Cell. Mol. Life Sci. 65 (2008) 2953 – 2956
1420-682X/08/192953-4
DOI 10.1007/s00018-008-8265-1

Cellular and Molecular Life Sciences

Visions & Reflections (Minireview)

Protein assemblies with palindromic structure motifs

N. Pinotsis[a, b] and M. Wilmanns[a, *]

[a] EMBL Hamburg, Notkestrasse 85, D-22603 Hamburg (Germany), Fax: +49-40-89902-149,
e-mail: wilmanns@embl-hamburg.de
[b] Section of Structural Biology, Institute of Cancer Research, Chester Beatty Laboratories, 237 Fulham Road, London SW3 6JB (United Kingdom)

Received 14 May 2008; received after revision 21 June 2008; accepted 14 July 2008
Online First 15 September 2008

Abstract. Symmetric DNA sequence motifs allow the formation of palindromic protein/DNA complexes. Although symmetric protein sequence motifs are less common, recent structural discoveries have unraveled a few protein/protein complexes with palindromic symmetry. Remarkably, symmetric protein/protein complexes can be generated either by adjacent or remote sequence motifs, which may be repeated or inverted. This contribution reflects and comments on recent findings of palindromic protein/protein complexes.

Keywords. Protein structure, palindromic symmetry, SH3 domain, titin, muscle.

Symmetry is one of the basic principles of living systems and is found in situations ranging from the structure of biological macromolecules to the architecture of viruses, organelles, cells and organisms [1 – 3]. In the RNA/DNA world, symmetric sequences have important functional implications in chromosome organization, homologous recombination, self-replication and protein evolution [4 – 8]. When dysfunctional, symmetric sequences may be linked to serious hereditary diseases and cancer [9]. Inverted chromosomal elements were first recognized in the early 1970 s and were subsequently referred to as "palindromic elements" (*palindromos* is Greek for "running back again") [10, 11].

At the level of molecular three-dimensional structures, palindromic DNA recognition elements allow the formation of symmetric protein/DNA complexes, which are induced by the double helical structure of DNA and a restricted set of possible base pairs (AT and CG) that promotes the frequent occurrence of symmetric DNA sequence elements [12] (Fig. 2A). Palindromic response elements favor the association of protein homodimers with important implications for specific regulation of transcription. In contrast, the availability of 20 different amino acids in protein sequences substantially reduces the likelihood of extended protein sequence motifs with palindromic symmetry. Because of the higher variability of amino acid composition and chiral nature of all amino acids (except glycine) proteins, therefore, display a broad range of three-dimensional folds that are inherently asymmetric.

Nevertheless, recent sequence searches have identified a limited number of palindromic sequence elements within available proteomes [13]. To date, however, knowledge of how such elements translate into symmetric three-dimensional structures has remained scarce. This contribution focuses on single chain protein structures with palindromic symmetry. Structural palindromes can either be generated directly by palindromic sequence motifs or by the

* Corresponding author.

symmetric arrangement of remote sequence motifs. However, only if these motifs are in remote positions can they be either repeated or inverted, as illustrated in Figure 1.

One of the longest palindromic protein sequence motifs has been found in prion proteins: VAGAAAAGAV (residues 111–121 in the mouse analogue MoPrP). Several studies demonstrated that this sequence segment is critically involved in conformational changes that may lead to amyloid fibril formation, which has been reported to have a role in prion-transmissible neurodegenerative diseases [14–16]. Biophysical measurements and simulation calculations of an extended peptide (residues 106–126) covering the palindromic sequence motif indicated an octameric complex in which two four-stranded, antiparallel β-sheets are packed against each other by methyl-methyl interactions [17, 18]. Markers that are sensitive to fold transitions within this sequence motif have recently been developed as a preclinical test to diagnose transmissible spongiform encephalopathy [19].

An alternative way to detect sequence-imposed structural symmetry may arise from the analysis of symmetry elements in protein structures. Probably the most classical example is the well established polyproline type II helix (PPII) conformation that has been found in a wide range of proline-rich sequence segments [20]. In contrast to conventional α-helices, the carbonyl groups of PPII sequence motifs are involved in β-sheet-like interactions with surface residues of a protein ligand, while the main-chain amido groups of the proline residues are N-substituted and thus not available for specific hydrogen bonds. Sequences in PPII conformation have almost exactly three residues per turn, thus generating a threefold symmetry along the principle helix axis. In the case of palindromic sequence motifs an additional twofold symmetry is generated vertically to the helix axis, generating 32 point symmetry. PPII or PPII-like conformations are also found in many other sequence motifs that are less rich in or even void of prolines [21]. However, except for collagen, where the role of the characteristic Gly-X-Y sequence pattern for triple helix formation has been well established [22], there are basically no structural data on the implications of those sequence motifs for symmetric protein assembly and, hence, they will not be further discussed in this contribution.

