Bioinformatics, 31(5), 2015, 767–769
doi: 10.1093/bioinformatics/btu700
Advance Access Publication Date: 22 October 2014
Applications Note

OXFORD

Sequence analysis

Waggawagga: comparative visualization of coiled-coil predictions and detection of stable single α-helices (SAH domains)

Dominic Simm, Klas Hatje and Martin Kollmar*

Department of NMR-based Structural Biology, Max-Planck-Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

*To whom correspondence should be addressed.

Associate Editor: John Hancock

Received on July 14, 2014; revised on October 17, 2014; accepted on October 20, 2014

Abstract

Summary: Waggawagga is a web-based tool for the comparative visualization of coiled-coil predictions and the detection of stable single α-helices (SAH domains). Overview schemes show the predicted coiled-coil regions found in the query sequence and provide sliders, which can be used to select segments for detailed helical wheel and helical net views. A window-based score has been developed to predict SAH domains. Export to several bitmap and vector graphics formats is supported.

Availability and implementation: http://waggawagga.motorprotein.de

Contact: mako@nmr.mpibpc.mpg.de

1 Introduction

Coiled coils are α-helical structural domains common to all domains of life and present in 2–12% of the proteins of a proteome (Liu and Rost, 2001). The classical coiled coils comprise dimerising long α-helices that wind around each other forming superhelices (Crick, 1952) as found in structural proteins such as α-keratins, muscle myosins and tropomyosin (KMTs). In the last years, this view has shifted to defining coiled-coil segments based on the presence of knobs-into-holes packing of side chains between α-helices resulting in many different architectures and topologies (Moutevelis and Woolfson, 2009). The basis for a coiled coil is an amino acid heptad ('abcdefg'), in which the 'a' and 'd' positions are occupied by hydrophobic residues.

Coiled coils were among the first structural domains to be predicted by algorithms (Lupas *et al.*, 1991). In this method a query sequence is compared with a database of known coiled-coil sequences, which were the KMTs available at that time. Subsequently, a similarity score is computed and a probability to form a coiled-coil calculated. In principle, this is an implementation of the basic ideas already presented nine years before, that residues show an asymmetric distribution within the heptad repeats and that this statistical data can be used to predict coiled coils in other proteins (Parry, 1982). This position-specific scoring matrix (PSSM) approach has been improved both on the database site and on the feature site trying to disfavor the assignment of high coiled-coil probabilities to hydrophilic sequences.

Another approach to predict coiled coils is based on the pairwise residue probabilities as implemented in Paircoil (Berger *et al.*, 1995) and Paircoil2 (McDonnell *et al.*, 2006). Here, pairwise frequencies of heptad residues are calculated from known coiled-coil sequences and the probability of a pair of amino acids in a given sequence for a certain combination in the heptad is scored. With this approach better predictions for long coiled-coil regions could be obtained compared to PSSM predictions. Improvement of the prediction of short coiled coils has been reached by using a hidden Markov model as in Marcoil (Delorenzi and Speed, 2002), and, recently, Markov Random Fields as used in Multicoil2 (Trigg *et al.*, 2011). These general coiled-coil prediction approaches have been extended by software to predict different oligomerization states as Scorer (Armstrong *et al.*, 2011), PrOCoil (Mahrenholz *et al.*, 2011), and LOGICOIL (Vincent *et al.*, 2013). However, all these approaches provide different results and fail to distinguish between coiled coils and stable single α-helices (SAH domains). A special case of SAH

domains consisting of alternating repeats of four glutamic acid (E) residues and four positively charged residues of either lysine (K) or arginine (R) has been termed ER/K motif (Sivaramakrishnan *et al.*, 2008). SAH domains are in most cases mispredicted as coiled coils but have been shown to form stable monomeric structures in aqueous solution (Knight *et al.*, 2005; Süveges *et al.*, 2009). SAHs are highly enriched in E, Q, K and R residues, which stabilize the α-helices through intrahelical salt bridges (Peckham and Knight, 2009).

Waggawagga is designed to provide a direct schematic comparison of many coiled-coil prediction tools. Users can inspect the predictions in classical helical wheel (Schiffer and Edmundson, 1967) and helical net (Dunnill, 1968) representations. Visualization of the coiled-coil predictions in helical wheel schemes provides the possibility to fast and easily identify potential hydrophobic seams in 'a' and 'd' and oppositely charged residues in 'e' and 'g' positions, respectively. In contrast, SAH domains have patterns of highly charged residues only occasionally interrupted by hydrophobic or hydrophilic amino acids. Therefore, charged residues are highly enriched in the 'a' and 'd' positions. The hydrogen-bonded and charged interaction network in SAH domains is best seen in helical net representations showing the potential interactions along the helix. Waggawagga provides layouts that can easily be used in presentations and manuscripts.

