Journal of Structural Biology 170 (2010) 192–201

Contents lists available at ScienceDirect

Journal of Structural Biology

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

The many types of interhelical ionic interactions in coiled coils – An overview

Markus Meier [a,b], Jörg Stetefeld [b], Peter Burkhard [a,*]

[a] The Institute of Materials Science and Department of Molecular and Cell Biology, University of Connecticut, 97 North Eagleville Road, Storrs, CT 06269-3136, USA
[b] Department of Chemistry, University of Manitoba, 144 Dysart Road, Winnipeg, Canada R3T 2N2

ARTICLE INFO

Article history:
Received 15 December 2009
Received in revised form 26 February 2010
Accepted 2 March 2010
Available online 6 March 2010

Keywords:
Coiled coil
α-helix
Interhelical ionic interaction
Ion pair
Salt bridge

ABSTRACT

Coiled coils represent the most frequent protein oligomerization motif in nature and are involved in many important biological processes. The prototype interhelical ionic interaction for coiled coils described in literature is an *i* to *i* + 5 ionic interaction from heptad position g to e′, but other possible ionic interactions have also been described. Here we use a statistical approach to systematically analyze all high-quality coiled-coil structures in the RCSB protein database for their interhelical ionic interactions. We provide a complete listing of all possible arrangements and analyze the frequency of their occurrence in the primary sequence together with their probability of formation in the quaternary structure of the coiled coils. We show that the classical *i* to *i* + 5 ionic interaction is indeed characteristic for parallel dimeric and trimeric coiled coils. But we also show that there are many more *i* to *i* + 2 ionic interactions in parallel tetrameric and pentameric coiled coils, and in antiparallel coiled coils the classical *i* to *i* + 5 ionic interaction is in none of the oligomerizations states the most frequently observed ionic interaction. We also demonstrate that many ionic interactions involve residues at the core positions that are usually occupied by hydrophobic residues and that such interhelical ionic interactions are a hallmark feature of dimeric coiled coils.

1. Introduction

Coiled coils represent an important and ubiquitous protein folding and oligomerization motif. This is highlighted by the fact that the first recorded X-ray fiber diffraction pattern of a protein (McArthur, 1943) was that of a coiled coil. Its fold, a.k.a. "knobs into holes packing", was predicted by Crick in 1953 (Crick, 1953) long before the first high-resolution structure of a dimeric coiled coil was published in 1991 by O'Shea (O'Shea et al., 1991).

Coiled coils are oligomers of two up to seven strands (O'Shea et al., 1991; Liu et al., 2006; Malashkevich et al., 1996; St Maurice et al., 2007; Stetefeld et al., 2000; Tao et al., 1997) of α-helices wound around each other to form a superhelical twist that is usually left-handed. Such left-handed coiled coils are characterized by a 7 amino acid repeat $(abcdefg)_n$ in their primary sequence, known as heptad repeat. However, also other repeat sequences such as hendecad (11) repeats (Stetefeld et al., 2000), pentadecad (15) repeats (Kuhnel et al., 2004), or even heptadecad (17) repeats (Bossus et al., 1997) exist, leading to straight or even right-handed superhelical twists. The helices may be aligned in either parallel or antiparallel fashion (Deng et al., 2006). Coiled coils are abundant in nature (for a review see Burkhard et al. (2001)), but there are also many *de novo* designed coiled-coil structures. Coiled coils can be very long proteins like tropomyosin, or they may be as short as two heptad repeats long to fold into a stable, three dimensional protein structure (Burkhard et al., 2002, 2000).

Due to its simplicity and regularity, the coiled coil protein folding motif has been one of the preferred model systems to study protein–protein interactions, protein folding and protein stability. Despite of its relative simplicity, it is far from being completely understood, and the ability to predict the correct coiled- coil protein fold *ab initio* has had only limited success. Replacement of only few amino acids can lead to different oligomerization states or unpredicted strand alignments (Deng et al., 2006), and sometimes only changing the refolding conditions may lead to a change of the oligomeric state (Burkhard et al., 2000).