Proline-rich sequence motifs with PPII conformation have an extraordinary ability to form protein complexes with a variety of modular protein domains, such as SH3 domains, WW domains, GYF domains, UEV domains and profilins [23, 24]. Due to the peculiar local symmetry imposed by this conformation, it is not surprising that a number of PPII peptide-mediated symmetric protein complexes have been found. For instance, a dimeric cortactin SH3 domain complex is assembled by the PPII peptide motif (KRPPPPPPG) in its ligand AMAP-1 [25] (Fig. 2B). However, in many other protein/PPII complexes, additional flanking residues, N-terminal or C-terminal to the proline-rich sequence motif, allow the formation of side-chain specific interactions that lead to preferential binding of one among the two possible orientations [23]. In the absence of a more extended structural database of PPII-protein/ligand complexes, beyond the large number of PPII-peptide/peptide complexes already known, the significance of the PPII conformation-imposed ligand self-assembly remains uncertain.

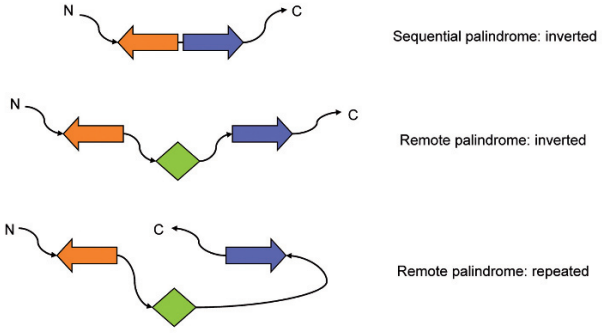

Figure 1. Scheme for structural palindromes in protein assemblies. Remote sequence motifs that allow the formation of structural palindromes can be either inverted or repeated along the sequence. Colors: structural palindrome-generating motifs are in blue and orange; inserted domains are indicated in green.

In contrast to the examples mentioned above, the palindromic structural arrangement found in a recent complex of the N-terminus of the giant muscle protein titin and its assembly ligand telethonin was unexpected [26] (Fig. 2C). Although the interaction between the two components was predicted to be binary [27], the crystal structure unraveled an assembly complex with 2:1 stoichiometry, in which telethonin mediates an antiparallel arrangement of the N-termini of two titin fibers. This symmetry is possible because of two repeated β-hairpin "wing" motifs in telethonin which have significant sequence and structural similarity [26]. The two telethonin wings are separated by an additional core β-sheet domain from the same protein, allowing the formation of the same type of twofold repeated titin-telethonin-titin β-sheet. Mutational analysis of the titin-telethonin complex *in vitro* and *in vivo* showed that the assembly, once formed, is very stable and resistant to the abolishment of specific side-chain interactions. However, recent analysis of patient data revealed that deletion of the entire codon for Glu13 in telethonin leads to rare non-isomorphous