2 Features

Waggawagga allows the comparative analysis of six coiled-coil prediction (Marcoil, Multicoil, Multicoil2, Ncoils, Paircoil, Paircoil2) and three oligomerization state prediction programs (Scorer, PrOCoil and LOGICOIL). In addition, Multicoil2 distinguishes dimers, trimers and non-coiled-coil oligomerization states. These tools can be run in any combination against single or multiple query sequences.

2.1 Domain view

The interactive domain view (Fig. 1A) is the main control element for setting the two analysis views, the helical wheel (Fig. 1B) and the helical net views (Fig. 1C). Separate schemes are generated for each of the results of the selected coiled-coil prediction programs visualizing the coiled-coil regions on top of the query sequence. For each region, the predicted oligomerization state is given depending on the selected tools. In the selected domain scheme, any specific region can be chosen by mouse clicks or by moving an interactive slider. The respective region is shown in detail in the helical wheel and the helical net views.

2.2 Helical wheel view

The exact sequence borders of the coiled-coil region and several different types of helix arrangements (parallel dimer, anti-parallel dimer, trimer) can be set in the configuration panel. In the helical wheel view, every α-helix is shown along the helix axis and the helices are arranged that the hydrophobic core is in the interface of the two or three helices. Each helical wheel represents 10 helix turns.

2.3 Helical net view

Intra-helical interactions are displayed in the helical net view, which is the representation of a helix split open along a line parallel to its axis and laid flat. Solid and dashed lines mark strong and week interactions, respectively. Strong and weak interactions are mainly formed by oppositely charged residues from subsequent turns resulting in hydrogen-bonded networks in addition to charge interactions, with the strength of the interaction given by the distance between the residues and their relative orientation. Interaction networks lead to extra stabilization in addition to that of the component pairs, and hydrophobic seams favor helix association in contrast to single helices. The so-called SAH-score is calculated as the sum of the

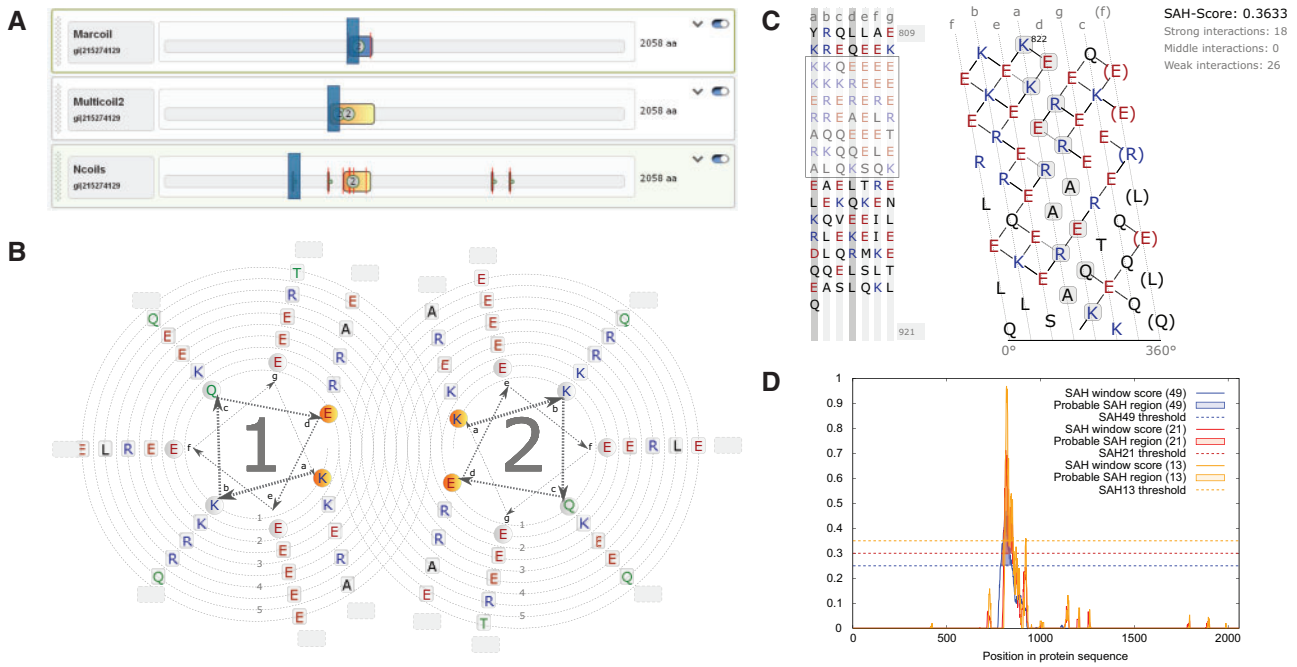

Fig. 1. (**A**) Interactive domain view of the coiled-coil prediction of human myosin-10, with the Marcoil prediction activated. (**B**) Helical wheel view. (**C**) Helical net view. 'a' to 'f' refer to the heptad positions. The 'f' residues are repeated on both sides of the net to better visualize intrahelical interactions with the residues in 'b' and 'c' positions. Solid and dotted lines between residues denote strong and weak intrahelical interactions, respectively, as have been introduced by others (Peckham and Knight, 2009). (**D**) SAH line plots for different window sizes

interactions divided by the window-size resulting in values from 0 to 1. SAHs typically have values greater than 0.25.

