Journal of Structural Biology xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Variability of the core geometry in parallel coiled-coil bundles

Krzysztof Szczepaniak[a], Jan Ludwiczak[a,b], Aleksander Winski[a], Stanislaw Dunin-Horkawicz[a,*]

[a] Laboratory of Structural Bioinformatics, Centre of New Technologies, University of Warsaw, 02-097 Warsaw, Poland
[b] Laboratory of Bioinformatics, Nencki Institute of Experimental Biology, Pasteura 3, 02-093 Warsaw, Poland

ARTICLE INFO

Keywords:
Coiled coils
Structure prediction
Machine learning
Hydrophobic core
Sequence design

ABSTRACT

In protein modelling and design, an understanding of the relationship between sequence and structure is essential. Using parallel, homotetrameric coiled-coil structures as a model system, we demonstrated that machine learning techniques can be used to predict structural parameters directly from the sequence. Coiled coils are regular protein structures, which are of great interest as building blocks for assembling larger nanostructures. They are composed of two or more alpha-helices wrapped around each other to form a supercoiled bundle. The coiled-coil bundles are defined by four basic structural parameters: topology (parallel or antiparallel), radius, degree of supercoiling, and the rotation of helices around their axes. In parallel coiled coils the latter parameter, describing the hydrophobic core packing geometry, was assumed to show little variation. However, we found that subtle differences between structures of this type were not artifacts of structure determination and could be predicted directly from the sequence. Using this information in modelling narrows the structural parameter space that must be searched and thus significantly reduces the required computational time. Moreover, the sequence-structure rules can be used to explain the effects of point mutations and to shed light on the relationship between hydrophobic core architecture and coiled-coil topology.

1. Introduction

Understanding the relationship between sequence and structure is crucial for protein modeling and design. Coiled coils represent a perfect model for such studies due to their regularity, allowing for accurate numerical description of their structures. Coiled-coil protein domains consist of two, three or more helices in parallel or antiparallel orientation that are wrapped around each other into superhelical bundles. The constituent helices of coiled coils interact via a geometry termed knobs-into-holes (Crick, 1953a, 1952), in which a residue from one helix (knob) packs into a cavity formed by side-chains of the facing helix (hole) (Lupas et al., 2017). Knobs-into-holes packing requires that the side-chains occupy periodically equivalent positions along the helical interfaces. Such an arrangement cannot be achieved with undistorted helices, which have periodicity around 3.6 residues per turn and thus the position of side-chains on their surface drifts continuously. Consequently, helices in coiled coils are bent and wound around each other to effectively (with respect to the bundle axis) reduce the number of residues per turn to 3.5. This allows the position of the side chains to repeat after two helical turns and gives rise to a seven-residue sequence repeat (heptad repeat) pattern. Positions in heptads are labeled *a*–*g*: the core-forming positions (*a* and *d*) are usually occupied by hydrophobic residues, whereas the remaining solvent-exposed positions (*b*, *c*, *e*, *f*, and *g*) are dominated by hydrophilic residues (middle panels in Fig. 1).

The regular nature of coiled coils means that their structures can be represented by parametric equations such as those originally developed by Crick (Crick, 1953b) that capture features of the bundle (degree of supercoiling, radius) and of the individual helices (rotation around the axis) (Lupas and Gruber, 2005). These parametrized models permits the calculation of coiled-coil backbone structures with predefined parameters, which subsequently can be used as templates for designs (Boyken et al., 2016). Likewise, experimentally determined structures can be described using Crick's parameters. This is especially useful in interpreting the effects of mutations, comparing snapshots of dynamic structures, and understanding the relationship between sequence and structure in coiled-coil proteins (Ferris et al., 2014; Grigoryan and Degrado, 2011; Szczepaniak et al., 2014). However, measuring the parameters of a coiled-coil structure involves a certain degree of averaging with respect to the local variability of the structure. For example, in the CCCP tool (Grigoryan and Degrado, 2011) the input structure is globally fitted to an idealized model generated using parametric equations. In contrast to CCCP, other programs such as TWISTER (Strelkov and Burkhard, 2003) or SamCC (Dunin-Horkawicz and Lupas, 2010) calculate the parameters for the individual positions in a structure, and thus permitting the detection of local changes. However, these programs also use averaging to determine bundle and helical axes. This