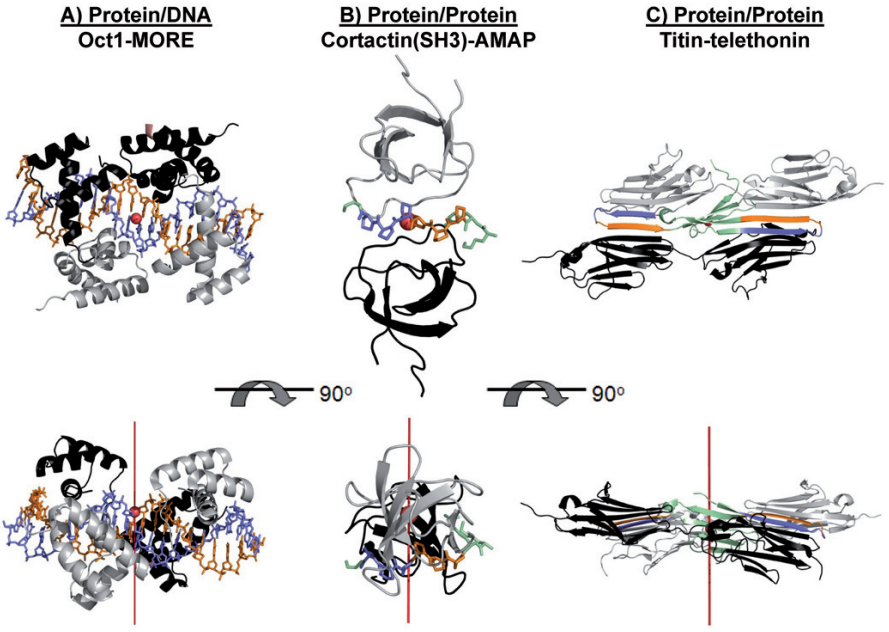

Figure 2. Palindromic assemblies in structural biology. (*A*) The Oct-1 POU domain/MORE complex (1E3O), an example of a palindromic protein/DNA assembly. (*B*) the cortactin (SH3)-AMAP complex (2D1X) and (*C*) the N-terminal titin-telethonin complex (1YA5), as examples of palindromic protein/protein assemblies. The palindromic segments in each complex are shown in orange and blue, with the remaining parts of each palindrome-mediating assembly component shown in green. The protein ligands with palindrome-imposed twofold symmetry are shown in light grey and dark grey. Upper panel, view along the palindrome-mediated twofold axis, which is indicated in each complex by a red sphere; lower panel, each complex is rotated by 90° along a horizontal axis in the paper plane, leading to an orientation of each complex in which the twofold axis is now vertical (red line).

polymorphisms [28, 29]. Since Glu13 is located within the first titin-binding β-hairpin of telethonin, deletion of one residue is expected to lead to a shift of the residue register, most likely interfering with the structure of the same β-hairpin. Interestingly, in the absence of titin, telethonin displays a high potential for aggregation. Both secondary structure prediction and circular dichroism measurements of apo-telethonin indicate a different structure than that observed in the titin-telethonin complex [26, 30],[Pinotsis & Wilmanns, unpublished].

From all the available structural data on palindromic protein/protein complexes, it seems premature to derive general conclusions. It is also unlikely that palindromic sequence/structure arrangements are as widespread in protein/protein interactions as palindromic DNA response motifs are in protein/DNA interactions. Nevertheless, it is remarkable that all known protein/protein assemblies involve hairpin-like structures with a high potential for disease-causing aggregation and fibril formation [31]. The available data indicate that these palindromes could become important molecular parameters in protein assembly and may lead to uncontrolled aggregation in dysfunctional cases.

Acknowledgments. This work has been supported by a FWF/DFG grant (P19060-B12) to M.W.

1 Blundell, T. L. and Srinivasan, N. (1996) Symmetry, stability, and dynamics of multidomain and multicomponent protein systems. Proc. Natl. Acad. Sci. USA 93, 14243–8.
2 Sobell, H. M. (1973) Symmetry in protein-nucleic acid interaction and its genetic implications. Adv. Genet. 17, 411–90.
3 Bamford, D. H., Gilbert, R.J., Grimes, J.M. and Stuart, D.I. (2001) Macromolecular assemblies: greater than their parts. Curr. Opin. Struct. Biol. 11, 107–13.
4 Leach, D. R. (1994) Long DNA palindromes, cruciform structures, genetic instability and secondary structure repair. Bioessays 16, 893–900.
5 Li, T. and Nicolaou, K. C. (1994) Chemical self-replication of palindromic duplex DNA. Nature 369, 218–21.
6 Zhou, Z. H., Akgun, E. and Jasin, M. (2001) Repeat expansion by homologous recombination in the mouse germ line at palindromic sequences. Proc Natl Acad Sci U S A 98, 8326–33.
7 Claverie, J. M. and Ogata, H. (2003) The insertion of palindromic repeats in the evolution of proteins. Trends Biochem. Sci. 28, 75–80.
8 Caburet, S., Conti, C., Schurra, C., Lebofsky, R., Edelstein, S.J. and Bensimon, A. (2005) Human ribosomal RNA gene arrays display a broad range of palindromic structures. Genome Res. 15, 1079–85.
9 Darlow, J. M. and Leach, D. R. (1995) The effects of trinucleotide repeats found in human inherited disorders on palindrome inviability in Escherichia coli suggest hairpin folding preferences in vivo. Genetics 141, 825–32.
10 Hamer, D. H. and Thomas, C. A., Jr. (1974) Palindrome theory. J. Mol. Biol. 84, 139–44.