2.4 Line plots and tables

Here, the prediction scores are presented as line plots and tables (Fig. 1D). These are particularly useful to evaluate the relevance of SAH regions.

2.5 Performance and limitations

The SAH prediction scheme and score have been developed and tested against previously reported regions from human proteins (Peckham and Knight, 2009), SwissProt proteins (Süveges *et al.*, 2009), and thousands of cytoskeletal and motor proteins from all across the eukaryotes as available from CyMoBase (http://www.cymobase.org). There are other proteins that are mispredicted as coiled coils although forming monomeric α-helical structures, such as stathmin, which, however, are not enriched in E, Q, K, and R residues and do not form stable structures in solution (Honnappa *et al.*, 2006). These types of proteins do not contain SAH domains according to their definition (Peckham and Knight, 2009; Süveges *et al.*, 2009), and are consequently not detected and characterized as monomeric α-helical proteins in the current implementation of Waggawagga.

3 Implementation

The web application framework is Ruby on Rails. In order to present the user with a feature rich interface the site makes extensive use of Ajax (Asynchronous JavaScript and XML) using jQuery (http://jquery.com) and FancyBox (http://fancybox.net). Interactive schemes are drawn as SVG and graphs are generated with the graphical toolkit GnuPlot (http://www.gnuplot.info). Figure export is done through an intermediary SVG file, which is converted into various output formats using the Inkscape graphics package (http://inkscape.org). User-uploaded data is stored temporary on the server and deleted when leaving the application.

Acknowledgement

We would like to thank Christian Griesinger for his continuous generous support.

Conflict of Interest: none declared.

References

Armstrong,C.T. *et al.* (2011) SCORER 2.0: an algorithm for distinguishing parallel dimeric and trimeric coiled-coil sequences. *Bioinformatics*, **27**, 1908–1914.

Berger,B. *et al.* (1995) Predicting coiled coils by use of pairwise residue correlations. *Proc. Natl Acad. Sci. USA*, **92**, 8259–8263.

Crick,F.H.C. (1952) Is alpha-keratin a coiled coil? *Nature*, **170**, 882–883.

Delorenzi,M. and Speed,T. (2002) An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics*, **18**, 617–625.

Dunnill,P. (1968) The use of helical net-diagrams to represent protein structures. *Biophys. J.*, **8**, 865–875.

Honnappa,S. *et al.* (2006) Control of intrinsically disordered stathmin by multisite phosphorylation. *J. Biol. Chem.*, **281**, 16078–16083.

Knight,P.J. *et al.* (2005) The predicted coiled-coil domain of myosin 10 forms a novel elongated domain that lengthens the head. *J. Biol. Chem.*, **280**, 34702–34708.

Liu,J. and Rost,B. (2001) Comparing function and structure between entire proteomes. *Protein Sci.*, **10**, 1970–1979.

Lupas,A. *et al.* (1991) Predicting coiled coils from protein sequences. *Science*, **252**, 1162–1164.

Mahrenholz,C.C. *et al.* (2011) Complex networks govern coiled-coil oligomerization – predicting and profiling by means of a machine learning approach. *Mol. Cell. Proteomics*, **10**, M110.004994.

McDonnell,A.V. *et al.* (2006) Paircoil2: improved prediction of coiled coils from sequence. *Bioinformatics*, **22**, 356–358.

Moutevelis,E. and Woolfson,D.N. (2009) A periodic table of coiled-coil protein structures. *J. Mol. Biol.*, **385**, 726–732.

Parry,D.A. (1982) Coiled-coils in alpha-helix-containing proteins: analysis of the residue types within the heptad repeat and the use of these data in the prediction of coiled-coils in other proteins. *Biosci. Rep.*, **2**, 1017–1024.

Peckham,M. and Knight,P.J. (2009) When a predicted coiled coil is really a single α-helix, in myosins and other proteins. *Soft Matter*, **5**, 2493–2503.

Schiffer,M. and Edmundson,A.B. (1967) Use of helical wheels to represent the structures of proteins and to identify segments with helical potential. *Biophys. J.*, **7**, 121–135.

Sivaramakrishnan,S. *et al.* (2008) Dynamic charge interactions create surprising rigidity in the ER/K alpha-helical protein motif. *Proc. Natl Acad. Sci. USA*, **105**, 13356–13361.

Süveges,D. *et al.* (2009) Charged single alpha-helix: a versatile protein structural motif. *Proteins*, **74**, 905–916.

Trigg,J. *et al.* (2011) Multicoil2: predicting coiled coils and their oligomerization states from sequence in the twilight zone. *PLoS One*, **6**, e23519.

Vincent,T.L. *et al.* (2013) LOGICOIL–multi-state prediction of coiled-coil oligomeric state. *Bioinformatics*, **29**, 69–76.