* Corresponding author.
E-mail address: s.dunin-horkawicz@cent.uw.edu.pl (S. Dunin-Horkawicz).

https://doi.org/10.1016/j.jsb.2018.07.002
Received 10 January 2018; Received in revised form 15 June 2018; Accepted 1 July 2018

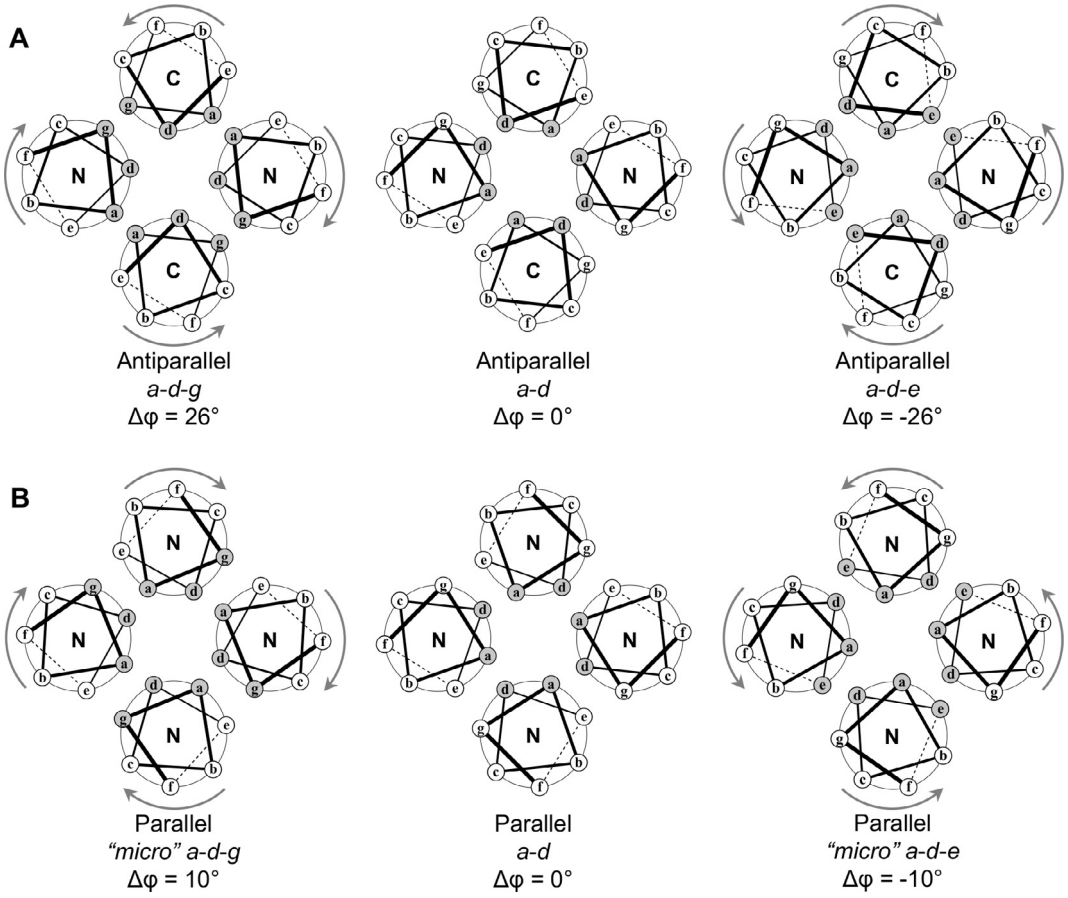

Fig. 1. Helical wheel diagrams of antiparallel (A) and parallel (B) coiled coils with the heptad positions labeled *a–g*. Heptad positions involved in the hydrophobic core formation are indicated by a gray background. Diagrams in the center correspond to the canonical *a–d* core, whereas those on the sides correspond to an extended *a–d–g* core (left) brought about by clockwise axial rotation of the helices, as viewed from the N-terminus, and the *a–d–e* core (right) by counter-clockwise rotation. The direction of the rotation relative to the canonical *a–d* core is indicated with gray arrows.

raises the question whether small fluctuations in measured parameters are noise resulting from the imperfection of the measurement methods or represent real features of the structure.

The helix axial rotation (Crick angle deviation, Δφ) is a parameter that describes the core packing geometry. In experimental coiled-coil structures the axial rotation ranges between −26° and 26° and values close to 0° are considered to correspond to a canonical knobs-into-holes packing in which two heptad positions are oriented towards the core of a bundle (*a–d* core; Fig. 1, middle panels). Some of the antiparallel structures employ an alternative packing mode, so-called complementary *x–da* packing, which is brought about by the global axial rotation of all helices by more than ± 10° (Fig. 1A; (Lupas et al., 2017)). When viewed form the N-terminus of the helix, clockwise rotation (Δφ > 10°) introduces an additional heptad position *g* to the hydrophobic core (*a–d–g* core), while counter-clockwise rotation (Δφ < −10°) introduces position *e* (*a–d–e* core). This *a–d–e* packing was identified in many homotetrameric antiparallel coiled coils, whereas *a–d–g* packing is rarely observed in this type of structures and is presumably energetically unfavorable (Szczepaniak et al., 2014).

