacklies, F. Gräter, *Biophys. J.* **2011**, *100*, 1298 – 1305.
[15] S. Keten, Z. P. Xu, B. Ihle, M. J. Buehler, *Nat. Mater.* **2010**, *9*, 359 – 367.
[16] B. L. Thiel, K. B. Guess, C. Viney, *Biopolymers* **1997**, *41*, 703 – 719.
[17] C. Riekel, C. Bränden, C. Craig, C. Ferrero, F. Heidelbach, M. Müller, *Int. J. Biol. Macromol.* **1999**, *24*, 179 – 186.
[18] A. H. Simmons, C. A. Michal, L. W. Jelinski, *Science* **1996**, *271*, 84 – 87.
[19] J. Graf, P. H. Nguyen, G. Stock, H. Schwalbe, *J. Am. Chem. Soc.* **2007**, *129*, 1179 – 1189.
[20] Z. Shi, K. Chen, Z. Liu, N. R. Kallenbach, *Chem. Rev.* **2006**, *106*, 1877 – 1897.
[21] J. Makowska, S. Rodziewicz-Motowidlo, K. Baginska, J. A. Vila, A. Liwo, L. Chmurzynski, H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 1744 – 1749.
[22] M. Mezei, P. J. Fleming, R. Srinivasan, G. D. Rose, *Proteins Struct. Funct. Bioinf.* **2004**, *55*, 502 – 507.
[23] M. P. Williamson, *How Proteins Work*, Garland Science, New York, **2011**, p. 9.
[24] J. Amiel, D. Trochet, M. Clément-Ziza, A. Munnich, S. Lyonnet, *Hum. Mol. Genet.* **2004**, *13*, R235 – R243.
[25] C. Messaed, G. A. Rouleau, *Neurobiol. Dis.* **2009**, *34*, 397 – 405.
[26] M. Sackewitz, H. A. Scheidt, G. Lodderstedt, A. Schierhorn, E. Schwarz, D. Huster, *J. Am. Chem. Soc.* **2008**, *130*, 7172 – 7173.
[27] T. Scheuermann, B. Schulz, A. Blume, E. Wahle, R. Rudolph, E. Schwarz, *Protein Sci.* **2003**, *12*, 2685 – 2692.
[28] R. Schneider, M. C. Schumacher, H. Mueller, D. Nand, V. Klaukien, H. Heise, D. Riedel, G. Wolf, E. Behrmann, S. Raunser, R. Seidel, M. Engelhard, M. Baldus, *J. Mol. Biol.* **2011**, *412*, 121 – 136.
[29] D. Sharma, L. M. Shinchuk, H. Inouye, R. Wetzel, D. A. Kirschner, *Proteins Struct. Funct. Bioinf.* **2005**, *61*, 398 – 411.

  