doi:10.1016/j.jmb.2010.09.034 J. Mol. Biol. (2010) 404, 232–246

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

Evolutionary Origin of a Secondary Structure: π-Helices as Cryptic but Widespread Insertional Variations of α-Helices That Enhance Protein Functionality

Richard B. Cooley[1], Daniel J. Arp[2] and P. Andrew Karplus[1]*

[1]*Department of Biochemistry and Biophysics, 2011 Ag and Life Sciences Building, Oregon State University, Corvallis, OR 97331, USA*
[2]*Department of Botany and Plant Pathology, 2082 Cordley Hall, Oregon State University, Corvallis, OR 97331, USA*

Received 24 May 2010;
received in revised form
15 September 2010;
accepted 16 September 2010
Available online
1 October 2010

Edited by M. Sternberg

Keywords:
π-helix;
α-aneurism;
protein evolution;
ferritin-like superfamily;
secondary structure

Formally annotated π-helices are rare in protein structures but have been correlated with functional sites. Here, we analyze protein structures to show that π-helices are the same as structures known as α-bulges, α-aneurisms, π-bulges, and looping outs, and are evolutionarily derived by the insertion of a single residue into an α-helix. This newly discovered evolutionary origin explains both why π-helices are cryptic, being rarely annotated despite occurring in 15% of known proteins, and why they tend to be associated with function. An analysis of π-helices in the diverse ferritin-like superfamily illustrates their tendency to be conserved in protein families and identifies a putative π-helix-containing primordial precursor, a "missing link" intermediary form of the ribonucleotide reductase family, vestigial π-helices, and a novel function for π-helices that we term a "peristaltic-like shift." This new understanding of π-helices paves the way for this generally overlooked motif to become a noteworthy feature that will aid in tracing the evolution of many protein families, guide investigations of protein and π-helix functionality, and contribute additional tools to the protein engineering toolkit.

Introduction

Diversification of protein function is an essential process that enables organisms from all kingdoms of life to thrive in an extraordinary range of environments. The goal of what has been called the functional synthesis of evolution is to gain a detailed understanding of how mutational changes in proteins give rise to novel functionality during this process.[1–3] However, how the first protein folds arose and the concrete mechanisms by which protein secondary structures and protein folds evolve have been difficult to define.

While α-helices (a dominant secondary structure in proteins) are defined by main-chain hydrogen bonds (H-bonds) between residues four positions apart in sequence, π-helices are protein secondary structures with main-chain H-bonds between residues five positions apart in sequence (referred to as π-type H-bonds). They were predicted in the 1950s,[4] and although extended π-helices of regular conformation do not occur in proteins,[5] short π-helical segments do. The empirical definition of a π-helix is the occurrence of at least two sequential π-type H-bonds.[6] The literature for these short motifs is rather muddy, evolving from early work indicating that π-helices do not occur[7] to a report documenting that π-helices as rare but have special functional

*Corresponding author. E-mail address: karplusp@science.oregonstate.edu.

Abbreviations used: H-bond, hydrogen bond; PDB, Protein Data Bank; MMOH, soluble methane monooxygenase hydroxylase; BMM, bacterial multicomponent monooxygenase; ToMO, toluene-4-monooxygenase; RNR, ribonucleotide reductase R2 subunit; PH, phenol hydroxylase.