This is because – despite of its simplicity – there are still many factors contributing to the stability and specificity of the coiled-coil structure. Undoubtedly, the main key driving force for protein folding, and hence also coiled coil formation, is the hydrophobic interaction. The hallmark in coiled coils is the hydrophobic interaction between residues at the so-called core positions (a) and (d) for the heptad repeat of coiled coils (Tripet et al., 2000; Wagschal et al., 1999). These positions are frequently occupied by leucine residues, hence the term leucine zipper. However, other interatomic interactions such as Van der Waals interactions (Zhu et al., 1993), hydrogen bonds, and ionic interactions (Burkhard et al., 2002; Matousek et al., 2007; Meier and Burkhard, 2006; Spek et al., 1998) also contribute to the thermal stability of coiled coils. Other factors that

* Corresponding author. Fax: +1 860 486 4745.
E-mail address: Peter.Burkhard@uconn.edu (P. Burkhard).

1047-8477/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jsb.2010.03.003

influence coiled-coil stability include the length of the particular coiled coil (Burkhard et al., 2000; Kammerer et al., 1998; Litowski and Hodges, 2001), which contributes to the stability in a kind of cooperative manner, the helical propensities of the residues (Chakrabartty et al., 1994; O'Neil and DeGrado, 1990), the helix dipole (Kohn et al., 1997), helix capping effects (Aurora and Rose, 1998; Lu et al., 1999; Tripet and Hodges, 2002), and translational and rotational entropy contributions (Yu et al., 1999).

One particular topic that has been extensively and controversially discussed, is the role of ionic interaction for folding and stability in coiled coils (Lavigne et al., 1996; Lumb and Kim, 1996). Ionic interactions play a critical role in many biological processes. They are important for the specificity of molecular recognitions, because the hydrophobic interaction, which is the main driving force for strong protein–protein association, is essentially unspecific. Thus, ionic interactions may determine the folding pathway and also the oligomerization state of proteins.

A better understanding of the contribution of ionic interactions, also called salt bridges, to protein folding and stability may therefore benefit the research fields of protein structure prediction and protein *de novo* design, since coiled coils are used as a scaffold for the design of many novel protein-based devices such as artificial fibers (Papapostolou et al., 2007), hydrogels (Wheeldon et al., 2008), protein purification tags (Litowski and Hodges, 2002), vaccines (Kaba et al., 2009; Pimentel et al., 2009), and many more.

There are several ways to elucidate the contribution of the ionic interactions to coiled-coil stability. Mutational analysis, for example, has been extensively applied by the lab of R. Hodges to investigate the contribution of electrostatic interaction to coiled-coil stability and formation (Lavigne et al., 1996; Kohn et al., 1995). Structural analysis by X-ray crystallography (Burkhard et al., 2000, 2002) or NMR (Matousek et al., 2007) has shed light into many structural aspects of these interactions. NMR is particularly suited to investigate the contribution of ionic interactions to coiled-coil stability due to its ability to monitor side chain titration curves in correlation to the structural stability of the protein (Matousek et al., 2007), which in turn allows to perform pH dependent stability and folding experiments.

Recently, we have used a statistical approach to derive information on intrahelical electrostatic interactions about the stability of coiled coils (Meier and Burkhard, 2006). Here we present a similar statistical approach for the interhelical ionic interactions in coiled coils. As a consequence, much of the language and methodology will be very similar to this previous publication. Especially one interhelical ionic interaction has been discussed in literature, which is the eg′ i to $i + 5$ interhelical interaction. Other ionic interactions have also been described (Burkhard et al., 2000), but a complete listing of the possible interactions is so far missing. Here we give a detailed analysis of all existing ionic interactions in coiled coils in the RCSB protein structure database (Berman et al., 2000) and describe their statistical occurrence.