11 Cavalier-Smith, T. (1974) Palindromic base sequences and replication of eukaryote chromosome ends. Nature 250, 467–70.
12 Higgins, C. F., McLaren, R. S. and Newbury, S. F. (1988) Repetitive extragenic palindromic sequences, mRNA stability and gene expression: evolution by gene conversion? A review. Gene 72, 3–14.
13 Giel-Pietraszuk, M., Hoffmann, M., Dolecka, S., Rychlewski, J. and Barciszewski, J. (2003) Palindromes in proteins. J. Protein Chem. 22, 109–13.
14 Dobson, C. M. (2003) Protein folding and misfolding. Nature 426, 884–90.
15 Nelson, R. and Eisenberg, D. (2006) Structural models of amyloid-like fibrils. Adv Protein Chem 73, 235–82.
16 Sawaya, M. R. Sambashivan S., Nelson R., Ivanova M. I., Sievers S. A., Apostol M. I., Thompson M. J., Balbirnie M., Wiltzius J. J., McFarlane H. T. (2007) Atomic structures of amyloid cross-beta spines reveal varied steric zippers. Nature 447, 453–7.
17 Kuwata, K., Matumoto, T., Cheng, H., Nagayama, K., James, T. L. and Roder, H. (2003) NMR-detected hydrogen exchange and molecular dynamics simulations provide structural insight into fibril formation of prion protein fragment 106–126. Proc. Natl. Acad. Sci. U S A 100, 14790–5.
18 Lee, S. W., Mou, Y., Lin, S. Y., Chou, F. C., Tseng, W. H., Chen, C. H., Lu, C. Y., Yu, S. S. and Chan, J. C. (2008) Steric Zipper of the Amyloid Fibrils Formed by Residues 109–122 of the Syrian Hamster Prion Protein. J. Mol. Biol.
19 Pan, T., Sethi, J., Nelsen, C., Rudolph, A., Cervenakova, L., Brown, P. and Orser, C. S. (2007) Detection of misfolded prion protein in blood with conformationally sensitive peptides. Transfusion 47, 1418–25.
20 Kay, B. K., Williamson, M. P. and Sudol, M. (2000) The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. Faseb. J. 14, 231–41.
21 Stapley, B. J. and Creamer, T. P. (1999) A survey of left-handed polyproline II helices. Protein Sci. 8, 587–95.
22 Brodsky, B. and Persikov, A. V. (2005) Molecular structure of the collagen triple helix. Adv. Protein Chem. 70, 301–39.
23 Zarrinpar, A., Bhattacharyya, R. P. and Lim, W. A. (2003) The structure and function of proline recognition domains. Sci. STKE 2003, RE8.
24 Li, S. S. (2005) Specificity and versatility of SH3 and other proline-recognition domains: structural basis and implications for cellular signal transduction. Biochem. J. 390, 641–53.
25 Hashimoto, S., Hirose M., Hashimoto A., Morishige M., Yamada A., Hosaka H., Akagi K., Ogawa E., Oneyama C., Agatsuma T., Okada M., Kobayashi H., Wada H., Nakano H., Ikegami T., Nakagawa A., Sabe H. (2006) Targeting AMAP1 and cortactin binding bearing an atypical src homology 3/proline interface for prevention of breast cancer invasion and metastasis. Proc. Natl. Acad. Sci. USA 103, 7036–41.
26 Zou, P., Pinotsis, N., Lange, S., Song, Y.H., Popov, A., Mavridis, I., Mayans, O.M., Gautel, M. and Wilmanns, M. (2006) Palindromic assembly of the giant muscle protein titin in the sarcomeric Z-disk. Nature 439, 229–33.
27 Gregorio, C. C., Granzier, H., Sorimachi, H. and Labeit, S. (1999) Muscle assembly: a titanic achievement? Curr. Opin. Cell. Biol. 11, 18–25.
28 Marziliano, N., Pilotto, A., Grasso, M., Pasotti, M. and Arbustini, E. (2006) Deletion of Glu at codon 13 of the TCAP gene encoding the titin-cap-telethonin is a rare polymorphism in a large Italian population. Mol. Genet. Metab. 89, 286–7.
29 Perrot, A., Posch, M. G. and Osterziel, K. J. (2006) Deletion of Glu at codon 13 in the TCAP gene encoding the Z-disc protein titin-cap/telethonin is a rare non-synonymous polymorphism. Mol. Genet. Metab. 88, 199–200.
30 Pinotsis, N., Petoukhov, M., Lange, S., Svergun, D., Zou, P., Gautel, M. and Wilmanns, M. (2006) Evidence for a dimeric assembly of two titin/telethonin complexes induced by the telethonin C-terminus. J. Struct. Biol.
31 Blanch, E. W., Morozova-Roche, L. A., Cochran, D. A., Doig, A. J., Hecht, L. and Barron, L. D. (2000) Is polyproline II helix the killer conformation? A Raman optical activity study of the amyloidogenic prefibrillar intermediate of human lysozyme. J. Mol. Biol. 301, 553–63.