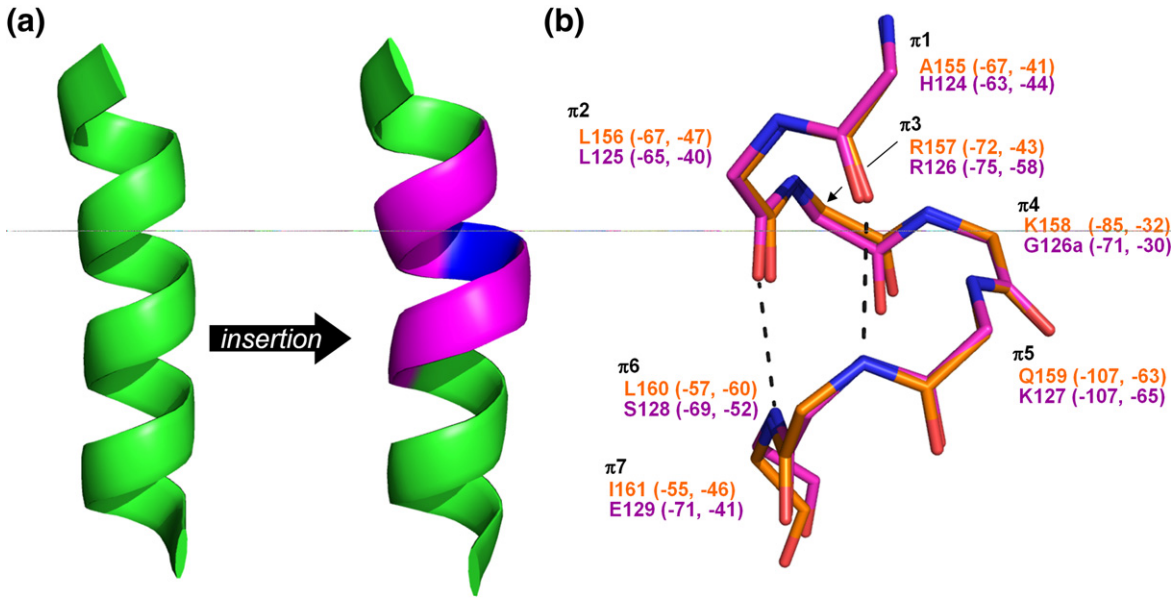

Fig. 1. The common two-H-bond π-helix is the same motif as an engineered α-aneurism. (a) An α-helix in *Staphylococcus aureus* nuclease (left, green cartoon; PDB code 1eyd) develops an α-aneurism (right, purple cartoon; PDB code 1sty) when a single amino acid (highlighted in blue) is inserted into it.[13] (b) An overlay of the π-helix from fumarase C of *Escherichia coli* (residues 155–161, orange carbon atoms; PDB code 1fur), and the same engineered α-aneurism from (a) (purple carbon atoms) demonstrates their equivalence. These α-aneurism-type helical distortions have been independently characterized as looping outs,[14] π-bulges,[15] and α-bulges,[16] but not as π-helices. Black broken lines represent the π-type H-bonds. (ϕ,ψ) torsion angles are written beside each residue. In (b), nitrogen (blue) and oxygen (red) atoms are indicated.

roles,[8] and continuing with the work of Fodje and Al-Karadaghi, who showed that π-helices are more common than suspected, occurring in "one of every 10 proteins."[6] Still, π-helices tend to be overlooked in protein structures. π-Helical conformations have been noted to occur in protein simulations,[9–11] although whether these occurrences give informative insight into the true occurrence of such helices or reflect a shortcoming of the molecular mechanics force fields is controversial.[12]

The large majority of naturally occurring π-helices consist of seven-residue segments with the minimal two π-type H-bonds.[6] What has not been previously noted is that these π-helices have the same conformation and H-bonding pattern seen in protein engineering studies when a single amino acid is inserted into an existing α-helix to create what was called an α-aneurism[13] or looping out[14] (Fig. 1). Also matching this conformation are α-helical distortions seen in some natural proteins, which have been called π-bulges[15] or α-bulges.[16] This striking similarity led us to hypothesize that natural π-helices arise during evolution via single-residue insertions into α-helices. Here, through an analysis of known protein structures, we provide compelling evidence that naturally occurring π-helices are indeed evolutionarily related to α-helices. We further show that this often overlooked secondary structure is a highly informative marker with important evolutionary and functional implications, giving new insights into the origin, evolution, and diversification of protein functionality, even for well-characterized protein families.

Results and Discussion

The evolutionary relationship between α-helices and π-helices

π-Helices are associated with α-helices

To explore our hypothesis that naturally occurring π-helices arose from the insertion of a single amino acid into an α-helix, we first investigated more closely the π-helices identified by Fodje and Al-Karadaghi.[6] Reclassification of their list of π-helices (see Fig. 2 and Materials and Methods) yielded 106 π-helical segments from 79 protein chains, with the longest ones having five consecutive π-type H-bonds (Table 1). Of these 106 π-helices, 102 were present in the midst of an α-helix, consistent with the insertion hypothesis. Interestingly, this observation completely explains why π-helices continue to be overlooked. When a π-helix is sandwiched in an α-helix, the first few residues and the last few residues of the π-helix are also part of the bordering α-helices, and because automated secondary structure assignment algorithms such as DSSP (*d*efinition of *s*econdary *s*tructure in *p*roteins)[17] and STRIDE[18] give precedence to the α-helical assignment, these residues are designated as α-