2. Materials and methods

We have used our previously developed program *SBSCC* to produce a statistics of the electrostatic interactions in all coiled-coil structures that were published in the FTP release of the Worldwide Protein Data Bank (wwPDB, Berman et al., 2003) until January 12, 2010. We further merged in all the coiled-coil structures listed in the CC+ database (Testa et al., 2009) since this database contained some structures that had not yet migrated from the main PDB repository at RCSB (Berman et al., 2000) to the wwPDB FTP archive. We found, however, that the protein data bank contains substantially more coiled-coil structures than those listed in the CC+ database. *SBSCC* identifies coiled-coil motifs in the protein database using Socket (Walshaw and Woolfson, 2001) with a default packing distance cut-off of 7.0 Å, assigns the heptad repeat using an algorithm derived from TWISTER (Strelkov and Burkhard, 2002), searches for ionic interactions within those motifs, and classifies the interactions according to their type and counts their occurrence. For details and the SBSCC algorithm see reference (Meier and Burkhard, 2006).

For our statistics, we used all high-quality coiled coils with a minimal length of 13 residues from X-ray crystallographic structures with a resolution better than 2.5 Å and a free *R*-factor lower than 30%. From NMR structure ensembles we used the most representative conformer indicated by the author. If that information was not included, we used the first model of the ensemble.

2.1. Helix length compensation

The frequency of each type of salt bridge configuration in the dataset is not only dependent on the physics of the coiled-coil motif but also on the frequencies of the amino acid types it consists of and the length of each coiled coil. The latter is simply due to the fact that, within a coiled coil of finite length, there are more pairs of amino acids with a shorter spacing between them than pairs with a larger spacing.

The data presented here contain the original numbers reported by *SBSCC*. These data are biased towards the amino acid composition of the structures and the distribution of coiled-coil lengths.

We did the following adjustment for the coiled-coil length effect on the frequency of the salt bridges for the data presented in panels C and D of Fig. 2: Each dimeric coiled coil of length L can accommodate $2(L - i)$ amino acid pairings with spacing i. The total number $n_{cfg,i}$ of amino acid pairings with spacing i in each dataset was calculated from the coiled-coil length distribution $\boldsymbol{f}(L_{cc})$. The correction factor c_i for amino acid pairings of spacing i corresponds to the number of coiled-coiled residues in the dataset divided by $n_{cfg,i}$. The adjusted number of amino acid pairings $n_{c,i}$ with spacing i is therefore $n_{c,i} = c_i * n_i$, where n_i is the observed number of amino acid pairings with spacing i.

2.2. Curating the datasets

We removed all five right-handed coiled coils (PDB entries 2O6N, 1TGG, 1FE6, 1RH4, 1YBK) because they contain hendecad repeats instead of heptad repeats. We further excluded the single seven-helix coiled coil (PDB entry 2HY6), because this structure does not display regular seven fold symmetry.

2.3. Non-redundant datasets

The non-redundant datasets of parallel and antiparallel two-stranded coiled coils were compiled from coiled coils with less than 50% sequence identity. The corresponding lists of PDB IDs were obtained from the CC+ database (Testa et al., 2009). We checked the relatedness of the sequences by aligning and sorting them according to their similarity using the program STRAP (Gille and Frömmel, 2001). We removed remaining duplicates from the PDB files using pdbcur (CCP4, 1994; Krissinel et al., 2004). We also excluded *de novo* designed sequences. To confirm that the automatically assigned heptad repeats were correct, we inspected all coiled coils visually using Pymol (DeLano, 2002).

2.4. Redundant datasets

The redundant datasets were derived from the FTP archive from wwPDB and the CC+ database (Testa et al., 2009) as described above. In these datasets we inspected all trimeric, tetrameric and pentameric coiled coils with Pymol (DeLano, 2002). The datasets of parallel and antiparallel two-stranded coiled coils were too large for visual