In this study, we show that parallel homotetrameric coiled coils utilizing canonical *a–d* core display a continuum of helix axial rotation states (Fig. 2) and that this variability is not an artifact of structure determination or measurement procedures. Importantly, the individual variants of *a–d* packing can be associated with well-defined sequence features and predicted directly from the sequence with ~2° accuracy. Considering the above points, we propose that the *a–d* core can be further subdivided into the canonical *a–d* core (Δφ ~ 0°), the "micro" *a–d–g* core (0° < Δφ < 10°), and the "micro" *a–d–e* core (0° > Δφ > −10°; Fig. 1B). The amino acid composition of the hydrophobic core defines not only the geometry of packing interactions, but also the oligomerization state and orientations of the helices (Lupas et al., 2017; Szczepaniak et al., 2014). Therefore, understanding the rules that link sequence with structure are essential for successful designing and modelling coiled coils. To exemplify the applicability of the proposed approach (available at http://lbs.cent.uw.edu.pl/ccpred) we demonstrated that the predicted packing mode can be used to aid modelling of coiled-coil structures.

2. Material and methods

2.1. Measurement of global helix axial rotation in structures

The Crick angle (φ) quantifies the position of a given residue relative to the bundle axis (Lupas and Gruber, 2005) – φ = 0° indicates that residue points directly towards the center of a bundle, whereas φ = ± 180° indicates that it points directly outwards the center of a bundle. The helix axial rotation (Δφ) for a given residue is defined as the difference between measured Crick angle value (φ) and Crick angle value expected at a given heptad position in a canonical *a–d* core (Dunin-Horkawicz and Lupas, 2010). The following reference values were used: 19.5°, 122.35°, −134.78°, −31.92°, 70.92°, 173.78°, and −83.35° for heptad positions *a*, *b*, *c*, *d*, *e*, *f*, and *g*, respectively. The reference values were selected based on the energy landscape (right panel in Fig. 2) such that the ideal *a–d* core (Δφ = 0) corresponds to the center of the Rosetta energy peak. Δφ = 0 denotes that measured and expected Crick angle for a given position are equal, and thus the residue